Coherence in typeface design: visual similarity of characters in Cyrillic, Devanagari, and Latin

David Březina
September 2018

Thesis submitted for the degree of Doctor of Philosophy
Department of Typography & Graphic Communication

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

David Březina

This research project was funded by the Arts & Humanities Research Council under the auspices of the Design Star Doctoral Training Centre.

Thesis typed and typeset by the author in typefaces
Yrsa, Skolar Sans, and Input.

To mum, dad, and Kuba.
And to Anna who supported me
all the way, although I made it hard.

Acknowledgements

I would like to thank Dr Mary C. Dyson, Dr Fiona G. E. Ross, and Keith Chi-Hang Tam for being the best supervisors I could ask for.

Also, I will always remember the hospitality and encouragement of the ANRT, Thomas Huot-Marchand and Bart Lamiroy in particular.

To Andrea Churchill Wong, Chris Earl, Maxim Zhukov, Maria Doreuli, Nils Frahm, Kiasmos, and my PhD colleagues: thank you for all your help and moral support.

Abstract

This thesis explores the visual similarity that underlies the coherence in the design of individual typefaces. Typeface designers aim to achieve a unifying coherence in their typefaces, so that characters can be identified individually as well as belonging together giving rise to an overall style. The objective is to determine whether the coherence perceived by readers differs from the coherence intended by designers. The research is cross-disciplinary, combining empirical studies of readers' perceptions with a computational model that is based on relevant typeface design knowledge.

Character similarity is studied in multiple different typefaces (fonts) intended for continuous reading in Cyrillic, Devanagari, and Latin scripts. The studies were conducted online to collect a large number of responses. The participants were presented with a sequence of character triplets. They were asked to identify the odd one out in each of these triplets judging by their visual similarity, thus making a statement about the similarity of the two complementary characters. This method studies the similarity in context, which provides more refined details about participants' similarity judgements.

The model interprets characters using two kinds of features: more specific parts and more general roles. The model learns the relative saliences of these features from a subset of the data collected in the studies. This allows the model to predict participants' responses to the triplets from the studies and for other, unseen triplets. Additionally, the model can provide explanations of the criteria participants used in their similarity judgements and can generate similarity matrices.

The model achieved high scores when predicting response probabilities and identifying the overall odd ones out. A view of coherence that is supported by readers' perception can be used to assist designers in their creative process, help with fonts' quality assessments, and contribute to readability research and multi-script typography.

Contents

1. Introduction 8
1.1. Research questions 11
1.2. Method 12
1.3. Impact 14
1.4. A few notes about the scope and style of this thesis 17
1.5. Note on notation and figures 17

2. Theory and definitions 18
2.1. Character 18
2.2. Character concept 20
2.3. Writing and script 25
2.4. Typeface 26
2.5. Similarity and coherence 28
2.6. Style 29

3. Overview of the scripts studied 31
3.1. Latin 32
3.2. Cyrillic 36
3.3. Devanagari 40
3.4. Representative characters 46

4. Approaches to character conceptualization 48
4.1. Shapes as contours 49
4.2. Shapes as strokes 51
4.3. Transformation of shapes 56
4.4. Composition of shapes 58
4.5. Generalization of shapes 66
4.5.1. Typographic attributes 66
4.5.2. The role hypothesis 67
4.6. Towards modelling of coherence 70

5. Studies with readers 71
5.1. Asking about similarity 71
5.1.1. Enquiry about parts 72
5.1.2. Enquiry about similarity 74
5.2. The approach taken 76
5.2.1. Triplet sequence and character pair similarity 77
5.2.2. Presentation 78
5.3. Organisation of the studies 81
5.3.1. Questionnaire 81
5.3.2. Triplet trials 84
5.3.3. Language variants 86
5.4. Scope 86
5.4.1. Typefaces 86
5.4.2. Characters and triplets 87
5.4.3. Kinds of triplets included 95
5.5. Technical implementation 96
5.6. Pilot study 97
5.7. Participants 97
5.7.1. Recruitment 97
5.7.2. Conversion rate and dropout 97
5.7.3. Devices used 98
5.7.4. Comments from participants 98
5.8. Pre-processing 99
5.9. Data set overview 100
5.10. Participants' background data 102
5.11. Response data 106
5.11.1. Density of the data 106
5.11.2. Response frequencies 107
5.11.3. Participants' agreement and triplet difficulty 108
5.12. Specific character pair similarity measure 111
5.13. Generalized character pair similarity 113
5.14. Character distinctiveness and perceivability 119
5.15. Comparison of responses 120
5.15.1. Comparison of repeated triplets between typefaces 120
5.15.2. Agreement between groups of participants 123
5.15.3. Comparison between experts and non-experts 124
5.15.4. Comparison between native and non-native, and fluent and non-fluent participants 128
5.16. Visual vs. linguistic interpretation 132
5.17. Possible methodological improvements 134
5.18. The use of the data 135

6. Modelling character similarity and coherence 136
6.1. Modelling approaches 137
6.2. The proposed model of coherence 145
6.3. Character conceptualization 149
6.3.1. Salience 152
6.3.2. Parts 152
6.3.3. Roles 161
6.4. Concepts' similarity 167
6.4.1. Generalized approach 167
6.4.2. Contextual approach 169
6.5. Model training 171
6.6. Technical implementation 172
6.6.1. File storage 172
6.6.2. Data structures 174
6.6.3. Calculating predictions 175
6.7. Model assessment 179
6.7.1. Convergence of the model 181
6.7.2. Cross-validation 183
6.8. Predictions and explanations 187

7. Conclusion 193
7.1. The theory 195
7.2. Study format and method 196
7.3. Data set 197
7.4. Discussion 197
7.5. Harmonious and essential future prospects 198

Bibliography 199

Appendix 1: character shapes studied 213
Appendix 2: observations and predictions 228
Appendix 3: similarity matrices 334
Appendix 4: cluster plots 340
Appendix 5: distinctiveness measures 346
Appendix 6: typeface comparisons 350
Appendix 7: concept diagrams 372
Appendix 8: typeface mappings 405
Appendix 9: Python code 415
Appendix 10: similarity matrices compared 443

1. Introduction

Whenever there is a series of products, be it furniture, icons, books, or letters, designers aim to achieve some level of unifying coherence among the individual items in the series. In typeface design, one of the important qualities of professional typefaces is that the shapes of characters, although they can be identified individually, can also be perceived as belonging to a single set. This fundamental attribute of quality typefaces has already been mentioned in early treatises dedicated to the craft, such as those by Joseph Moxon (1703) and Pierre-Simon Fournier le jeune (Fournier, 1706; Carter, 1995). It was mentioned more recently by Fernand Baudin, who stated: "For the graphic designer and the typographer, [the alphabet] is a system of signs with a visual logic all of its own." (Baudin, 1989, p. 23). The ability to create a sense of coherence across different character shapes is an essential skill that typeface designers need to acquire together with the ability to create characters that are legible and readable.[1]

Besides adhering to tradition rooted in handwriting and other forms of manual lettering, there are at least three motivations for designers to create typefaces with a visual coherence. The first one is the association between coherence and high-quality typefaces already mentioned above, i.e. that coherent typefaces are seen as being of quality. Note that whether coherence constitutes a positive aesthetic value depends on the worldview of a particular typeface designer and contemporary design trends. The second motivation seems to be user-centric. In any typographic design, typefaces are important building blocks contributing towards the overall style of a document. In the complex environment of typographic design, they become tools, and need to function predictably so that the designers who use them have firm control over the aesthetics and atmosphere of the whole work. For example, it would be a challenge for a graphic designer to use a typeface that would sometimes appear formal and other times informal, depending on the text it is used to set. The third motivation is related to effectiveness of production. The process of designing a typeface can be very tedious since it typically involves drawing and redrawing hundreds of character shapes. Occasional reuse of existing shapes and sub-shapes within a typeface helps to reduce the workload. Such reuse may involve sub-shapes being copied verbatim from one character to another, but in professional text typefaces this typically involves subtle modification to adapt the copied sub-shapes to a new visual context. In this respect coherence can be seen as a side-effect of striving for increased effectiveness.

1. The definitions of legibility and readability used here are by Walter Tracy (2003, p. 31). *Legibility* refers to the ease of recognition of character shapes while *readability* refers to the ease of comprehension of texts or documents set using character shapes produced using a particular typeface.

This does not necessarily mean that all typefaces have to be as coherent as possible. Sometimes, incoherence is called for by the designer or the project brief. It is simply to say that there are certain perceived benefits to coherent typefaces. Furthermore, it should be noted that it remains to be reliably determined whether coherent typefaces are easier to read. Psychological studies dealing with so-called font tuning (Sanocki, 1987; Sanocki, 1988; Sanocki, 1992; Gauthier et al., 2006; Walker, 2008; Sanocki & Dyson, 2012) provide evidence suggesting that there might be a correlation between typefaces' uniformity and speed of reading. Experiments attempting to link coherence and legibility or readability, however, require a clear, formal description of coherence in the first place. This research does not try to question the motivations described above. Instead, it attempts to analyse how typeface coherence is realized, and provide a formal description so that more productive discussions about the utility of coherence can be had.

In this work, coherence is understood as an overall quality that refers to a group of shapes. Similarity, on the other hand, refers to relationships between individual shapes in pairs in particular. The main challenge is to provide a detailed description of similarity judgements involved in perceived coherence. Rather than stating that letters a, b, d look coherent, the research aims to clarify what it is that makes the individual character pairs a, b, b, d, and a, d look similar in this context, to what extent they are similar, and what this means for the overall level of coherence. It is important to emphasize that there can be different degrees of coherence (see figure 1.1). Moreover, the term coherence in this work refers specifically to similarity relationships among character shapes from a single typeface. Coherence and similarity among characters from different typefaces is not studied.

FIGURE 1.1: The degree of apparent coherence among the character shapes in each line decreases from top to bottom. The effect is caused by deliberate mixing of different typefaces in each line. The first line uses shapes from a single typeface while the second and third lines mix multiple different typefaces. The differences in the second line are subtle and hard to spot at first. The third line also includes shapes with different stroke thicknesses which makes spotting incoherences easier.

To be applicable in practice, such descriptions should be relatable to typeface design processes. The aim is to also describe coherence in a general way that would work for more than one typeface or script. Therefore, such a description should be as independent of a particular technology as possible. This also suggests that the viewpoint from which coherence is studied should be that of the human, not of the machine. The research considers the reader's perspective and the way readers perceive and interpret the appearance of characters. The typefaces are studied as perceived by experienced readers and not, for example, by beginner readers.

The aspiration for a general theory is also a reason for conducting the work with multiple scripts. Three scripts – Cyrillic, Devanagari, and Latin (see figure 1.2) – have been selected for analysis in order to set up a challenging domain for the theory: Latin as a script for which numerous character representations and typographic systems have been devised; Cyrillic as a script that bears partial resemblance to Latin, yet remains significantly different in some respects; and Devanagari as a script clearly distinct from Latin and Cyrillic, both visually and linguistically, as Devanagari is a syllabic script.

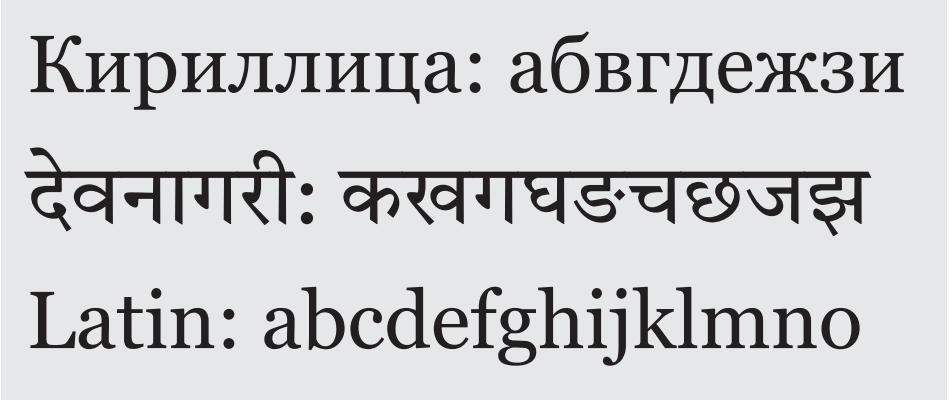


FIGURE 1.2: From top to bottom: Cyrillic, Devanagari, and Latin scripts in typefaces Georgia (Cyrillic and Latin) and Murty Hindi (Devanagari).

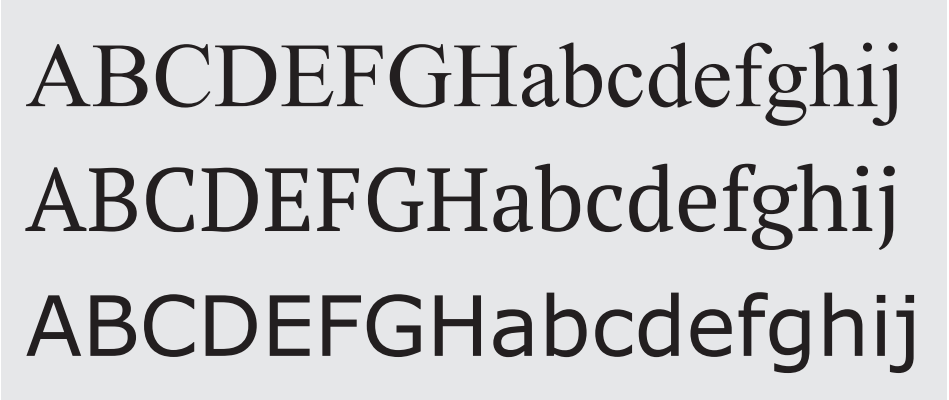


FIGURE 1.3: Common text typefaces represent more conventional designs: Times New Roman (top), PT Serif (middle), and Verdana (bottom).

The focus is on text typefaces, that is, typefaces that are generally recommended for longer, continuous reading, usually set in sizes around 9 to 14 points and used, for example, for longer articles, books, newspapers, or theses, printed as well as on the screen. The main reason for this choice is that text typefaces are considered to be conventional interpretations of the script they represent. Therefore, there is a greater chance that their study will contribute more towards fundamental ideas about scripts than would a study of unconventional designs (compare figures 1.3 and 1.4).

Notably, the research is not attempting to create a new way of designing typefaces or a model of character recognition. The aim is to describe the coherence in existing typefaces and draw more general conclusions about conventional relationships between characters within the selected scripts. These relationships will be referred to as *script conventions* in this work.

FIGURE 1.4: More unconventional typefaces: Alcoholica (top), Sutturah (middle), and Anglaise (bottom).

1.1. Research questions

The aims of this research can be summed up by the following research questions:

i. What are the principles underlying the coherence in design of individual typefaces as perceived by readers?
ii. Can a single theory describe perceived coherence in typeface design in different scripts? In other words, is it possible to develop means of commenting on character similarity relationships in typeface design for more than one script?
iii. Does the coherence perceived by readers differ from the coherence intended by designers?

1.2. Method

To explain typeface coherence, i.e. coherence of character shapes produced using the same typeface, one has first to explain how character shape similarity works. The study of similarity between character shapes is therefore central to this enquiry. The research consists of two main components:

i. testing with readers as participants who provide data about their perception of similarity,
ii. a theoretical, computational approach to similarity and coherence related to expert knowledge about character shape relationships and construction.

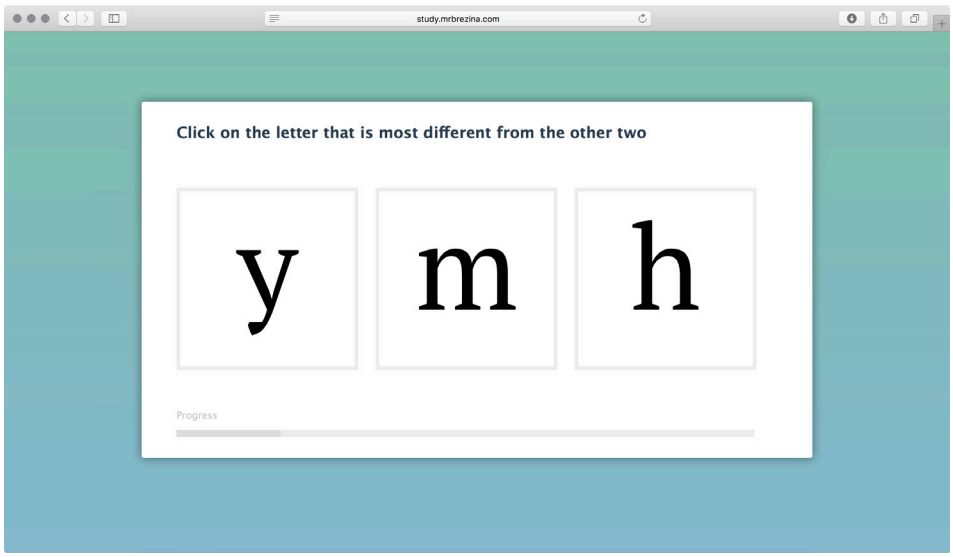


FIGURE 1.5: Participants in the studies were asked to select the most different character (the odd one out) from each in a series of character triplets.

The first component, testing with participants, is further discussed in chapter 5: *Studies with readers*. By way of introduction it should suffice to say that in studies conducted online, experienced readers performed a series of simple discrimination tasks on triplets of character shapes produced using the same typeface (see figure 1.5). For each triplet, e.g. y, m, h, they were asked to identify the character they found visually most different from the other two (the odd one out). The collected data has been analysed and used to derive response frequencies describing perceived similarity of character pairs within the context of the individual triplets. This approach to testing allowed the presentation of the character shapes without any degradation and at a comfortable size.

The second component involves a construction of a theoretical, computational model of coherence based on expert knowledge. Characters are interpreted using mental concepts and techniques typeface designers use when creating typefaces. Character shapes are formally conceptualized in terms

of their features: parts, which are more specific, and roles, which are more general. The resulting concepts are included in the computational model and used to calculate predictions regarding participants' responses from the study. Assessment of the model is by extension also a test that investigates to what extent expert knowledge can predict the observed similarity judgements of experienced readers. The typeface-design and typographic expert knowledge is further discussed in chapter 4: *Approaches to character conceptualization* while the approach to modelling and the assessment of the resulting model is detailed in chapter 6: *Modelling character similarity and coherence*. The overall research design is in figure 1.6.

It is important to consider that the model is developed independently of the designers of the studied typefaces. As such, it is only an approximation of the designers' original intentions. Umberto Eco (2015) distinguishes between the intention of the author, i.e. the typeface designer, the intention of the reader, and the intention of the text, i.e. the typeface in this context. In this sense, the results of studies with participants represent the perspective of the reader while the theoretical model stands between the perspective of the designers and that of the typefaces themselves.

As will be discussed in the conclusion, aggregation of the data collected in the studies and the generalisation of this data using the model can inform our understanding of coherence in the typefaces studied as well as the scripts' conventions.

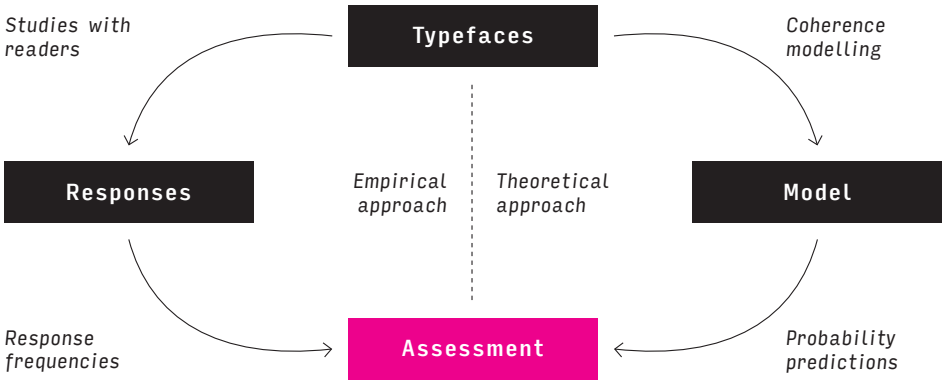


FIGURE 1.6: The design of the research consists of studies with participants (empirical approach) and modelling character similarity and coherence (theoretical approach). Results from both of these components are compared in the final assessment of the model.

1.3. Impact

One of the aims of this research is to develop a theory for a systematic analysis of coherence. A theory that would be related to perception as well as typeface design practice and could provide a basis for a subsequent formal perspective of typeface design. A perspective that could be used alongside the already well-established historical and technological views. More generally, the study of coherence can contribute to any field dealing with the question "What makes an 'a' an 'a'?". The font-tuning studies mentioned above (Sanocki, 1987; Sanocki, 1988; Sanocki, 1992; Gauthier et al., 2006; Walker, 2008; Sanocki & Dyson, 2012) as well as other studies (Palmer, 1999) show that the recognition of characters is influenced by context, both visual and linguistic (see figure 1.7). In this respect, the study of coherence can be seen as a study of the visual context created by typefaces and their designers. Although not providing the complete explanation of the processes involved in character recognition, better understanding of coherence could help to address the question about the essence of characters which is an important topic in epigraphy, linguistics (graphemics), psychology (letter-recognition and readability research), and cognitive science.

FIGURE 1.7: Stephen Palmer (1999, p. 429) shows that the identification of characters is influenced by context. English readers should not have a problem reading the two words above as "THE CAT". Note, however, that during the reading process, a single shape has been identified as two different characters depending on its context. H in "THE" and A in "CAT" both use the same shape. This shows that lexical knowledge can be an important factor in character identification, among others.

One of the challenges in readability research is the verification or rejection of the aforementioned question of whether coherent typefaces are easier to read. A theory of coherence that would correspond to typeface-design practice and allow measurement of coherence should help with articulating the question in a formal and testable way.

From a typeface-design perspective the theory could feed into font production processes, typeface design education, quality assessment of fonts, as well as multi-script typeface design. A few directions for possible contributions are suggested in the following paragraphs.

Greater formality for discussing coherence supported by empirical evidence will help to make communication of visual relationships between characters clearer. This will be beneficial in typeface design education, as

well as in practice, especially when dealing with large projects that require a coordinated design effort. The methodology described in chapters 5 and 6 could also help in situations where design conventions and processes for particular scripts are currently not well documented.

In his article *Lessons learned from Metafont*, Donald Knuth (1985) states that Metafont's "ability to capture the designer's intentions rather than just the drawings that result from these intentions [...] will prove to be much more important than anything else". He seems to be referring to intentions visualized in character drawings, but not necessarily captured in a formal way. While it is questionable whether any computer system can ever capture the sheer complexity of any designer's intentions, Knuth's observation that typeface design is more than a set of independent drawings is acute. A tool that ought to assist designers during their creative process might, and perhaps should, go beyond mere assistance with construction of individual drawings. Coherence that spans across different character shapes definitely belongs among designers' intentions which are not usually captured in font production tools.

Understanding coherence as it is perceived by readers will help designers to better distinguish their design-specific ideas from generally accepted and expected similarity relationships. This will lead to a greater effectiveness in their creative process.

Contemporary font-development software might implement the presented approaches on two levels. First, tools could provide real-time auditing and suggestions by comparing the work in progress with predetermined script conventions. Second, tools could use these conventions in an implicit way to generate new character shapes based on the ones that have already been designed. A reasonable implementation of such an approach is up for discussion. On the one hand, software tools could become more effective and remove or at least alleviate some of the tedious and repetitive aspects of typeface design work. On the other hand, each typeface design tool must consider design freedom and intrinsic variety of script conventions. Insensitive application of the theory might lead to overly restrictive tools that make interrogating conventions more challenging if not impossible.

Independence from technology is an important attribute of the theory as it allows for discussion of typefaces and their appearance across different platforms. Although contemporary font formats promise to be transferrable across various reproduction environments, little control over their appearance in these environments results in incoherencies that may affect reading experience. This issue of translatability of typeface design has been discussed in detail by Richard Southall (1986, 1991, 2005). The ability to encode coherence among characters may be the first step for typeface appearance to become

truly cross-platform and consequently ready for any future technologies.[2] Further, it will improve quality assessment by allowing comparison of some of the designer's intentions with the reproduced results.

Better understanding of coherence is also of great importance in multi-script typeface design, for which one of the key questions is a stylistic harmonisation of two or more scripts. Formalizing principles of coherence in scripts individually and in a coordinated way should establish the conditions for addressing the script harmonization challenge, and improve methodology and discourse within multilingual and multi-script typography.

1.4. A few notes about the scope and style of this thesis

When writing this thesis, the main challenge lay in the fact that it is cross-disciplinary. However, this also constitutes what I hope will end up being its important contribution. Chapter 2 sets out terminology to make the work relatable to linguists, psychologists, and typographers alike. Chapter 3 clarifies the basics of the scripts studied both from the linguistic and typographic point of view. Chapter 4 contains an overview of the common typeface design approaches and tools for conceptualizing character shapes. These shorter chapters set the context while the central focus of the thesis is in the following two chapters. Chapter 5 describes the testing methodology and statistical analysis used for the studies with readers. Finally, chapter 6 uses data analysis with a little bit of machine learning to explain the cognitive processes involved in judgements of character similarity. This summary promises an exciting mixture of topics and perspectives. However, this mixture also makes it difficult to write a text that is both relevant and digestible to typographers, typeface designers, linguists, psychologists, and data scientists. I have made an effort to write a text that is as readable as possible in parts where approaches from potentially unfamiliar fields are presented, but that is not overly simplistic in parts where familiar approaches are discussed. More technical considerations that are used to support the main argument and are not essential for the theoretical narrative are extracted into individual sections, footnotes, or appendices, so those who are not keen on reading them can skip them easily and not lose the thread of the central argument.

2. For example, rasterization of contemporary fonts on low-resolution screens sometimes results in an incoherent treatment of features that should remain coherent, e.g. the thickness of strokes or serifs varies across different characters from the same typeface. In contemporary font formats based on contour descriptions this issue is partially solved by the introduction of an additional layer of hinting instructions. Even though they solve the rasterisation problem sufficiently, the hinting instructions are still an incomplete description of typeface coherence. Note that rasterization does not seem to be an issue in Metafont which deals with the issue in a more comprehensive way. Nonetheless, the various issues of rasterization are not the subject of this research.

1.5. Note on notation and figures

Throughout this thesis, specific character shapes the readers should look at as if they were illustrations are marked in text using `a grey background`. Note that `a, b, c` refers to a set of shapes (a triplet in this case) that should be considered together while `a`, `b`, `c` refers to three shapes that should be considered individually. When referring to general character categories or their associated sounds, single quotation marks are used in a similar fashion: ‘a, b, c’ for a set and ‘a’, ‘b’, ‘c’ for three individual categories. In case there is a need for a reference to a character in a figure, single quotes are used. The brackets are used for transliteration or pronunciation.

Unless stated otherwise, images from other publications are reproduced at 100 % scale.

2. Theory and definitions

The fields that deal with subjects related to this research – fields such as cognitive psychology, linguistics, computer science, epigraphy, and typography – use varied and often incompatible terminology. In order to to allow discussion of character similarity, typeface coherence, script conventions, and other ideas presented here, this chapter puts forward terminology that is based on existing notions within the aforementioned fields, but redefined to provide a clear perspective that is based on the visual aspects and accounts for multi-script and multilingual considerations. Corresponding terms from other fields are mentioned where appropriate.

2.1. Character

The term *character* is problematic as it is not always clear from its definitions whether it refers to a visual or non-visual entity. *The Oxford English dictionary* (2018) entry for the term is extensive, but effectively it defines character either as "a member of a set of symbols used in writing or printing to represent linguistic elements, as individual speech sounds, syllables, or words; any of the simple elements of a written language, as a letter of an alphabet, or an ideogram." or, in computing, also as "a representation of such a symbol by means of a small number of bits, holes in punched tape, etc.". *The Unicode standard* (2014, p. 6) states the character identified by a Unicode code point is an abstract entity (e.g. Devanagari digit two) and the mark made on screen or paper, called a glyph, is a visual representation of the character. *The Unicode Standard* definition and the *Oxford English dictionary* definition for computing are non-visual. On the other hand, in typography or epigraphy, characters are usually visual representations.[1] This work adheres to this meaning as well.

Another question is whether a character is a general or specific entity. This is resolved here by stating that these are essentially two different modes of use of the same term. In a wider sense, a character is general and refers to a category of symbols representing the same information, typically a linguistic unit, e.g. all letters 'b' or any letter 'b'. In the most narrow sense, it refers to a specific symbol, a material object, e.g. the letter `b` that you have just read. To distinguish between these two modes of use when necessary, two compound terms are used: the general *character category* and the specific *character shape* (see figure 2.1). The set of all character shapes from the same character category is called a *character category domain*.

1. To be precise and not to discriminate against blind people, one should in fact speak of visual or tactile representation in typography or epigraphy. However, this work focuses only on the visual perception of characters.

It is worth noting that character shapes from the same category can look very different depending on their styling (see figure 2.2), colour, scale, distance, geometric and other manipulations (e.g. those provided by DTP software), but also depending on the reproduction technology used. Various kinds of screens, rasterizers, and printers each have different effects which show in the resulting shapes.

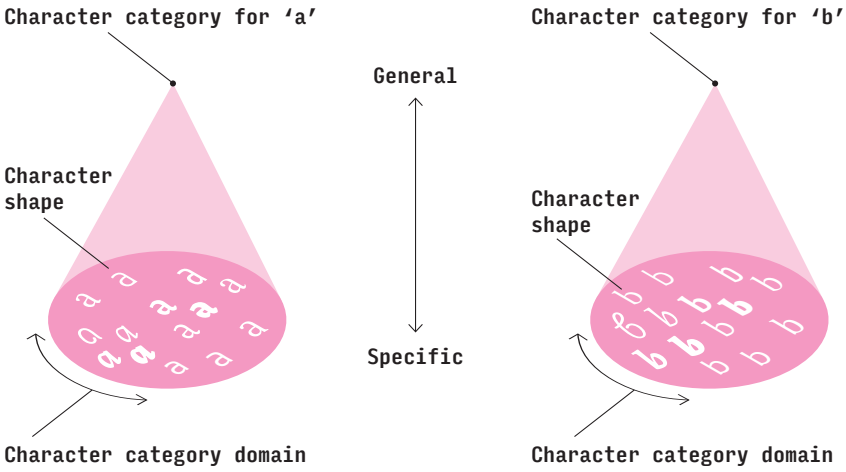


FIGURE 2.1: Two modes of using the term character. General character categories (illustrated as points of the cones) contain specific character shapes. The character category domain (bases of the cones) comprises all character shapes that can be recognized as belonging to the character category.

Importantly, a character always represents a conventional linguistic unit. Based on the structural level of the associated linguistic unit, one can distinguish several kinds of characters (Coulmas, 1989). There are three kinds that are used in this research:

i. a *letter* is a character representing one or more of the sounds used in speech, e.g. the Latin-script letter b,
ii. a *syllabic character* (or *syllable*), is a character representing a syllabic sound, e.g. the Devanagari syllables ब [Ba] or ड्ड [DdDda],
iii. an *orbital sign* is a character that cannot be used on its own, but can be combined with other characters to modify their meaning, e.g. the caron ˇ accent as used in letters such as ř or the Devanagari vowel sign ी [-ii] as used in combinations such as पी [Pii].

This list is not meant to be complete. It leaves aside logograms or ideograms such as those used in Chinese and Japanese scripts, for example, because these scripts are not discussed in this work.

In a multilingual environment, the situation is more complex as the same character can represent different linguistic units depending on a language, e.g. the character shape c represents a different sound in English compared to Czech. Nevertheless, this seems to be a linguistic issue as the identity of a character remains the same: c would be identified as the same symbol by both

English or Czech readers no matter what language context it were used in. In other words, a character is a token associated with a linguistic unit, but this association is merely conventional (see also Saussure, 1972). The linguistic unit itself does not prescribe the appearance of its representative shape nor vice versa, the script conventions do. Should the c go missing from the repertoire of a favourite font, another symbol could be used in its stead if this was agreed upon (conventionalized).

It is also important to note that an uppercase (or capital) letter and a lowercase (or small) letter representing the same sound are considered two distinct character categories in this research. In other words, the linguistic unit they represent is different. For example, in the Latin script, character shapes A and a belong to two different categories, while a and *a* belong to the same category.

There are similar terms in other fields. The term character category roughly corresponds to the linguistic term *grapheme* when this is defined as the smallest functional unit of writing (Coulmas, 1996, p. 174). On the other hand, cognitive psychology often uses the term *abstract letter identity* (see for example: Kinoshita & Kaplan, 2008) which seems to refer to all letters representing a particular sound regardless of case, i.e. it could include multiple character categories. For example, A and a would have the same abstract letter identity.

2.2. Character concept

Readers that are not very familiar with typeface design and typography often assume that there is a single shape-like idea prescribing the appearance of all character shapes from a particular category. However, after careful considertation, a single invariant prototype based on shapes seems untenable.

There are several problems that make the task of describing all shapes in a character category really difficult. Firstly, character categories, although illustrated here with simple diagrams with well-defined borders between them, are ever-evolving social constructs. Thus, they also depend on perceptions of individual readers. What might pass as an ‘a’ for some, others might not perceive as one. Secondly, designers keep on producing new typefaces and letterings. New radical designs constantly push the boundaries of what are acceptable members of character categories. Thirdly, characters exist in the context of other characters. Sometimes knowing how the other characters look may help identify previously unseen characters, especially in less formal typefaces where characters might be easy to confuse. And lastly, the linguistic context has an effect on character identification as well. What is an ‘H’ in one context, can be an ‘A’ in a different context (see figure 1.7).[2] Thus, describing all the shapes in a character category is a task that

2. Note that character categories can also overlap when the same shape can be interpreted as belonging to two different character categories.

would involve not only describing the social construct, it would also require a working model of human perception and creativity in the context of character design, including all possible production processes. Clearly, this would be a difficult task.

Douglas Hofstadter (1982) challenged an idea that all character shapes belonging to a particular category could be described with a parametric shape-based prototype on the grounds of mathematical logic. Illustrating the difficulty of parametric approach with many variants of an 'a' (see figure 2.2), he suggested an interconnected web of general criteria that make up the idea of 'a' itself (Hofstadter, 1982). The exploration of the idea of 'a' and the automated creation of a coherent style were some of the objectives of the Letter Spirit project lead by Hofstadter within the Fluid Analogies Research Group in the 1990s and early 2000s (Hofstadter & McGraw, 1998; Rehling, 2001).

While it seems unfeasible to describe the whole domain of all shapes from a particular character category, it is possible to describe at least some subsets of the domain in order to make general comments about the character shapes included. Subsets of character shapes from the category domain can be described using *character concepts*. A character concept is yet another mode of using the term character. Essentially, it is any generalized description of character shapes. A character concept is not as general as a character category and not as specific as a character shape. There is no assumption being made regarding an approach taken to describe the shapes other than it has to focus on the visual aspects. It can be more or less general. The more general the concept, the larger the set of character shapes it describes, and so the larger the *character concept domain* (see figure 2.3).

Some character concepts can be used in the production of character shapes, e.g. in fonts which determine the shapes of characters in the context of reproduction technologies. However, other character concepts might not be suited for production, especially those that are very general. That being said, character concepts can always be used to judge whether a character shape belongs to their domain, that is, whether its shape complies with the concept's definition.

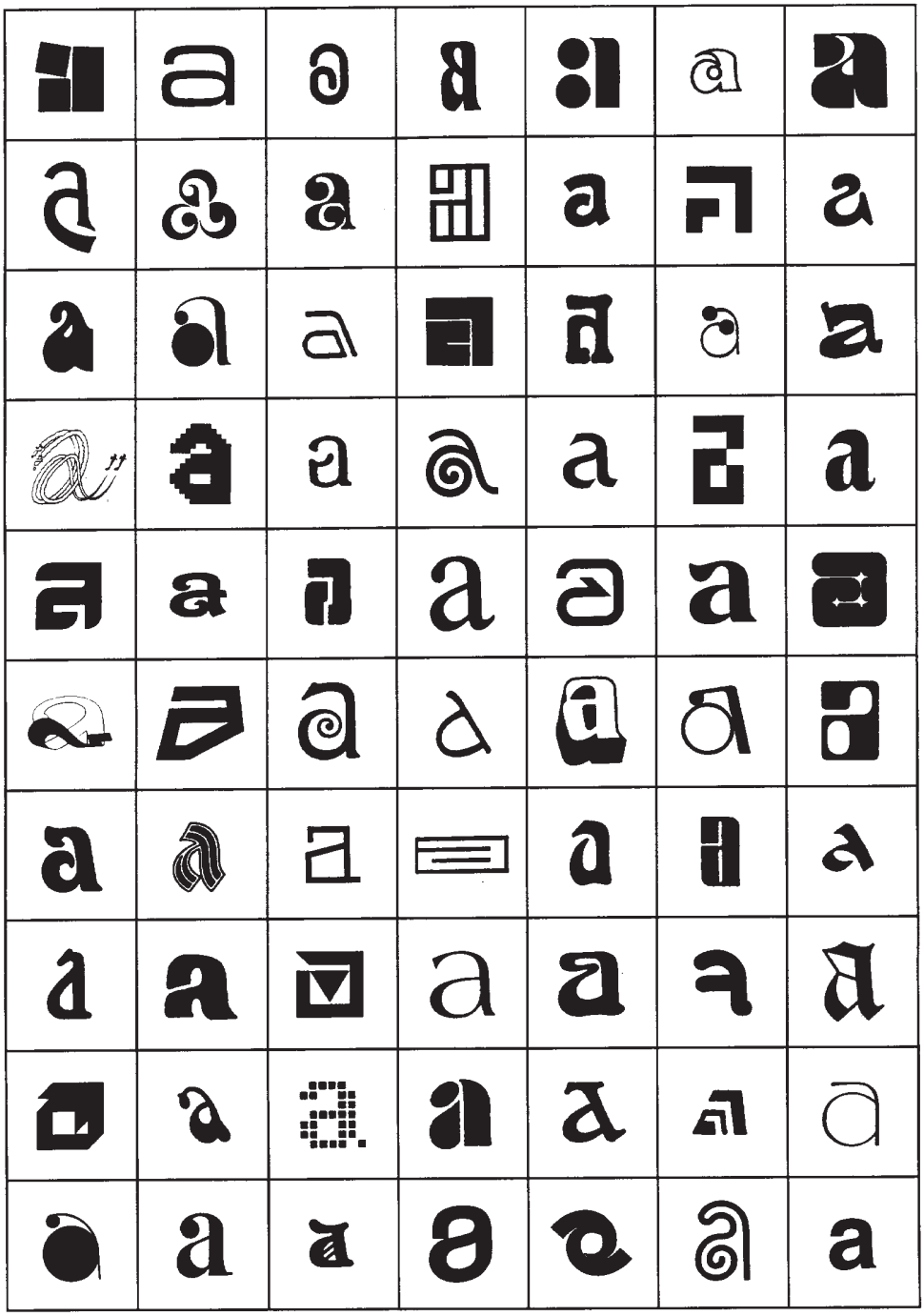


FIGURE 2.2: The variety of shapes among 70 different letters 'a' shows how challenging it is to find a unifying description for all possible shapes of 'a' (from: Hofstadter and McGraw, 1998, p. 413).

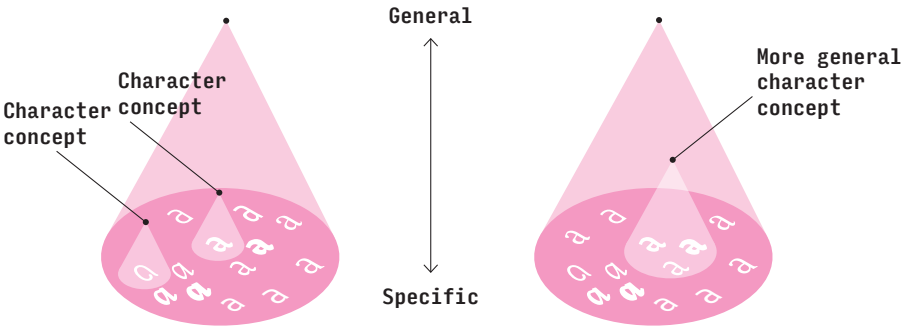


FIGURE 2.3: The domains of all potential character shapes (bases of the large cones) for particular character categories (points of the large cones) are compartmentalised by character concepts (points of the small cones) into smaller character concept domains (bases of the smaller cones).

For example, the character shape `a` could be conceptualized as "having an arch on the top that is smoothly connected to a vertical stem on its right side with a small loop tucked underneath the arch on the left side of the stem ..." either verbally or in some kind of formal system. Clearly, many other shapes of an 'a', would satisfy this fairly general description and thus fit within the domain of this concept.

On the other hand, an 'a' could be conceptualized by means of a much more precise geometric description of its contours, e.g. bézier curves in an OpenType font (Microsoft, 2018), which would significantly limit the number of all potential shapes of 'a' fitting this concept (see the 'a' on the left in figure 2.4).

It is worth noting that although geometric character concepts may seem very precise, they can still realise a set of different character shapes depending on the reproduction technology used. This has been thoroughly discussed by Richard Southall (2005) and others. Some coarse printers, for example those used in cash registers to print receipts, may produce very different results from high-resolution screens based on the same concept.

The geometric descriptions can be more general too: the concept of a letter 'a' in a variable OpenType font (Microsoft, 2018) that allows variation spanning from regular, to bold, and to black weight variant would be implemented differently from the example in the previous paragraph. The contours of the letter would be described using bézier curves and a set of values that specify how these curves change between the weight variants (see the three 'a's on the right in figure 2.4). The set of realised character shapes contains a multitude of weight variants of 'a' between regular, bold and black and their interpretations in various reproduction technologies. Thus, this character concept is more general.

Note that the more general character concepts can theoretically include other character concepts. In the examples above, all the character shapes of an 'a' realised using the single-style OpenType font can also be realised using the character concept for an 'a' from the Variable OpenType font (compare the concepts on the left with the concept on the right in figure 2.3 and both with figure 2.4).

Various ways of conceptualizing characters used in typeface design and font production will be discussed in chapter 4: *Approaches to character conceptualization*.

The idea of character concept is related to Debra Adams and Richard Southall's (1989) term "character appearance specification", Southall's (1991) terms "model" and "pattern", and McGraw's (1995, pp. 6–7) terms "letter concept, letter conceptualization, and letter plan". All of these are character concepts with varying levels of generalization.

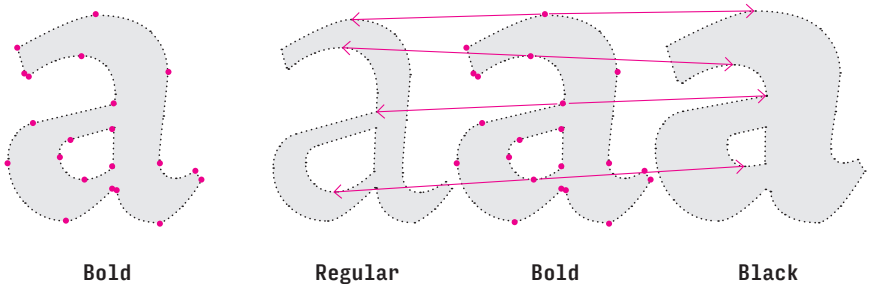


FIGURE 2.4: Single-style OpenType fonts (the single 'a' on the left) represent character concepts as a set of bézier curves. Variable OpenType fonts (the three 'a's on the right) represent character concepts as a set of bézier curves for one character with associated values that specify how these curves change between character variants (illustrated by the arrows), in this case changing the weight of the characters. Only some of the control points (marked in magenta) defining bézier curves are shown and the contours are dotted to illustrate that only the control points are included in the concept, not all the points on the curves. Only some of the arrows indicating the variability are shown here for the sake of simplicity.

The model proposed in chapter 6 uses simple character concepts based on expert knowledge in an attempt to describe criteria readers might use when judging similarity of character shapes. Clarifying how these concepts relate to readers' similarity judgements will help answer the third research question: whether the coherence perceived by readers differs from the coherence intended by designers (see chapter 1). Understanding the difference in perception of coherence between designers and readers is an important step towards a better understanding of perception of typefaces.

2.3. Writing and script

The term *writing* in this work includes any use of characters to convey a message, in print, on screen, or in handwriting. For writing to be deciphered, one has to know the linguistic units conventionally associated with characters. In the words of Gerrit Noordzij (1973, p. 80): "Writing is a convention. [...] Unconventional writing is no writing at all."

Writing system and *script* are two terms that are often used interchangeably although they can have distinct meanings. This work adheres to the definitions provided by Florian Coulmas (1989). He uses the term *writing system* to differentiate systems depicting linguistic units of different structural levels, e.g. alphabetic writing system, syllabic writing system, etc. Following his definition, the term *script* refers to a visual instantiation of a writing system, e.g. Latin script, Devanagari, etc. Further, he notes that the term *script* is often used as if scripts were inherently related to a given language and that this is only partly true. There are cases where one script is used by multiple languages, e.g. Devanagari is used for Hindi, Marathi, Nepali, and other languages. Admittedly, there are cases where there is just one language using a given script, e.g. Greek, or cases where one language uses multiple scripts, e.g. Serbian uses Latin as well as Cyrillic scripts. In addition, the rules for applying a script (its *orthography*) to a particular language may vary as well.

For the purpose of this work, and in line with Coulmas's notions, a script can be defined as a set of character categories used for writing, together with the general system of their use, i.e. orthography (see figure 2.5).[3] As the focus is on the visual, the scripts' orthographies will be discussed only briefly in chapter 3: *Overview of the scripts studied*.

3. The definition used in this work interprets a script as a contrast set. A contrast set is a set of categories that are mutually exclusive, i.e. contrasted, (Watson, 2005). Contrast sets are commonly used to describe various domains of human interest, e.g. temperature scales, colour palettes, countries, alphabets, and therefore should be a straightforward concept to work with. Using the concept of a contrast set will also allow for a clearer discussion about how individual character categories are formed and how they relate to each other.

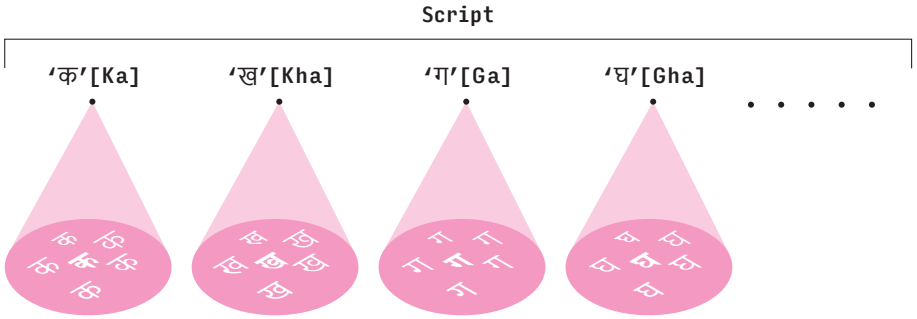


FIGURE 2.5: An example of a script represented as a set of character categories. The Devanagari script is represented here with only a few character categories and shapes. The orthography is not illustrated.

2.4. Typeface

According to Adams and Southall (1989, p. 213), "a typeface is a set of appearance specifications for the objects that realise the characters of a script." Note that the term character as it is used in their article refers to what is called a character category in this work. Defining a typeface this way is practical as it clearly links characters to typefaces and to scripts. This allows for descriptions of relationships between characters at different levels of generalization.

Adapted to the already established terminology and rephrased to work within the context of multi-script typography, Adams and Southall's definition would read as: A typeface is a set of character concepts, each of them specifying character shapes for one of the character categories from one or more scripts (see figure 2.6).[4] Notably, a typeface defined this way can contain character concepts from multiple scripts. For example, the typeface Arial contains character concepts used for Latin and Cyrillic scripts.

Character concepts for a typeface are typically implemented in a *font*. The approach to conceptualisation is usually shared for all character concepts implemented by a font and is determined by the *font format*.

Character concepts included in typefaces and fonts are the results of typeface designers' work. It is an inherent part of typeface designer's job to make sure the character concepts determine character shapes that can be recognized as belonging to the corresponding character category. Their failure at this task leads to less readable fonts which may lead to readers misreading the intended message.

4. It is worth noting that some designers (e.g. Bigelow, 1985) consider typefaces as coherent by definition. With such a definition, to be able to say precisely what is and what is not a typeface would require a definition of coherence and the criteria used to assess the presence or level of coherence in a typeface. In order to avoid any assumption regarding coherence, this research uses a definition where coherence is not considered a necessary quality of a typeface.

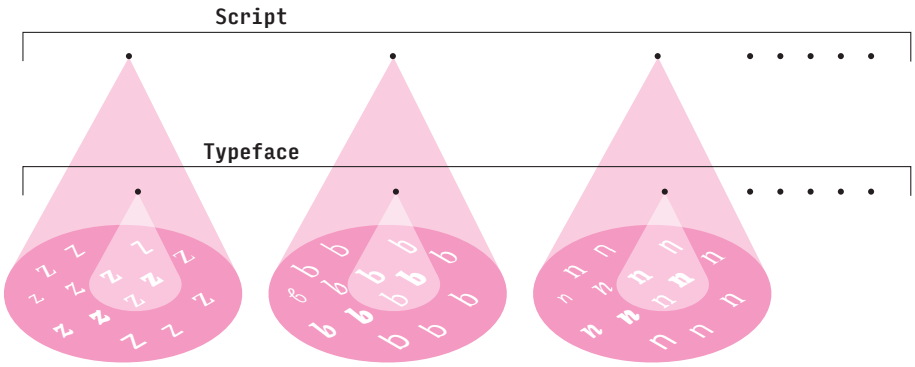


FIGURE 2.6: Relationships between a script consisting of character categories (points of the large cones), a typeface consisting of character concepts (points of the small cones), and character shapes (within the bases of the cones).

To avoid ambiguity, it is worth mentioning that the term *typeface* is commonly used in a broad sense by designers. The scope of this use has been thoroughly discussed by Robin Kinross (2002) without a definitive resolution. Fred Smeijers's (1996, p. 16) definition of a typeface as "a set of types over a whole range of sizes" seems to be the most generally accepted. According to him, the idea of a typeface emerged gradually around the 16th century when individual fonts for specific sizes began to be joined by a consistent, unifying notion of their forms. Later, variants, such as italic, small capitals, bold, light, and more, were included within the notion of a typeface as well. Thus, the term came to be used as a user-oriented umbrella term to suggest design coordination or compatibility of its constituents. To avoid confusion, the term *type family* is used for this broader meaning of the term (see figure 2.7).

Arek Regular
Arek Italic
Arek Semibold
Arek Semibold Italic
Arek Bold
Arek Bold Italic
Arek Extrabold
Arek Extrabold Italic

FIGURE 2.7: The type family Arek contains multiple variants, e.g. uprights (left) and italics (right) in the following weights (top to bottom): Regular, Semibold, Bold, Extrabold.

2.5. Similarity and coherence

Where similarity refers to relationships between individual objects, coherence refers to an overall quality. In this work *coherence* is viewed simply as a structure of individual similarity relationships between characters. Thus there can be a coherence of a group of character shapes, coherence of character concepts (i.e. *typeface coherence*), and coherence of character categories. The conventional coherence of character categories from the same script is called *script conventions*.

Defining characters as visual entities with different modes means that conclusions can be made about character concepts within a typeface based on the evidence about material character shapes. And analogically, conclusions about general character categories within scripts can be based on evidence about the character concepts or character shapes. Thus, the similarities between character shapes can be aggregated and taken as evidence of inter-concept and inter-category similarity relationships and coherence (see figure 2.8). Note that none of this would be possible if characters or character categories were defined as abstract entities.

The key to understanding script conventions and typeface coherence is character similarity which is studied empirically in chapter 5 and theoretically in chapter 6.

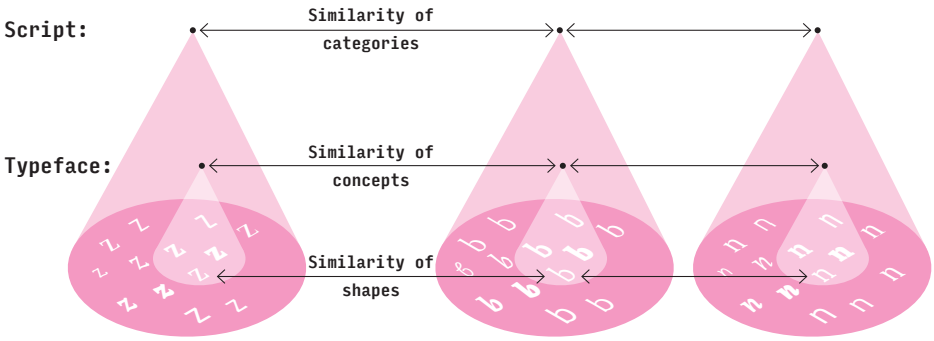


FIGURE 2.8: Similarity relationships between character shapes constitute the basis for similarity relationships between character concepts in typefaces and conventional similarity relationships between character categories from a script.

2.6. Style

The term *style* is commonly used with various meanings. It may be used to suggest something exceptional ("This typeface has style!"); it may refer to a particular period of time ("It has been designed in the art deco style."); it may refer to an artist's personal voice ("He has a peculiar style."); or it may refer to a particular variant of a typeface or a single-style font from a type family ("Let's use the bold-style font for the headline."). In addition, it may be understood as a non-essential part of a shape's appearance, particularly in visual communication and web design where styles are often defined independently of the shapes or content they are assigned to ("Make sure the footnotes and sidenotes have the same style applied to them, including the colour and rounded section border."). The many different usages of the term style make it difficult to use in academic writing. Nonetheless, it is an important term in design practice and it is useful to relate the present study of coherence to it.

For visual arts and for art history in particular, style is one of the central issues and there are several key texts that discuss it. In particular, Ernst Gombrich (2009) and Meyer Schapiro (1994) have achieved sufficiently general definitions which permit their transfer to typeface design.

Gombrich (2009, p. 129) defines style as: "a distinctive, and therefore recognisable, way in which an act is performed or an artifact made or ought to be performed and made". The stress here is on recognizability and distinctiveness rather than on similarity to other acts or artefacts. Gombrich develops this further by quoting Stephen Ullman (1957, p. 6): "The pivot of the whole theory of expressiveness is the concept of choice. There can be no question of style unless the speaker or writer has the possibility of choosing between alternative forms of expression."[5] In this view, an object can have a style only when it is clearly distinct from others it is compared to and when there is a variety of alternative forms—this is referred to as a *stylistic space* in this work.

In order to compare objects or forms to say whether they are similar or distinct, there has to be a set of criteria that allows for the comparison—a system that allows describing different objects in a compatible way in order to compare them. For example, one can imagine comparing a chair and table by saying they share the same upholstery and wood carving. In this case, the different pieces of furniture have been described as "objects with upholstery and carving" and that permits further discussion. It might

5. The word 'choice' may seem to suggest an author's ultimate understanding of the final object and control over its interpretation. It was already discussed in the introduction that this is not always the case. There is a difference between the intention of the designer and the intention of the reader.

seem that nothing in Gombrich's definition suggests that style can link objects together, but the need for comparison implies some form of linking.

Schapiro (1994, p. 51), on the other hand, puts the requirement for similarity in a more straightforward way: "By style is usually meant the constant form—and sometimes the constant elements, qualities, and expression—in the art of an individual or a group." Investigating style, according to Schapiro (1994, p. 59), is "often a search for hidden correspondences explained by an organizing principle that determines both the character of the parts and the patterning of the whole." And in the words of typeface designers Erik van Blokland and Just van Rossum (2004), "style in the broader sense could be the art of finding common threads in seemingly unrelated things and defining groups for them."

Thus *style* is defined in this work as a set of distinctive visual attributes of forms which permits their grouping with other forms. Style in typeface design is often described by means of typographic attributes (see section 4.5.1) such as the differences between thin and thick parts of the strokes (the contrast), the weight of the thick parts, height, etc. Nonetheless, these attributes are not applied equally to all characters from the same typeface (see figure 2.9). Different character categories make different allowances for the implementation of the global attributes and style as described this way. In order to understand style better, one ought to study individual characters and the attributes they can carry. In this respect, coherence can be seen as a structure describing the ways the attributes of a particular style are implemented in different characters.

Approached from a different point of view, typefaces do not have to be coherent, but their coherence is often imposed on them by a designer. And one of the motivations to do so, as discussed already in the introduction, is the appeal of a uniform style. In order to produce a uniform style, designers need to distribute the attributes of style consistently across the different character concepts. Coherence, in this sense, is the result of the work of a designer: a coordination of character concepts in order to produce coherent character shapes.

FIGURE 2.9: Global attributes are not applied equally to all characters from the same typeface. The tapering of strokes is present in i, n, x, but not in o. The difference between the thin and thick strokes is most apparent in o, x, less so in n, and not present in i. The example is in the typeface Optima.

3. Overview of the scripts studied

An introduction to the scripts studied is provided in this chapter with the goal of supporting some of the arguments regarding the testing with participants in the forthcoming chapter 5: *Studies with readers*. A brief glance over the scripts' historical backgrounds is given to provide context. However, the intention is not to paint a comprehensive linguistic and historical picture, but to focus on presenting the scripts as systems of shapes. To start with, a clarification of the reasons behind the selection of scripts is due.

The main motivation for looking at three different scripts is the potential transferability of the theory being tested. Diverse material enables one to draw more general conclusions, and by extension encourages a more flexible theoretical framework. Since different scripts challenge assumptions about writing tools and proportions differently, a theory tested with multiple scripts is more likely to be transferable to other scripts than a theory tested with a single script.

Three scripts were selected: Cyrillic, Devanagari, and Latin. The Latin script is used for English, Spanish, and many other languages. Cyrillic is a script used to write many Slavic and Asian languages while Devanagari is used to write many languages of the Indian subcontinent. See table 3.1 for an indication of the scripts' popularity based on the language survey Ethnologue (2018).

In terms of design principles, Latin and Cyrillic represent scripts with a shared legacy that are likely to be easily interpreted using the same theory. On the other hand, there are very few principles shared between these two scripts and Devanagari. To conclude, these particular scripts were selected not only because they are major world scripts, but also because they provide diverse and challenging study material.

Script	Language	Number of users
Cyrillic	Russian	265,026,130
Devanagari	Hindi	534,271,550
Latin	English	1,121,806,280

TABLE 3.1: Counting the number of users of a script is problematic as one has to consider various languages that make use of the script, censuses, and literacy rates. In order to provide some indication of the popularity of each of the scripts, the numbers of users of the three major languages, one for each of the scripts studied, are provided in the table. The numbers taken from Ethnologue (2018) are for users that have the language as their first or second language in all countries of the world. Note that these numbers may include illiterate users and that each of the scripts is used for more languages.

In terms of specialized literature, both theoretical and professional, there are various theories regarding the conceptualization and construction of Latin characters (see chapter 4: *Approaches to character conceptualization*). Comparatively, there seems to be fewer resources about the design concepts for Cyrillic or Devanagari. However, there is enough helpful informative documentation, for example Gordon's (2013) extensive analysis of Cyrillic or Naik's (1971) analysis of Devanagari.

It seems safe to assume that readers of this text are familiar with the design qualities of the Latin script. Yet, some are repeated in order to establish a shared terminology. There is no assumption about the reader's understanding of Cyrillic and Devanagari.

The general focus is on the scripts and their principles. When there is a need to be language-specific, rather than to tackle the sheer number and complexity of all languages each script supports, a major representative is used instead. English is used for the Latin script, Russian for Cyrillic, and Hindi for Devanagari.

3.1. Latin

The Latin script is an alphabetic writing system. It is usually written from left to right with lines of text arranged from top to bottom. Each character (letter) represents a consonant or vowel.

Latin is a *bicameral* script. This refers to the fact that it includes two variants for most characters, one uppercase (majuscule forms, also called capital letters) and one lowercase (minuscule forms, also called small letters). These two variants represent the same sound, but serve different roles in writing. The *basic Latin alphabet* contains 26 uppercase and 26 lowercase letters (see figure 3.1).[1]

Without any amendments, the basic Latin alphabet is used only for a limited number of languages. Additional characters can extend its symbolic repertoire to cater for a greater variety of sounds and languages. The shapes of additional characters are frequently conceived through a modification of an existing character shape rather than an invention of a completely new one. This can be done by manipulating the character shape itself, as seen in the letter ŋ (Latin small letter eng) for example. An alternative, yet equally common, approach is to add an orbital sign (accent) into the proximity of the base shape to create an *accented character*, as seen in the letter ř (Latin small letter r with caron) for example.

1. This character set is based on the ISO basic Latin alphabet and the Basic Latin as defined in the Unicode Standard (2014) in the block Controls and Basic Latin and in the ISO/IEC 8859-5 (1999). The alphabet for the Latin language would not require all of these characters.

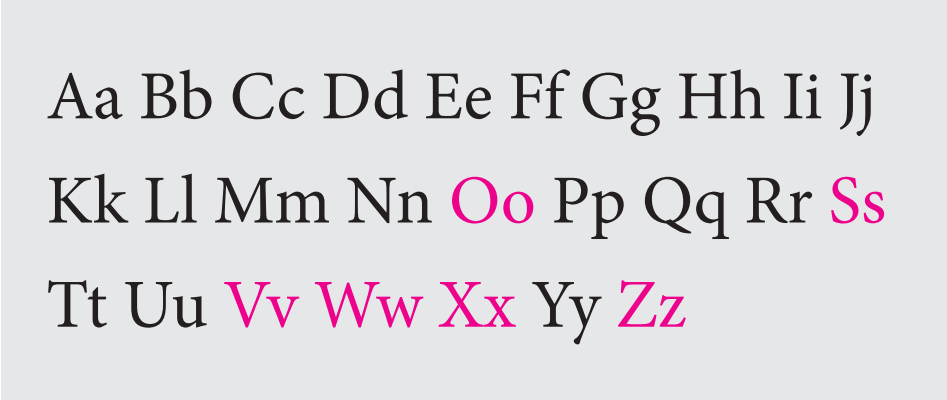


FIGURE 3.1: Uppercase and lowercase characters of the basic Latin alphabet presented in the typeface Minion. Character shapes that have the same spatial configuration of strokes and stroke endings between uppercase and lowercase are shown in magenta. Note that this correspondence may differ to some extent, depending on a particular typeface's design.

Other characters required for contemporary use of the script are punctuation characters, including word spaces and Latin numerals.[2]

As a system of shapes, the Latin script is a marriage of two approaches. The uppercase letters follow a convention established in Roman capitals. Those were drawn using a brush, each letter executed in multiple strokes, and then chiseled in a stone following a grid made of two guidelines (Catich, 1968). This form of production gave them a rather static appearance with often abrupt transitions between straight and rounded strokes.

In contrast, the lowercase letters derive their forms from Carolingian minuscule and similar writing styles (Clayton, 2014). The minuscule is written on a grid made of four guidelines. Unlike the majuscule, it is more fluid and the character construction makes use of fewer interruptions. Another typical characteristic is ascenders and descenders which extend above and below the base defined by the two central guidelines (see figures 3.2 and 3.3).

2. The numerals originated in India. Having been adopted by Persian and Arabic scholars first, they were, centuries later, popularised in Europe by Leonardo Fibonacci in the 13th century (Britannica, 2016). Deriving their name from this odyssey, they have been commonly called Hindu-Arabic or Indo-Arabic. Unfortunately, such names are confusing in the multilingual context. Hence, they are called Latin numerals in this work.

a b c d e f g h i j k l m n

o p q r s t u v w x y z

FIGURE 3.2: Latin lowercase letters from the basic Latin alphabet set in the typeface Minion. Letters with ascenders or descenders are marked with magenta. Accented letters are marked with green.

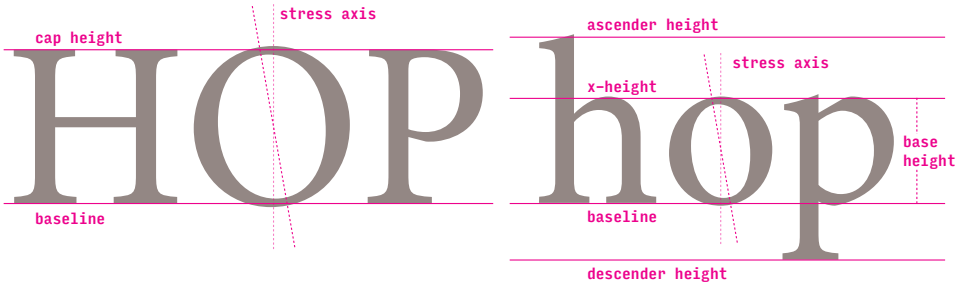


FIGURE 3.3: Latin uppercase letters (left) are aligned between two guidelines (cap height and baseline), the lowercase (right) makes use of four guidelines (ascender height, x-height, baseline, and descender line). The prominent metric, base height, is marked on the right side. The stress axis is conventionally diagonal or vertical. The example is set in the typeface Minion.

It is customary to use the uppercase for capitalization of names and titles[3], for abbreviations, for initial letters in sentences, and for emphasis. The rest of the text is usually represented in lowercase. As a consequence, the lowercase characters are comparably more frequent in general use.

For the sake of completeness, it is worth noting that some of the characters in use today, particularly the characters for extended language support, were introduced long after the basics of the Latin scripts were established. Yet most of them are designed to belong to either of the cases and to follow the same design principles.

The origins of the Latin script in inscriptional lettering and calligraphy have an impact on every kind of writing, including typography. In typography the conventions became even more constrained. This is often attributed to the fact that the printing press is an “innately more conservative medium” (Clayton, 2014, p. 58).

3. In some languages, such as German, the uppercase is used to capitalise nouns as well. This means that their frequency is higher when compared to English for example. However, it is still significantly lower when compared to the lowercase.

The conventional design principles of typical typefaces for continuous reading were established in the early Renaissance (Carter, 1969) and they are, for the most part, still used today. This happened under the significant influence of the then-contemporary formal humanist calligraphy which referenced both Roman capitals and Carolingian minuscule. First used in 1470 in a *Preparatio evangelica* (Eusebius, Georgius & Nicolaus, 1470), Jenson's Roman type is a representative of these already established design principles (see figure 3.4).

Notably, both Roman capitals and Carolingian minuscule make use of tools with a broad nib, such as brush or quill. The broad-nib model, or what would be in Gerrit Noordzij's (1991) terms the *translation* sort of contrast, defines the conventional distribution of thick and thin parts of strokes in the Latin script in calligraphy, epigraphy, or typography (see section 4.2 for more details). The stress of the stroke modulation is often diagonal or vertical with the thin parts being in the top-left and bottom-right direction (see the stress axis in figure 3.3).

However, the conventional design principles go beyond the broad-nib model. The formal humanist calligraphy of lowercase was adapted for typography by adding serifs that were already present in uppercase (Smeijers, 1996, p. 51). The spatial configuration of strokes and the form and configuration of stroke endings became an inherent part of the conventional design principles as well (Carter, 1969).

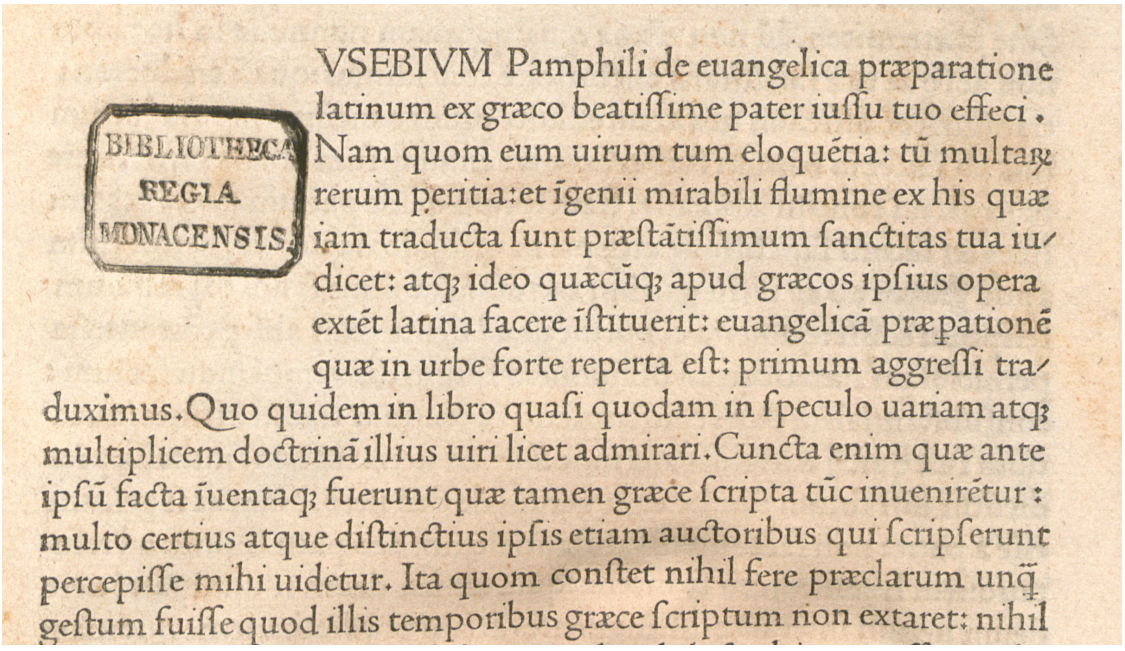


FIGURE 3.4: Jenson's Roman is representative of typefaces used for continuous reading today. The image shows a portion of a page from *Praeparatio evangelica* (Eusebius, Georgius & Nicolaus, 1470, p. 1) where this typeface was used for the first time. Image courtesy of the Bayerische Staatsbibliothek.

In the context of visual communication, it is sometimes important to get a measure of the text's appearing height. Getting an exact measure of a particular text would be too complicated and would indeed vary depending on the specific text and the frequency of ascenders, descenders, and accents in it. Instead, designers use the *x-height* (see figure 3.3) to approximate the appearing height of realistic texts. The zone between baseline and x-height contains most of the character shape information which makes the x-height prominent. The term *base height* is used in this work as a useful generalisation of x-height that can be applied across different scripts.

3.2. Cyrillic

Like Latin, Cyrillic is an instance of an alphabetic writing system. It is typically written from left to right with lines of text arranged from top to bottom. Each character (letter) represents a consonant or vowel. It is also bicameral.

Cyrillic is used by numerous Slavic and non-Slavic languages, e.g. Ukrainian, Russian, or Bashkir. As a representative subset of the script's characters, the contemporary *basic Cyrillic alphabet* is used in this work. It contains 33 uppercase and 33 lowercase letters (see figures 3.5 and 3.6).[4]

In the following text and figures, the Cyrillic characters are complemented by their names based on character descriptions from the Unicode standard. This naming is used consistently throughout this work instead of more standard transliterations using accented characters. It is convenient to avoid accented characters in the context of computation and data analysis tools as these might rely on only the basic Latin alphabet and punctuation in some contexts.

Similar to Latin, Cyrillic adds numerous characters to its basic alphabet. It uses the same numerals as Latin and most of the punctuation is identical or similar to Latin.

An important aspect of many scripts is the use of *regional character variants*. In various geographical and linguistic regions, the shapes representing the same character category can vary significantly. This aspect seems to be of less importance in Latin than in Cyrillic or Devanagari where major languages make use of regional character variants (see figure 3.7).

4. For the sake of simplicity, the 66-character basic Russian alphabet specified in the Cyrillic block of the Unicode Standard (2014) is used instead of the 96 characters suggested in the ISO/IEC 8859-5 (1999).

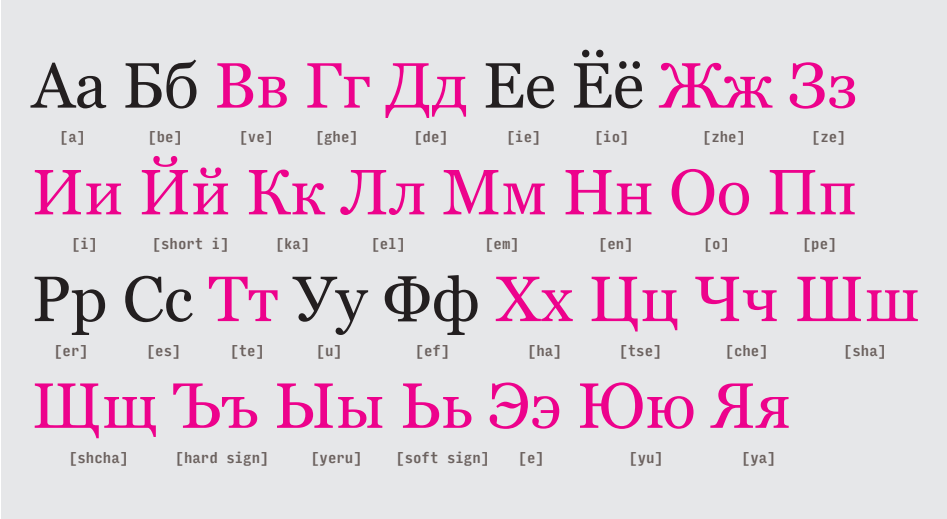


FIGURE 3.5: Uppercase and lowercase characters of the basic Cyrillic alphabet presented in the typeface Georgia. Character shapes that use the same spatial configuration of strokes and and stroke endings between uppercase and lowercase are marked in magenta. Note that this correspondence may differ to some extent, depending on a particular typeface's design.

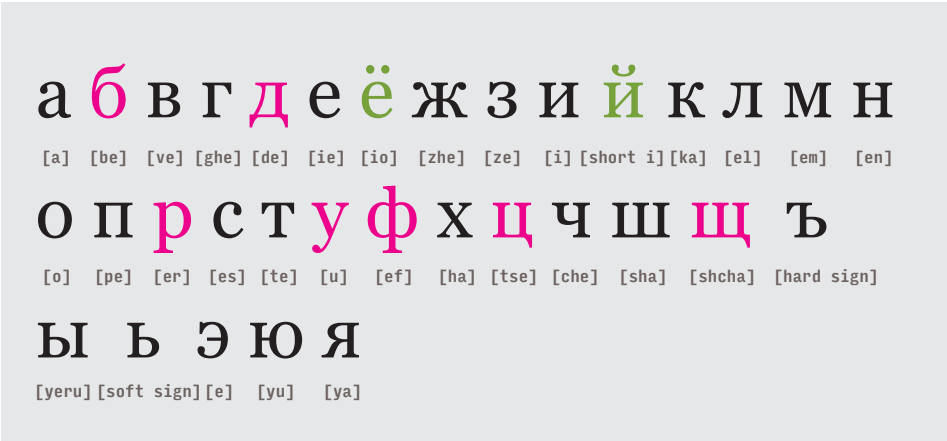


FIGURE 3.6: Cyrillic lowercase letters from the basic Cyrillic alphabet set in the typeface Georgia. Letters with ascenders or descenders are marked in magenta. Accented letters are marked in green.

FIGURE 3.7: The Russian Cyrillic (top) and the Bulgarian Cyrillic (bottom) make use of different regional character variants. These are only some of them. The example is set in the typeface Skolar.

While Cyrillic characters originated from Greek and Glagolitic, the script was drastically reformed by Peter the Great in the 18th century (Zhukov, 1996; Yefimov 2002). In the reform he seems to have followed the then-contemporary Latin typefaces, borrowing some of the Latin characters directly, and even inventing some of the letters himself following the Latin example. Therefore, many overlaps can be found between the design principles of the Latin and Cyrillic scripts (see figure 3.8).

Perhaps as a consequence of these reforms, conventions for some of the Cyrillic characters have been significantly influenced by Didone-like typefaces or what is also referred to as *modern* style in Latin (old-style being the Renaissance letters). When it comes to the stress of stroke modulation, Cyrillic often adheres to the vertical stress of modern-style typefaces, but can also use a more diagonal stress angle as in Latin (see figure 3.9).

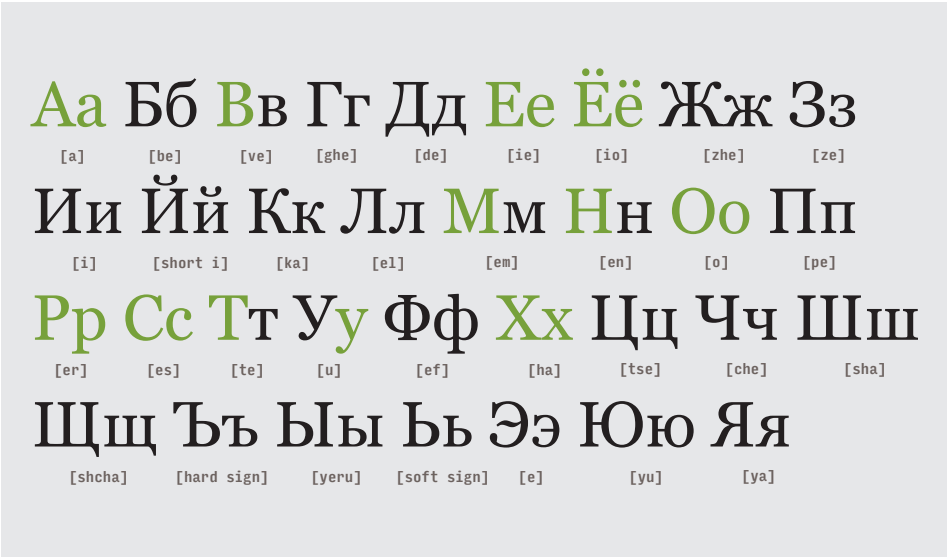


FIGURE 3.8: Typefaces for contintuous reading typically share some character shapes between Cyrillic and Latin (marked in green). These character shapes may correspond to different character categories, though. The example is set in the typeface Georgia.

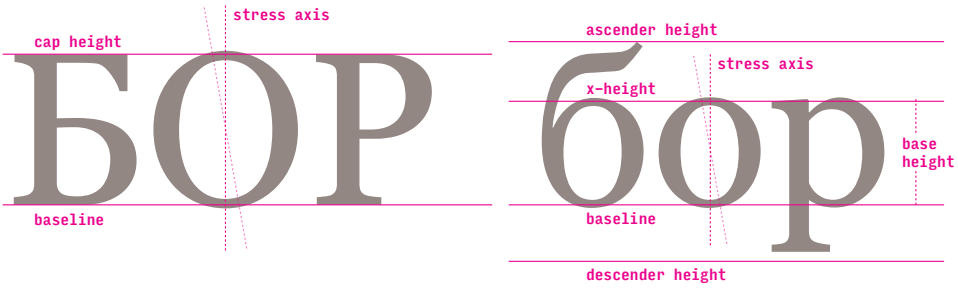


FIGURE 3.9: Cyrillic uppercase (left) and lowercase (right) can use the same guidelines as Latin. The stress axis is conventionally vertical or diagonal. The example is set in the typeface Georgia.

The early Cyrillic writing styles did not distinguish between different cases. When Cyrillic developed into a bicameral script – it is widely assumed this followed intervention by Peter the Great – the unicameral, majuscule letters remained to be used as uppercase and at the same time they were adapted to act as lowercase (Zhukov, 1996). This seems to be the reason behind the apparent lack of differentiation between the cases, apart from the relative size, that can be still observed in contemporary Cyrillic (compare figures 3.5 and 3.1). The uppercase letters are used in a similar fashion to the Latin script (see section 3.1 for details), which makes the lowercase comparably more frequent in general texts.

See figure 3.10 for an interpretation of Peter the Great's concept. The majority of the design principles present in his concept are still applied to typefaces for continuous reading today. Zhukov (1996) considers Peter the Great's typeface proposal as idiomatic for the Cyrillic script as the 15th-century Italian typefaces are for the Latin script.

FIGURE 3.10: Typeface by Galina Bannikova from 1950 based on Peter the Great's civil type (from: Zhukov, 1996, p. 13).

3.3. Devanagari

Unlike Latin and Cyrillic, Devanagari is an example of a syllabic writing system. Each character represents a syllable, i.e. a single vowel alone or together with one or more consonants. Similar to Latin and Cyrillic, it is typically written from left to right with lines of text arranged from top to bottom. It is unicameral, i.e. there is no distinction between lowercase and uppercase.

Devanagari is used for numerous Indian languages today, e.g. Hindi, Marathi, Nepali, and Sanskrit. The basic *Hindi syllabary* contains 11 independent vowel characters, 33 consonant characters, 10 dependent vowel characters (also called vowel signs or matras), and other signs (also called modifiers) (see figure 3.11).[5]

In order to make the explanation of the script behaviour more convenient, the Devanagari characters are labelled with their names or transliterations based on character descriptions from the Unicode standard (for reasons see section 3.2). The names of the syllables are capitalized.

Note that characters for the consonantal syllables represent a combination of a consonant sound, e.g. 'k', 'p', 'h', and an inherent vowel sound 'a'. To represent a consonant without the inherent vowel, a special sign *virama* (also called a *halant*) is positioned below a consonant character.

A crucial concept of the script is the ability to combine basic characters into more complex ones, so-called *conjuncts*, that represent syllables made of clusters of two or more consonants and a vowel, e.g. छ्र [ChRa].

The shapes of conjunct characters are mostly formed through a combination of elements from basic characters (see figure 3.12).

The inherent vowel of a syllable represented by a consonant or conjunct character changes as a result of a dependent vowel character being attached either before, after, below, or above the character, e.g. हि [Hi], ही [Hii], हू [Huu], है [Hai] (see also figure 3.13). These characters, orbital signs, are called "dependent" as they are not typically used on their own. They create meaningful combinations only with vowel, consonant, or conjunct characters.

5. This specific syllabary has been selected for the purpose of this work based on the Devanagari block of the Unicode Standard (2014) and Lambert (1953). Characters used for Hindi have been included. Characters that are used only for Marathi, Sanskrit, transliteration of English or Dravidian languages, and characters which are made of other characters with the addition of orbital signs, e.g. consonant ख [Khha], are not included.

FIGURE 3.11: Characters of the basic Hindi syllabary presented in the typeface Murty Hindi. Characters made as a combination of another character and an orbital sign are marked in green. Dependent vowel characters and orbital signs cannot stand on their own, hence they are shown in the figure as attached to the syllable क [Ka] or, alternatively, ट [Tta].

FIGURE 3.12: Different ways of forming conjuncts. On the left are the character shapes that correspond to the keyed-in sequence, on the right is the expected conjunct character shape. The conjuncts' sub-shapes can be combined either horizontally त्स्य [TSYa] or vertically ड्ड [DdDda]. Less often, the conjuncts have their own distinct design क्ष [KSsa]. There are several special cases that involve a consonant र [Ra]. When र [Ra] starts a conjunct, it changes form to become a crescent-like sign and is placed above the character र्क [RKa]. When it is at the end of a conjunct, it changes form and goes below छ्र [ChRa] or creates a distinct design with the preceding consonant character प्र [PRa]. The examples are set in the typeface Murty Hindi.

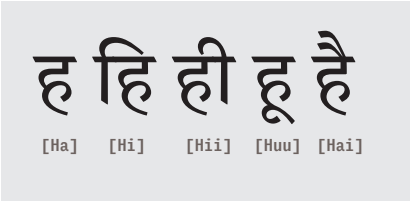


FIGURE 3.13: Dependent vowel characters change the inherent vowel of the syllable ह [Ha] to form syllables हि [Hi], ही [Hii], हू [Huu], है [Hai]. The example is set in the typeface Murty Hindi.

Different kinds of texts require different numbers of conjuncts. In modern languages, the conjuncts are relatively infrequent. Nonetheless, a professional contemporary font would include over 400 precomposed conjuncts and partial shapes used to form conjuncts automatically during typesetting (see figure 3.15).

Both Devanagari numerals and Latin numerals are used nowadays. The punctuation is related to the Latin punctuation, but it might be modified to work well in the context of Devanagari characters. There are several additional punctuation symbols as well.

There are regional character variants of basic consonant characters, conjuncts, and numerals required for some of the languages, e.g. there are differences between Hindi, Marathi, and Sanskrit. These however, constitute a small portion of the whole character set required for each of these languages.

The design principles of Devanagari are considerably different from both Latin and Cyrillic. The script "hangs down" from the so-called *headline* (in Hindi and Marathi called *shirorekha*) and occupies different vertical space (see figure 3.14). Conventionally, the script is based on broad-nib tools as it was originally written with a reed pen. The stress is diagonal, but it is typically opposite to that of Latin and Cyrillic, i.e. with the thin parts being in the top-right and bottom-left direction (Lambert & Summers, 1953).

FIGURE 3.14: Devanagari characters do not "sit" on the baseline, instead they are aligned as if "hanging" from the headline (shirorekha). The prominent metric, base height, is marked on the right side. The stress axis is conventionally diagonal, but it is tilted in the opposite direction from the Latin or Cyrillic stress axis. The example is set in the typeface Murty Hindi.

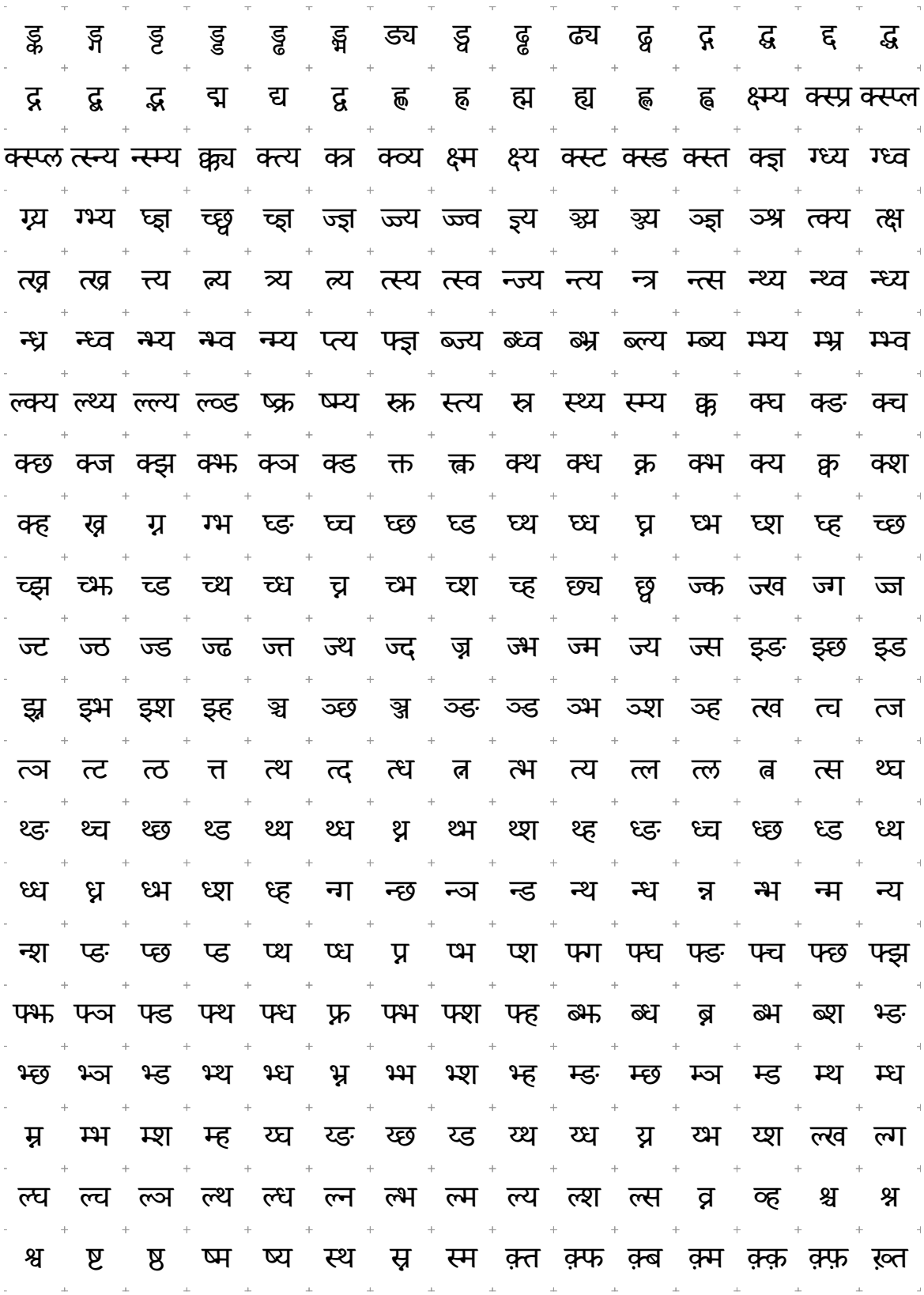


FIGURE 3.15: Some of the precomposed conjuncts included in the fonts for the typeface Skolar Devanagari.

There does not seem to be general agreement on a single, epitomic Devanagari typeface that would be equivalent to what Jenson's Roman is to Latin script. However, the late 18th century and the 19th century have seen several mature attempts at a typographic representation of the script. The following are a few examples. In 1795 Charles Wilkins created his first Devanagari typeface which although well executed, contained some idiosyncratic forms (see figures 3.16 and 4.18). Typefaces created in various mission presses in India in the first half of the 19th century showed qualitative improvements. And, in the second half of the 19th century, Ranoji Raoji Aru and Jaoji Dadaji created several exceptional typefaces for the Nirnaya Sagar type foundry and press (Naik, 1971) (see figure 3.17). Typefaces from this era definitely set the direction for Devanagari typefaces for continuous reading. Note that by way of comparison, this sets the establishment of conventional design principles for typographic Devanagari around 325–400 years after Jenson's Roman and over one hundred years after Peter the Great's Cyrillic type.

2. *Potential*	1. देदिवात्	देदिवातां	देदिवुः &c.
3. *Imperative*	1. देद्योतु or देदिवीतु	देद्योतां	देदिवतु &c.
4. *Firft Pret.*	1. अदेद्योत् or अदेदिवीतु	अदेद्युतां	अदेदिवुः &c.
5. *Sec. Prèt.*	1. देदिवां चकार	देदिवां चक्रतुः	देदिवां चक्रुः &c.
6. *Firft Fut.*	1. देदेविता	देदेवितारौ	देदेवितारः &c.
7. *Second Fut.*	1. देदेविथति	देदेविथतः	देदेविथन्ति &c.
8. *Precative*	1. देदिवात्	देदिवास्तां	देदिवासुः &c.
9 *Conditional*	1. अदेदेविथत्	अदेदेविथतां	अदेदेविथन् &c.
10. *Third Pret.*	1. अदेदेवीत्	अदेदेविष्टां	अदेदेविषुः &c.

So सिव् Serve, and others in व् .

FIGURE 3.16: An example of Devanagari typography of Charles Wilkins's Sanskrit grammar published in 1808 (from: Naik, 1971, p. 264). The types used had been revised from their original version made in 1795.

आरूंचा हस्तस्पर्श होतांच त्यांना सौंदर्य प्राप्त होत असे. जणूं काय त्यांच्या हाताचीं बोटें जादूनें भारलेलीं होतीं. एकादा अभिजात प्रतिभासंपन्न शिल्पकार दगडाला किंवा मातीला स्पर्श करतांच तिला हां हां म्हणतां जीवन - कला आणतो कि प्रेक्षकांना थक्क करून सोडतो; त्याचप्रमाणें मुद्रिकांसाठीं लागणाऱ्या धातूच्या तुकड्याला आरूंचा हस्तस्पर्श होतांच त्यांतून नयनरम्य, बोलक्या

FIGURE 3.17: An example of a Nirnaya Sagar Devanagari typeface Balbodh (from: Naik, 1971, p. 313).

3.4. Representative characters

In order for studies with participants to be feasible, it is necessary to reduce the number of characters included in the studies. The rationale for this is described in detail in section 5.4; however, it is worth mentioning here the criteria for including or not including particular characters in relation to the design principles of the corresponding scripts.

The practical goal is to select around 20 to 40 characters that are most representative of the appearance of the script as used in text. Naturally, a script can be used by multiple languages in various different ways. In this regard, the focus is on English, Russian, and Hindi, as described above. Therefore, the Latin and Cyrillic basic alphabets or Devanagari syllabary respectively suggest themselves as ideal candidates for the character sets for the studies. Unfortunately, these are still too large and need to be reduced.

An approach based solely on the frequency of characters in text in each of these languages has been avoided. A script is a system of inter-related character shapes. But the character shape relationships are not based on or derived from how often they are used in text. For example, the frequency of the letter 'z' is low in English (Mayzner & Tresselt, 1965; Norvig, 2013), but it is an important part of the system nonetheless and should be included. By selecting the most frequent characters one might end up creating a set that is unrepresentative of the script as a system of shapes. Therefore, the intention is to include the conventional character sets such as English lowercase or Hindi consonantal syllables as complete as possible to maintain the whole spectrum of shape relationships within those sets.

This is also the reason why the total numbers of characters studied differ for each script. An arbitrary limit would mean that, for example, some characters from the basic set of Hindi consonantal syllables would not be included.

Accented characters are not included for Latin. Only two accented characters are included for Cyrillic ('ё', 'й' [io, short i]) and one character distinguished by an orbital sign is included in Devanagari('ङ' [Nya]), as these are essential to the basic Russian alphabet and the Hindi syllabary, respectively. In Devanagari, for the most part, character combinations with dependent vowels and orbital signs, e.g. 'है' [Hai], are not included. This is an acceptable compromise given that dependent vowels, orbital signs, and accents are, in fact, independent shapes that attach to many different characters in various combinations systematically. It seems reasonable to assume that, for example, ही [Hii] is as similar to डी [Ddii] as ह [Ha] is to ड [Dda], or á is as similar to é as a is to e. Results of similarity studies with accented characters that were conducted by Ian C. Simpson et al. (2013) suggest this is the case. A study focused on similarity among orbital signs or accents could be conducted additionally in the future to complement the results of this research.

Punctuation and figures have less influence on the appearance of the text in each script, hence they are not included for any of the scripts. As mentioned

above, in Latin and Cyrillic, the uppercase is generally less important for the overall appearance. Therefore, only the lowercase characters from the basic alphabets for these scripts are used in the studies (see figure 3.18).

In Devanagari, the conjuncts are not included as they are comparatively less frequent in contemporary texts. The dependent vowels and orbital signs are not included on their own either. Only simpler, independent vowels are included. The characters studied for Devanagari therefore include only the consonants and selected vowel characters from the basic Devanagari syllabary (see figure 3.18).

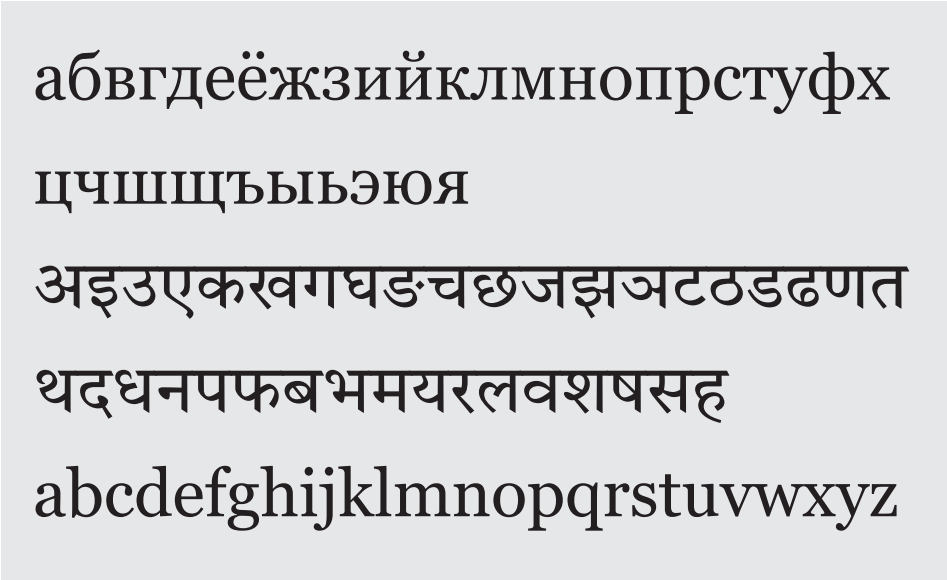


FIGURE 3.18: Representative characters selected for the studies with participants. Cyrillic (top) has 33 characters, Devanagari (middle) has 37 characters, and Latin script (bottom) has 26 characters. The examples are set in the typefaces Georgia (Cyrillic and Latin) and Murty Hindi (Devanagari).

4. Approaches to character conceptualization

The following is a discussion of the main approaches to making and thinking about shapes in typeface design. The goal is to capture the gist of expert knowledge that could help create the theoretical model described in chapter 6: *Modelling character similarity and coherence* and used to predict and explain readers' responses when judging similarity in conventional typefaces for continuous reading.

Individual techniques used to create character shapes as they emerged during history have been detailed in various publications (for example Harvey, 1996; Smeijers, 1996; Southall, 2005) and include, among others: handwriting, calligraphy, scribbling, carving, cutting, sculpting, bending wires or neon-light tubes, repurposing other objects, stamping, drawing, digital drawing using mathematically defined contours, stencilling, signpainting, graffitti, or algorithmic generation of shapes. Some of these techniques are used to produce character shapes directly, while others create character concepts that are used in combination with reproduction technologies, as in the case of various kinds of printing, stencilling, light projections, or digital displays. Sometimes, there is even a series of techniques and reproduction technologies that leads to the final character shapes. For example, letter punches are cut and used to produce metal matrices, which in turn are used to cast metal type, which is then used for printing (Smeijers, 1996). Producing character shapes in the digital environment seems significantly simpler in this respect. Designers typically master digital fonts on computers using specialized software while visualizing them on screens. The resulting fonts can be used in other software to display texts on screens, or to prepare them for other kinds of reproduction.

When faced with an already-made character shape, and bearing in mind the techniques enumerated above, one can conceptualize its shape by assuming the ways in which it was produced. Note that depending on the assumptions made, a single shape can thus be conceptualized in multiple different ways (see figure 4.1). Thus, it is also possible to notionally conceptualize characters in terms of one technique while using a different technique for their production. For example, a designer developing a font with characters described as contours can still think about the characters in terms of strokes. The ways in which shapes are notionally conceptualized depends on conventions (see chapter 3) and convenience. The ways they are conceptualized in production naturally depends on the font-development tools used. Some of these are described below.

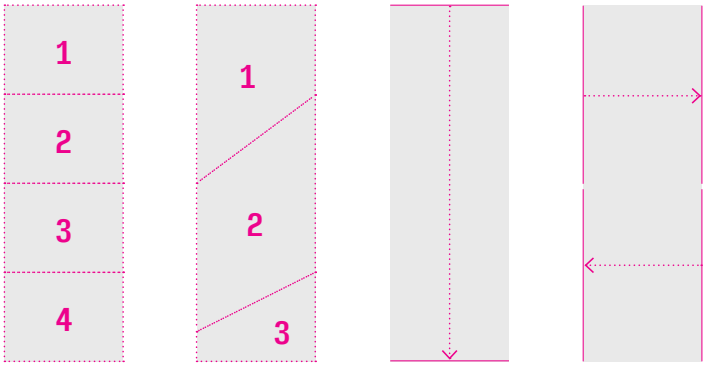


FIGURE 4.1: A single shape (the grey rectangle) can be conceptualized in different ways, for example (from left to right): as composed of four identical sub-shapes, as composed of three different sub-shapes, as a single vertical stroke made with a thin flat pen, as two smaller horizontal strokes.

Five general approaches to shape conceptualization are discussed in the following sections. This list is not meant to cover the whole spectrum of approaches one could possibly use.[1] It only covers the common ones that are likely to be used in the design of conventional typefaces today. These approaches can then be used to inform the model described in chapter 6.

4.1. Shapes as contours

Character shapes can be conceptualized using their contours. The shapes are then created by filling the area within these contours when they are rendered on screen or printed (see figure 4.2). Approaches that use mathematically-defined curves to describe the contours have been used in most contemporary font formats. Today, various kinds of bézier curves are used most of the time (Kamermans, 2011; Haralambous, 2007). Bézier curves are used in OpenType, a format developed jointly by Microsoft and Adobe (Microsoft, 2018) as an extension to the TrueType format that was originally developed by Apple (Apple, n. d.). They are also used in the older Type 1 format developed by Adobe (1993) and in the interchange font format UFO (Leming, van Rossum & van Blokland, 2004). The IKARUS format takes a slightly different, more general, route; it describes the contours as a series of points without any statement regarding the mathematical functions that will connect those points to form a continuous curve (Karow, 1994). Contour-based approaches are also supported in most of the contemporary font-production tools such as FontLab VI (FontLab, 2017), Glyphs (Seifert & Scheichelbauer, 2016), and RoboFont (Berlaen, 2018).

1. One could hypothesize that original designs often present new ways of thinking about shapes that challenge this list.

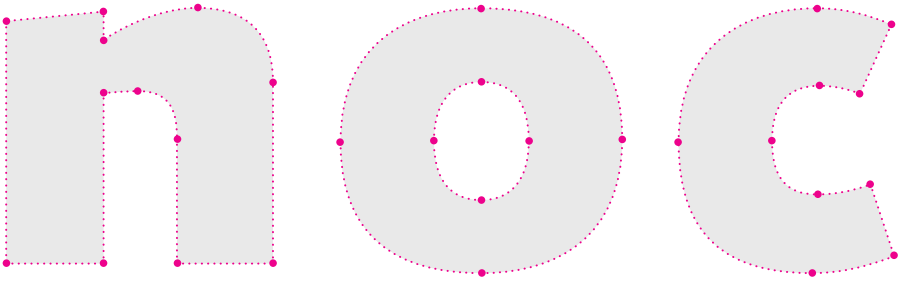


FIGURE 4.2: Character shapes for 'n, o, c' (in grey) conceptualized as contours using bézier curves (the dotted line). The on-curve control points are shown as the larger magenta dots. The off-curve control points are not shown.

Working with contours has become routine to contemporary typeface and graphic designers. The approach achieves precision and a large degree of resolution-independence, and provides considerable flexibility by allowing a great variety of shapes to be described. It is worth noting that describing characters with closed counters (see figure 4.3) requires more than one contour and systematic approach to filling, which introduces further complexity. The reproduction of contour-based concepts on screens or in print often requires their rasterization, i.e. their continuous contours must be adapted to a discrete grid made of pixels.

Representing shapes through contours instead of by body seems to go against intuition. For example, the letter 'o' is usually made of two contours, one contour describing the outer edge and another one describing the inner edge (see figure 4.2). This goes against the intuition that o is a single ring-like shape. On the other hand, the ability to control shapes of counters independently of the outer contours and the ability to shape various visual compensations directly are important advantages of this approach.

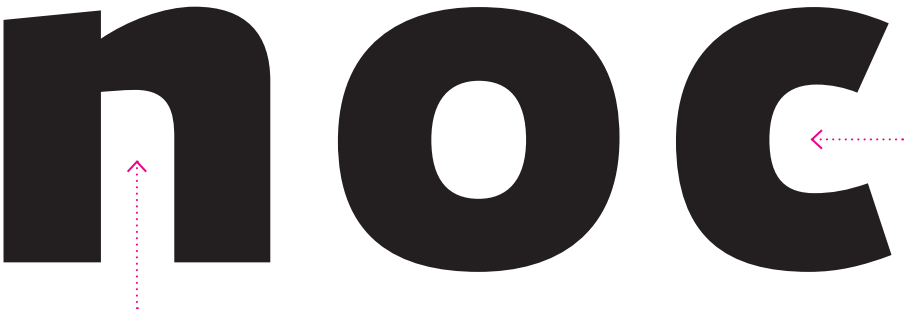


FIGURE 4.3: *Counters* are the non-printed spaces within character shapes. A counter is called *closed* when it is not connected to the outside area of the character shape as in the letter 'o' in this illustration. It is called *open* when it is connected to the outside area as in the letters 'n, c' in the illustration.

4.2. Shapes as strokes

The stroke-based approach attempts to approximate calligraphic and sign-painting practice, and handwriting. These techniques are collectively referred to as *chirographic* in this chapter to avoid the decorative and casual connotations related to calligraphy and handwriting, respectively, while adhering to the notion that a writing hand could have been involved. The stroke-based approach represents characters as sequences of strokes. This sequence can be continuous or interrupted. Edward Johnston (1958) illustrates this, and so do many other calligraphers in their writing manuals (see figure 4.4). Importantly, the approach assumes that there has been a movement across the writing plane involved in the making of the resulting shape. According to Gerrit Noordzij (1991, 2005), a stroke is an uninterrupted trace of some marking tool on the writing plane.

FIGURE 4.4: Writing illustrations provided by Johnston (1958, pp. 40–41, 90% scale).

A simplified geometric conceptualization of a stroke (see figure 4.5) is used further in this research. It defines a stroke using:

i. a marking tool, called a *pen*, described by its shape,
ii. a notional line the pen was centered on while making the stroke shape, called a *midline*,[2]
iii. a *pen behaviour* which also includes the manner in which the pen was used.

The shape of a pen at a particular point along the midline as affected by the pen behaviour is called *pen impression*. Thus, a stroke can be also implemented as an interpolation between different pen impressions along the midline.

Variations of such a geometric description have been used in many contemporary vector-based drawing programs like Adobe Illustrator (Adobe, 2018), Inkscape (Inkscape, 2018), and CorelDRAW (Corel, 2018), where the pen shape is represented by a contour or a group of contours and the midline is represented as another contour.

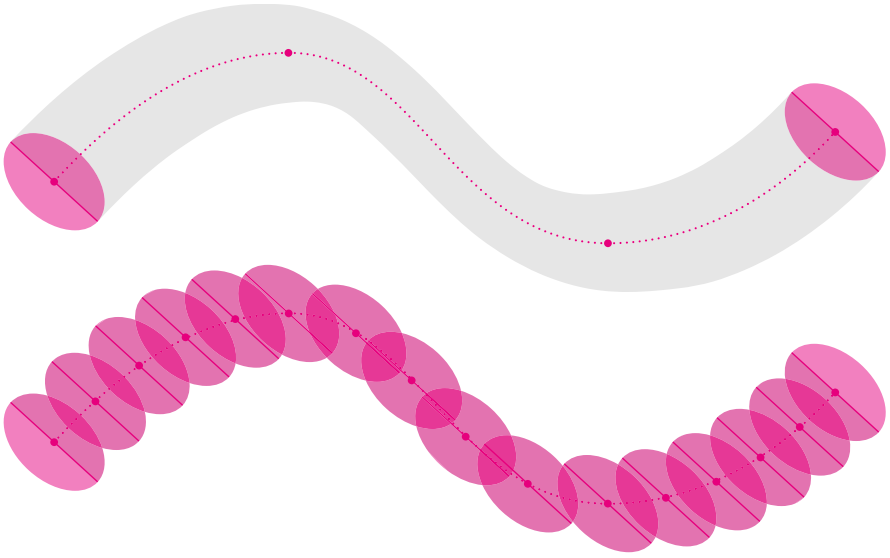


FIGURE 4.5: A simplified stroke shape can be specified by describing the pen shape and midline it follows (the dotted line). In these examples, the pen behaviour corresponds to the translation sort of contrast (see below), thus the pen shape corresponds to the pen impressions (magenta ovals). The top example shows the pen shapes at the beginning and at the end of the stroke, midline, and the final stroke shape (in grey). The bottom example shows how the final shape results from the pen impressions made along the midline.

2. Noordzij (2005, p. 27) uses the term *heartline* and defines it as "described by the midpoint of the advancing counterpoint". A counterpoint is a pair of points on a stroke's contour created at the same time with the writing tool (ibid, p. 20). A set of stroke midlines for the whole character shape is often called a *character skeleton* (Majoor, 2004).

In order to discuss the pen behaviour more clearly, one can use Noordzij's (2005, p. 26) definition of *contrast* (see figure 4.6) as the ratio between thin and thick parts of the stroke and three *sorts of contrast* (see figure 4.7) which can be combined.

In the *translation* sort of contrast, the contrast of the stroke is the result of changes in the direction of the stroke alone as the orientation and size of the pen shape remain constant. When using the translation sort of contrast, the angle of the pen shape influences the placement of the thin and thick parts (also called a *stress*) directly. Oblique pen shape produces contrast with a diagonal stress, upright pen shape produces strokes with a vertical or horizontal contrast.

In the *rotation* sort of contrast,the contrast is the result of changes in the direction of the stroke, but also of the changes in the orientation of the pen shape while its size remains constant.

In the *expansion* sort of contrast,: the contrast is the result of changes in the direction of the stroke, but also of the changes in size of the pen shape while its orientation remains constant. The changing size of the pen is often used to replicate the effect of a changing pressure when writing with a flexible real pen.

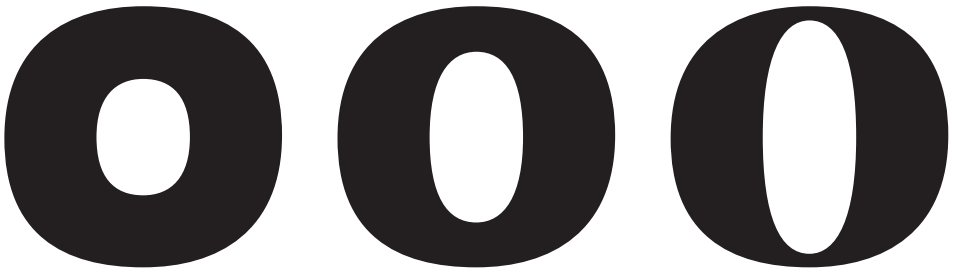


FIGURE 4.6: Left to right: letter 'o' with increasing level of contrast.

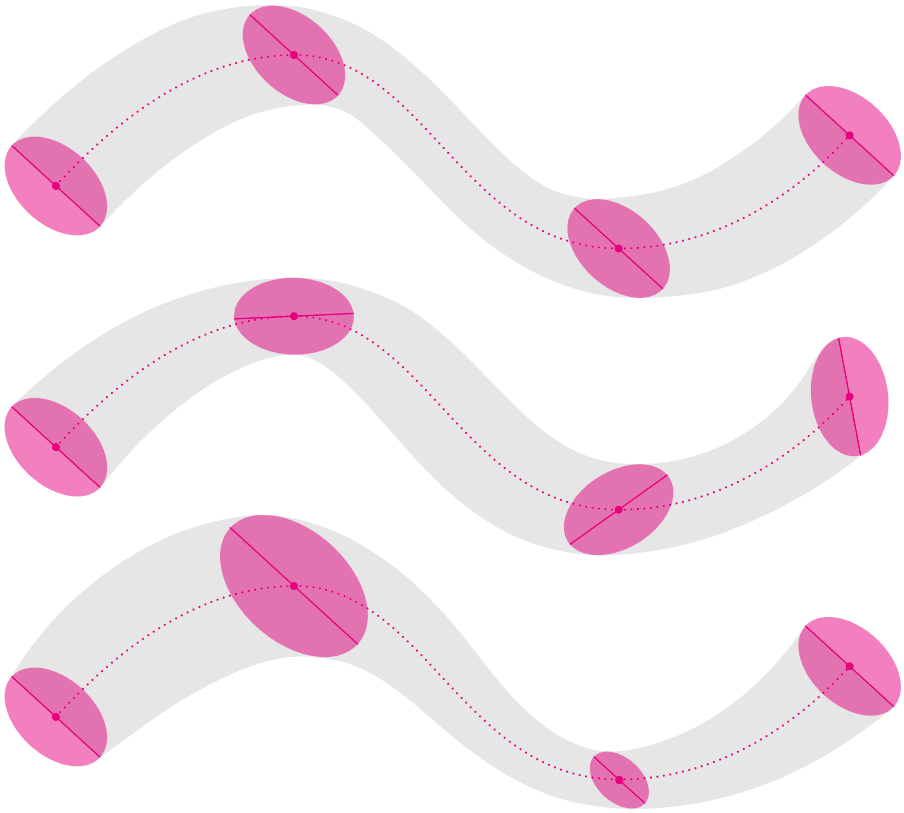


FIGURE 4.7: Noordzij's sorts of contrast (top to bottom): translation, rotation, and expansion. Different sorts of contrast affect the same pen shape in different ways and produce various pen impressions (the four magenta ovals in each of the strokes above).

Contemporary font editors and their extensions often implement some kind of stroke-based approach, but only as a means of producing contours. Probably the most well-known implementation of the stroke-based approach is Metafont by Donald Knuth (1982, 1985, 1999), which fully embraces the approach and uses strokes to conceptualize characters within the final fonts.

Knuth came up with the concept of a meta-font, a generic tool that would allow the description of fonts. His Metafont[3] is one example of such a tool. Metafont is a programming language that uses strokes to conceptualize characters (see figure 4.8) in a way that allows effective generation of additional styles and effective rasterization in various sizes on screen. Originally, it supported only strokes. It was only later that support for contours was added to allow for a greater variety of shapes. Today, it can be considered a hybrid system with support for both stroke-based and

3. The capitalization and hyphenation is important here: it distinguishes the concept (meta-font) and the actual implementation (Metafont). The latter is usually represented as an all-caps logotype.

contour-based approaches. It is also worth noting that Metafont has been used to create fonts for the Arabic script (Sherif & Fahmy, 2008) as well as for various Indian scripts (Mohanty, 1998; Pandey, 1998). Among other scripts, the stroke-based approach is also commonly used to conceptualize Chinese characters (Hu & Fischer, 2004).

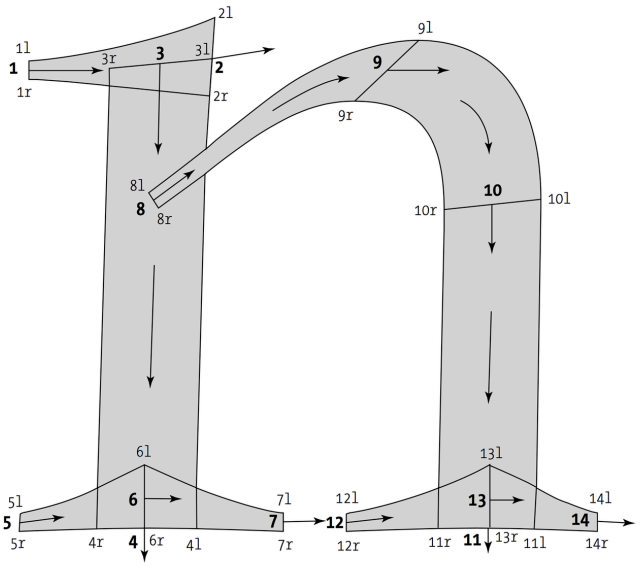


FIGURE 4.8: A letter in Metafont is built using strokes (from: Haralambous, 2007, p. 919).

The stroke-based approach provides a convenient way to implement chirographic conventions in one's design. As mentioned in chapter 3. all of the three scripts studied derive some of their basic principles from traditions that are chirographic.

However, shapes are not always easily conceptualized as simplified strokes. When their precise description involves too many changing pen impressions, it is often easier to use a contour-based approach (see figure 4.9). Keeping a contour-based approach for the precise description, one can still use a stroke-based approach notionally, for example to decide how to distribute the thin and thick parts according to a relevant convention.

Using ideas such as a pen shape, midline, or sort of contrast allows one to categorise a variety of different stroke shapes to represent their similarity. It is possible to talk about all stroke shapes made by a particular pen shape, all stroke shapes with a particular kind of midline, or stroke shapes with a particular sort of contrast.

The main doubt raised within the context of this research is, that although the reading and writing are often taught together in schools, it is unclear whether readers actually perceive characters as if they are written as some evidence against this was given by Sue Walker and Linda Reynolds (2002).

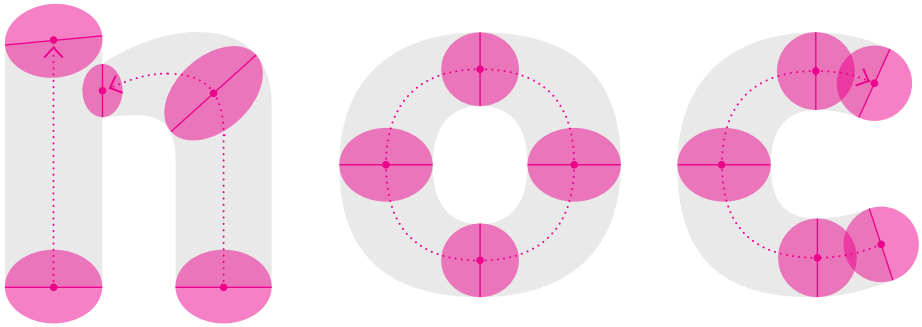


FIGURE 4.9: Stroke-based conceptualizations of the character shapes for 'n, o, c' (in grey). While 'o, c' can be represented with a single stroke, conceptualizing 'n' in this case requires two strokes. Note that concepts for 'n, c' make use of several different pen impressions as the orientation and scale of the pen shape changes. The pen impressions are represented by the magenta ovals; the midlines are represented with the dotted lines. The strokes are trimmed flat at their ends in order to obtain the resulting character shapes in grey.

4.3. Transformation of shapes

Shapes can be seen as results of transformations applied to other shapes. These transformations could include basic affine operations such as translation, rotation, reflection, or scaling (see figure 4.10), but also other more complex transformations such as outlining or casting a shadow (see figure 4.11). Affine transformations, i.e. transformations which preserve parallel relationships, are supported by most font editors and are commonly used. Notice how a sub-shape transformed using an affine tranformation is still easily identified (see figure 4.12). This is an important consideration in the model described in chapter 6.

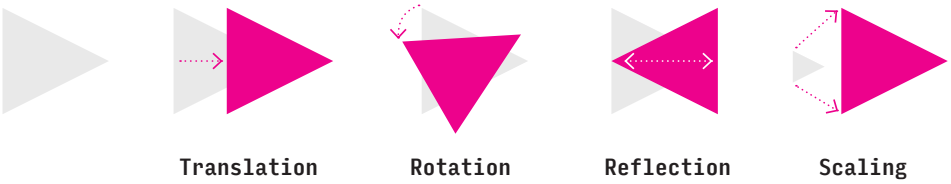


FIGURE 4.10: Affine transformations: translation, rotation, reflection, and scale. The grey shape is being transformed into the magenta shape.

Table Table Table

FIGURE 4.11: A few examples of more complex tranformations that can be applied to character shapes (from: Karow, 1994, p. 83).

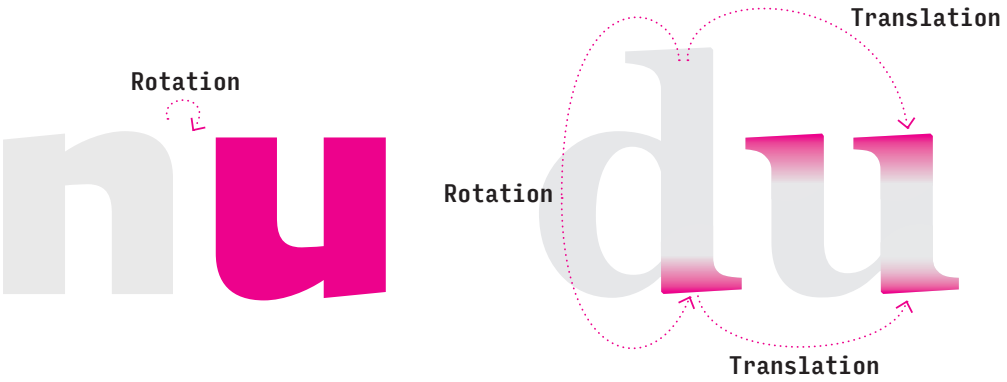


FIGURE 4.12: Using affine transformations to produce whole character shapes (left) and to produce character sub-shapes (right). The arrows indicate the shape transformations. Note that in practice, the transformed sub-shapes would be usually adapted to a new visual context by the designer.

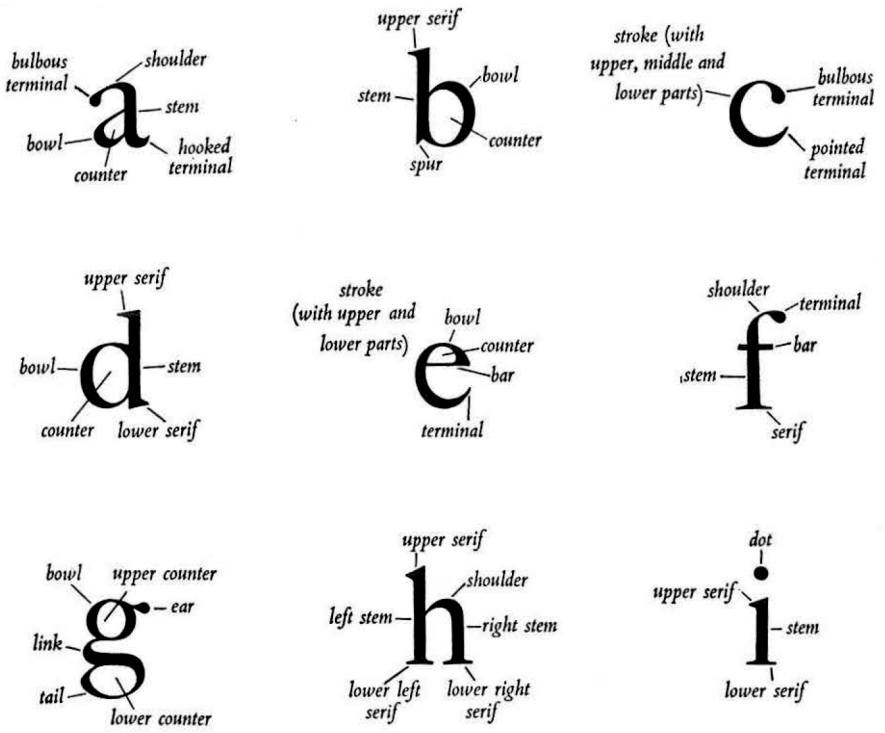


FIGURE 4.13: English nomenclature for Latin characters (detail from: Gaskell, 1976, p. 48).

4.4. Composition of shapes

Character shapes are often viewed as being made up of one or more component shapes. For example, Fernand Baudin (1989, p. 30) states that the graphical order "is determined by the similarities between characters made up of mainly straight lines, circles or obliques". When discussing style, Schapiro (1994, p. 54) makes the following point: "... in general the description of a style refers to three aspects of art: form elements or motifs, form relationships, and qualities."

Further evidence supporting the decomposition approach can be found in various type anatomy charts and nomenclatures (Menhart, 1957; Gaskell, 1976; McLean, 1992; Baines & Haslam, 2005; Pecina & Březina, 2008). Four of them are shown in figures 4.13—4.16. In these charts, similar – but not necessarily geometrically identical – sub-shapes are given the same name. However, one has to keep in mind that some of these charts were not compiled by designers. Thus, rather than merely being evidence of production processes, they are evidence of perception.

FIGURE 4.14: Czech nomenclature for Latin characters (detail from: Menhart, 1957, p. 141).

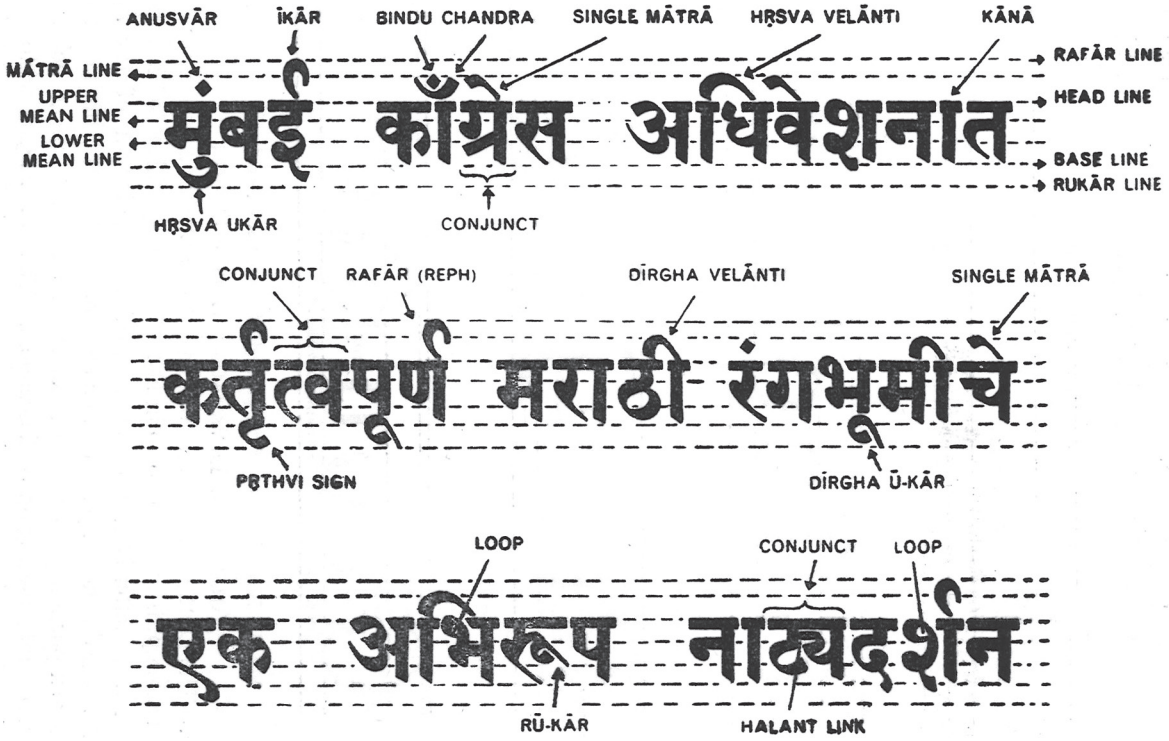


FIGURE 4.15: Basic anatomy of Devanagari typefaces, its guidelines, and nomenclature for categories of characters (from: Naik, 1971, p. 214).

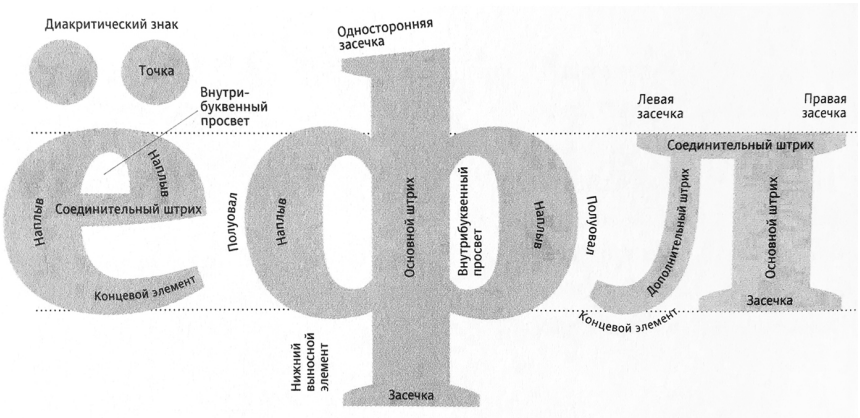


FIGURE 4.16: Nomenclature for Cyrillic typefaces (detail from: Gordon, 2013, p. 52).

Building up character shapes from multiple components has been an important aspect of the design process of some typeface designers (see figures 4.17 and 4.18). Also, it has been hypothesized that Johannes Gutenberg made his type using component punches (Agüera y Arcas, 2003). Contemporary designers often draw characters as multiple overlapping contours instead of one merged contour in font editors (see figure 4.19). This is done in order to simplify sharing of these components across multiple glyphs and different weight variants and to speed up drawing and editing. Note that these components can be modified after copying to adapt them to the new visual context. It also makes the contour-based conceptualization of shapes (see section 4.1) closer to the stroke-based approach (see section 4.2). Note that when characters are represented as merged contours (for example see the letter 'n' on the left in figure 4.19) any information regarding their shared components gets lost.[4]

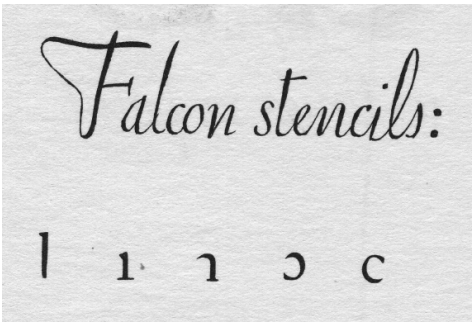


FIGURE 4.17: In his letter to typeface designer Rudolf Růžička, William Addison Dwiggins (1940, p. 5) explains how he used stencils of character components to draft his typeface Falcon.

4. The notion of a component in this research is different from that used in many font editors today, where a component is a special entity that cannot be modified freely (only using some basic affine transformations). Such components can be preserved in published TrueType fonts, for example.

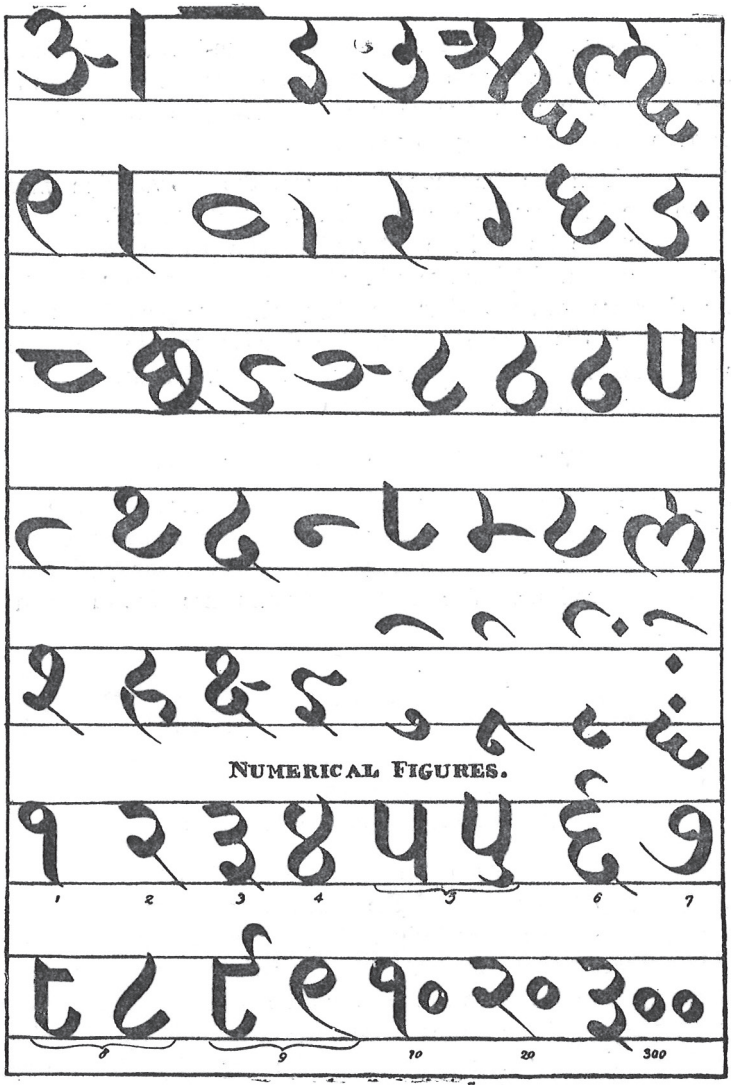


FIGURE 4.18: Charles Wilkins' chirographic analysis of Devanagari (from: Naik, 1971, p. 216). Note that the diagram shows some idiosyncratic forms specific to the author.

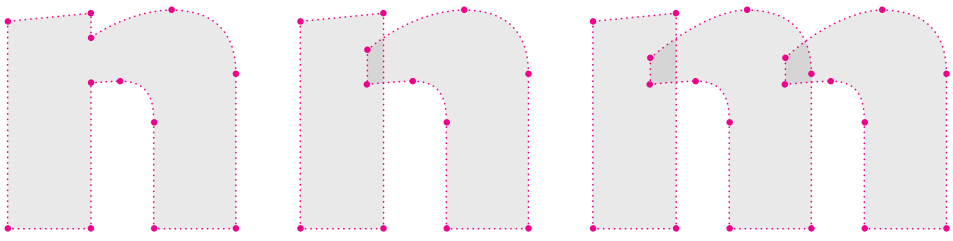


FIGURE 4.19: Merged ('n' on the left) and non-merged ('n, m' on the right) contours. The 'm' shows a reuse of contours from 'n'. Note that the reused contours were slightly modified and differently spaced after copying.

Depending on what can be done with character components, one can distinguish two kinds of approaches. In the *module-based* approach, the components (called modules) remain constant in size and shape, geometrically. The approach is commonly used to create so-called modular typefaces (see figure 4.20), where each character is made up of modules. The set of modules is limited and shared across different character concepts from the same typeface. This way a great degree of coherence is achieved. In the *element-based* approach, on the other hand, the components (this time called elements) can be modified. They can be rotated, stretched, or scaled, for example. Each kind of element can be modified in different ways to adapt it to the visual context of a particular character. In this sense, the element-based approach is more flexible than the module-based approach.

FIGURE 4.20: Module-based approach as implemented by the tool FontStruct (Meek, 2010).

There is no assumption being made about the way the components are conceptualized. They could be conceptualized as strokes or contours, further transformed or decomposed. Typically, components that also appear in simple handwriting models of corresponding characters are interpreted as strokes (see figure 4.21) while other sub-shapes can be interpreted using contours.

Bitmap fonts deserve a sidenote in this context. In the era of phototypesetting and in the early days of digital typography, characters started to be represented as bitmaps in order to control their appearance on low-resolution screens precisely (see figure 4.22). Bitmap is a matrix of bits which prescribes which screen pixels should be turned on or off. Thus, bitmap fonts can be seen as module based, with individual modules corresponding to screen

pixels. However, bitmap conceptualizations are not considered in more detail here since the shape decomposition is an adaptation to limits of a particular type of technology rather than a means of creating coherent typefaces.

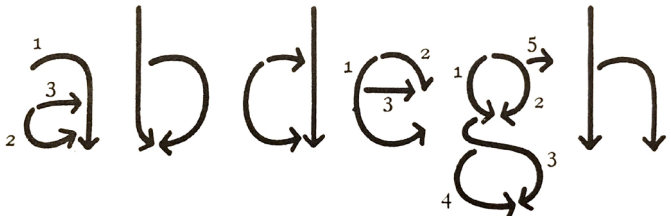


FIGURE 4.21: A simple handwriting model (Korger, 1992, p. 52).

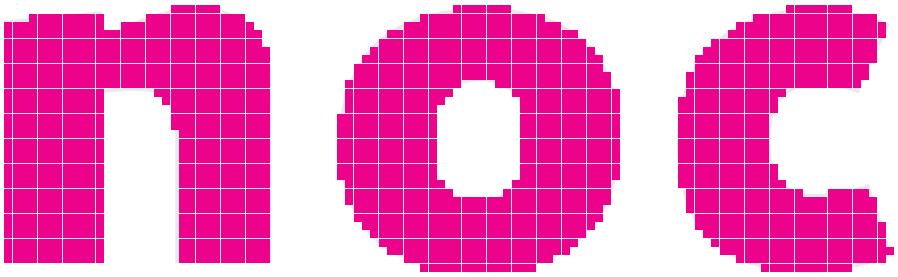


FIGURE 4.22: Bitmap conceptualizations of the letters 'n, o, c'. The squares in this illustration correspond to screen pixels a bitmap font.

An example of an element-based concept is the experimental tool *abcdefg* [ab kuh def gee] that was being researched and developed at Xerox PARC in the late 1980s (see figure 4.23). The tool was based on Debra Adams's Master's thesis (Adams, 1989). Initially, the tool was intended to streamline prototyping of new fonts. However, the long-term aspiration was to create a complete font-development tool that would speed up typeface designers' work. The idea was that the designer would first design several control characters which would then be decomposed into individual components. These components would be used to derive preliminary designs for the remaining characters automatically. It was expected that derived shapes would be retouched by a designer.

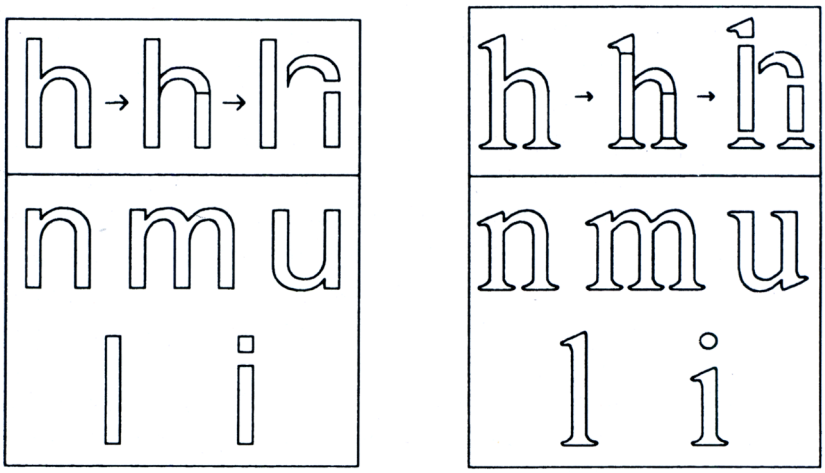


FIGURE 4.23: Characters generated by the *abcdefg* tool are derived by applying transformations to components of the 'h' (from: Adams, 1989, p. 65).

In order for the derivation process to work, there had to be a system describing the decomposition of control characters to elements, as well as the composition of the remaining characters from the elements. The system included a list of elements needed for a particular character, but also their spatial relationships, element transformations, and manner in which they would join. The elements themselves were represented as contours, i.e. bézier outlines (see section 4.1). There are four control characters considered in Adams's dissertation: 'o, h, p v', but those would certainly not contain all of the design information needed to construct the remaining characters of the Latin alphabet. Despite its many qualities, including an attempt to adhere to designers' creative processes, *abcdefg* remained an experiment.

Between the years 1998 and 2001, Changyuan Hu (1998) and Roger D. Hersch (Hu & Hersch, 2001) developed another system based on elements called *CPFPage*. Later on, the system was revisited by Tamir Hassan (Hassan, Hu & Hersch, 2010) and tested by faithfully reconstructing the typeface Frutiger and exploring new styles generated by interpolating and extrapolating control parameters of the system.

In CPFPage, characters are built up from elements, such as sweeps, stems, and serifs. While in *abcdefg* the elements were represented as contours, CPFPage system is hybrid as there are several kinds of elements conceptualized in different ways (see figure 4.24). Namely, sweeps are akin to simplified strokes as they were defined in section 4.2 (see figure 4.25). All the elements can be combined to form characters, which allows a great flexibility. Furthermore, various parameters are used in order to keep track of relationships between characters. Unfortunately, this system has not been developed further; from the available documentation it promises to be a very effective font development environment.

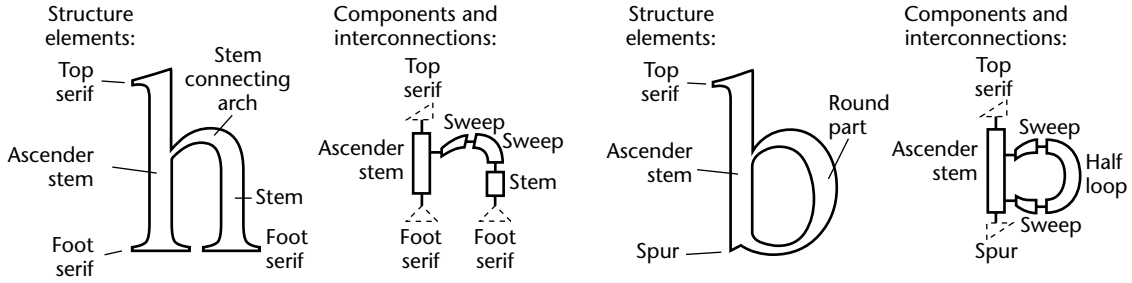


FIGURE 4.24: Composition of characters in the CPFPage system. CPFPage uses an idea of a sweep that is close to the stroke as it was defined in section 4.2 (from: Hu and Hersch 2001, p. 71).

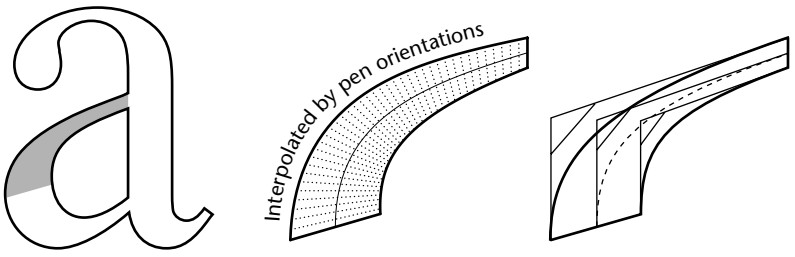


FIGURE 4.25: The implementation of the stroke-based approach in CPFPage, called a sweep, interpolates simplified pen impressions along the midline (from: Hu and Hersch 2001, p. 75).

Hassan et al. (2010) mention one aspect of the system that deserves emphasis: the underlying knowledge about character constructions which is needed for various genres of typefaces. This is not unlike the system describing characters in *abcdefg* or the notion of script conventions in this research.

Uwe Schneider's system DaType (Schneider, 1998) is another experimental tool building character shapes from elements. It shows a hierarchical composition from strokes. Later, Schneider (2000) presented a hybrid approach that combined the stroke-based and contour-based approaches in one.

Conceptualizing characters using components promises to approximate the designers' creative processes well. However, in all the systems mentioned in this section, there is no guarantee that the decomposition into shapes maintains some important holistic qualities. These qualities could include relationships between individual components or overall qualities of the characters being decomposed. For example, the quality of a square being a closed shape is lost when it is represented as four individual lines. The systems thus rely on their users to make sure the resulting character shapes look as they should according to a script's conventions. This is discussed in greater detail in sections 6.3.2 and 6.3.3.

4.5. Generalization of shapes

4.5.1. Typographic attributes

Instead of focusing on detailed character descriptions, one can focus on more general visual qualities shared across multiple character shapes. *Typographic attributes*, as they are called here, are a way to describe coherence of character shapes from a particular typeface. Commonly, attributes worth considering are expressed in most of the character shapes studied. Nonetheless, sometimes, when it is exceptionally distinctive or when it helps to identify a particular typeface, it may be an attribute that is relevant to only a handful of character shapes, e.g. the treatment of joining diagonals in the letters 'K, k' (see figure 4.26). Also note that, as shown in figure 2.9, some attributes that are often assumed to be global, may not be present in all character shapes.

FIGURE 4.26: Different connections of the diagonals to the vertical stem in the letters 'K, k' in typefaces (left to right): Minion, Georgia, PT Serif, Palatino.

The use of typographic attributes is a very common approach in typeface classifications. Some comprehensive examples for Latin are the PANOSE System developed by Benjamin Bauermeister (1988) or the classification of typefaces by Jan Solpera (Solpera, 2009) (see figure 4.27).

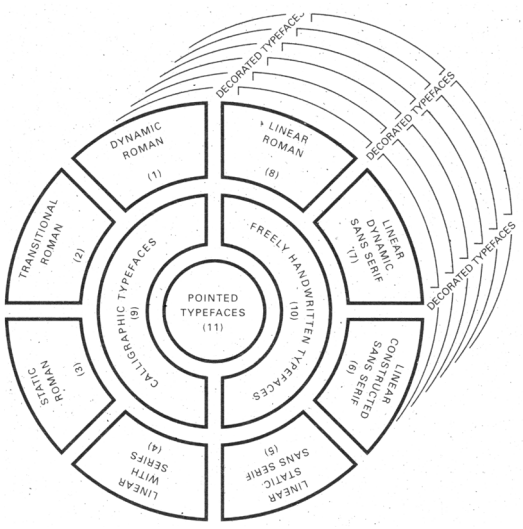


FIGURE 4.27: Jan Solpera (2009, p. 3) attempted a systematic and comprehensive classification of typefaces for Latin.

Common typographic attributes have been concisely summarised by Baudin (1989); a list of the attributes he considers follows (compare these with basic guidelines for the scripts studied illustrated in figures 3.3, 3.9, and 3.14):

- weight/colour,
- treatment of serifs,
- width/horizontal proportion and rhythm,
- contrast between thick and thin parts,
- size and closure of counters,
- number of strokes (complexity),
- guidelines/vertical proportions,
- relative proportions of characters to each other.

Typographic attributes help in relating typeface design to the requirements of document design, e.g. typefaces with a very high contrast are generally not recommended for continuous reading. But at the same time, this approach remains limited as it provides only a vague description of the individual character shapes. It provides a direction rather than a description.

Furthermore, a focus on typographic attributes is problematic in terms of flexibility as the attributes need to be defined before typefaces are created. It is impossible to preempt all typographic attributes for all typefaces in the world (Hofstadter, 1982). Pursuing this exercise is, however, useful for educational purposes. Typographic attributes can help define a common interface in parametric typeface design systems such as Prototypo (Mathey & Babé, 2014), Adobe Multiple Masters (Adobe, 1995), Infinifont (McQueen III & Beausoleil, 1993), variable OpenType fonts (Microsoft, 2018), and Metafont mentioned in section 4.2. In these tools, attributes are represented in terms of numerical parameters which can be changed by users of the final fonts.

4.5.2. The role hypothesis

Douglas Hofstadter, Gary McGraw, and John Rehling worked on an experimental tool called *Letter Spirit* that would create character shapes with a uniform style without human supervision (Hofstadter & McGraw, 1993; McGraw & Hofstadter, 1993; McGraw, 1995; Hofstadter & McGraw, 1998; Rehling, 2001; see also section 2.2). When given a few basic character shapes on input, the tool could distill their style, create the remaining character shapes from the lowercase of the basic English alphabet, and possibly even modify the input characters in order to output a complete, coherent set of new character shapes. The analysis of style was not based solely on decomposition into components. Rejecting the view that whole character-category domains can be represented using a parametric prototype, Hofstadter (1982) proposed a theory that character categories can be conceptualized as sets of general criteria called

roles. A role is not a shape, rather it is a set of general norms (see figure 4.28). By sufficiently filling a particular set of roles, a shape can be identified as a member of a corresponding character category. “Filling a role” means that one or more character components satisfy the norms of the role. At the same time, not all of the roles from the set for the character category need to be filled completely for a shape to be recognised as belonging to the category.

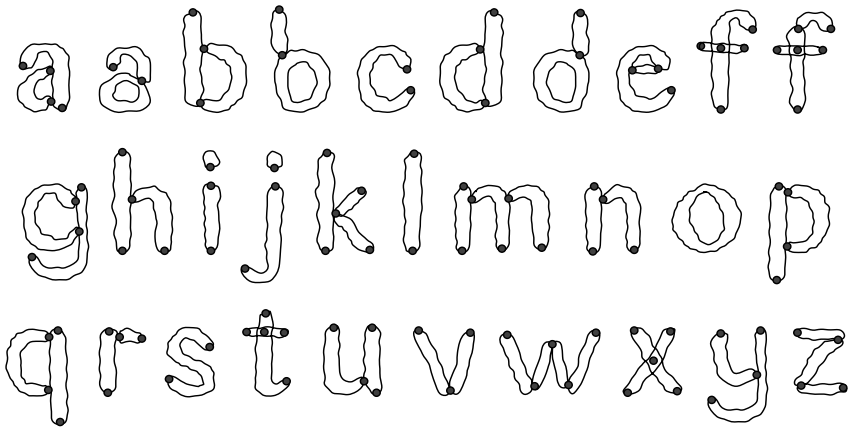


FIGURE 4.28: A visual representation of some of the most standard sets of roles for the lowercase letters of the English alphabet. The outlined areas correspond to roles, the dots mark where the roles end or join. There are other ways of conceptualizing many of the letters.
(from: McGraw 1995, p. 10).

For example, a hypothetical set of roles for ‘A’ might be: having two diagonal components that are connected at the top, being wide open at the bottom, having a component in the middle that connects the diagonals, etc. The role of the connecting component in the middle would normally be filled with a straight bar, but there could be a wavy bar or a dot instead too. At the same time, the connection of the diagonals at the top could vary as well. Either way, the character shape would be still recognizable as an ‘A’ (see figure 4.29) .

FIGURE 4.29: One possible conceptualization of an 'A' described in the text is visualized on the left; the grey ovals corresponding to the individual roles and their overlaps corresponding to the areas where these could join. The horizontal role in the middle of the character could be specified with these norms: i. shape component, ii. straight, iii. placed in the middle, iv. horizontal, v. connecting the other two roles roughly in the middle. The three character shapes on the right show different ways to fill this role. The first 'A' satisfies all the norms, the second 'A' satisfies all the norms except for ii., and the last 'A' satisfies only the norms i. and iii.

An important quality of this approach is that it provides hierarchical decomposition and interpretation of characters: from general character categories, through character concepts described as sets of roles, to more specific concepts described using components, to even more specific character shapes. It maintains the connection between components and the more holistic qualities and thus eliminates issues associated with direct decomposition into components mentioned at the end of section 4.4.

As the same role can exist in conceptualizations of different characters, this approach manages to describe a more general level of coherence (characters having a connecting component) instead of relying on specific shapes (having a specific kind of bar).

A related idea was explored by Blesser et al. (1973) in the domain of character recognition where they proposed that a distinction should be made between physical, perceptual, and functional attributes. The physical attributes refer to shapes, the perceptual attributes refer to the shapes' appearance as perceived by readers, and functional attributes describe the shapes' roles within characters.

Note that roles are, like typographic attributes, generalizations of shapes that appear in one or more characters. While the typographic attributes tend to be predominantly global and are either present or not, roles can be seen as prescriptions of local expectations which may be satisfied (filled) to a certain degree.

Letter Spirit's results were illustrated using a series of so-called grid fonts (see figure 4.30). This limited demonstration may raise doubts regarding the applicability of this approach to conventional typefaces possessing a greater degree of refinement. The authors argued that, in fact grid fonts, despite being seemingly simple, provide a "rich microdomain that focuses attention on cognitive issues of letter perception" (McGraw, Rehling & Goldstone, 1994). However, a tool applying this approach in the design of typefaces for continuous reading does not exist yet.

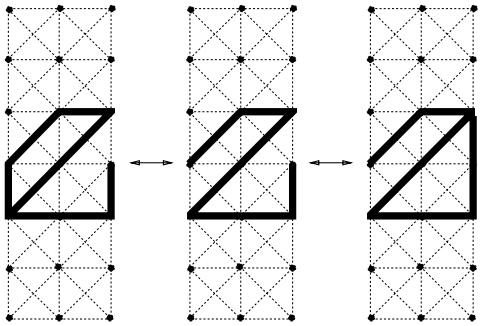


FIGURE 4.30: Three letters from grid fonts used in the Letter Spirit project (from: McGraw 1995, p. 21).

4.6. Towards modelling of coherence

The shape-making and shape-linking approaches discussed above show the diversity of creative, production, and mental processes typeface designers use in their work. While their personal preferences may differ, they undoubtedly use some of the conceptualizations discussed above. It also seems that they often use multiple approaches to benefit from their respective advantages. Of the approaches discussed previously, the stroke-based approach, the transformation, decomposition, and generalization of shapes are particularly useful as they can, in one way or another, help describe similarity between shapes from different character categories. They also seem transferable to other scripts. Thus, they can be considered when modelling coherence in chapter 6.

A note about character spacing and rhythm is appropriate, as designers might consider these factors important for overall coherence of a typeset text. The character shapes in the studies in chapter 5 are presented individually and not within a continuous text. Therefore, the participants have no way of assessing the rhythm or spacing of the characters. However, the participants might observe character counters and relative widths of characters – features that have an effect on rhythm. These are represented in the model, yet by themselves they are insufficient for a proper assessment of an overall rhythm of a text. Therefore, the coherence studied in this research is coherence without consideration for spacing or rhythm; it is the coherence in a collection of character shapes rather than the coherence of a text.

5. Studies with readers

The overall goal of studying coherence perception was tackled through a series of studies with human readers as participants. Each study consisted of a sequence of trials which elicited character similarity responses in the context of other characters. Moreover, this was done with respect to typeface design objectives so that the collected data can eventually inform typeface design practice. For this reason, and to have results based on diverse material, the studies were run on multiple representative typefaces for each script.

In order to reach a sufficiently large number of participants with diverse cultural and geographic backgrounds, the studies were conducted online using a custom-built website. The collected data was later used to calculate observed similarity measures and to verify and optimise a theoretical typeface model (see chapter 6: *Modelling character similarity and coherence*).

5.1. Asking about similarity

Some of the observations about character similarity may come across as obvious and leave one wondering whether there is a need for testing and elaborate verification. Consider, for example, how apparent the similarity between the shapes of letters `n` and `h` seems to be. However, there are more complex cases, e.g. letters `z` and `s`, where the criteria of similarity becomes more of a challenge and it is harder to assume that everyone would perceive the similarity in the same way. There are also cases where it may be very difficult to see any similarity at all, e.g. letters `o` and `i`. Testing the apparent cases might seem like an exercise in verifying the obvious to some, especially typeface designers and typographers. At the same time, testing the very difficult cases might seem like a futile effort providing little insight. The space between those extremes provides opportunities for non-obvious comparisons within a diverse domain.

When looking at all of these cases, there is still a need to distinguish between personal opinion and general agreement. Even when it comes to something as familiar as characters, what is clear to one person may not be as clear to someone else. Depending on their personal predispositions, cultural context, familiarity with the script studied, and education, different people might conceptualize characters in different ways and make different similarity judgements. And more importantly, some might not be aware of their decision-making process when assessing similarity.

Likewise, opinion based on expert knowledge needs to be distinguished from the experience and perception of the general public. While expert knowledge may provide insights about characters and their construction,

it may also influence their perception of similarity. Researchers found that experienced readers discriminate characters more effectively than inexperienced readers (Wiley, Wilson & Rapp, 2016) and that design expertise has an effect in some situations; designers are able to abstract from stylistic variations better than non-designers (Dyson, Tam, Leake & Kwok, 2015). As a result, there can be a considerable discrepancy between similarity as perceived by experts and non-experts, and by extension, between a model built on expert knowledge and knowledge of a general audience. After all, it is one of the research questions, to see whether there is a difference between these two views.

At this point, it is also important to note that pre-selecting combinations of characters that seem to bear similarity above a certain threshold should be avoided as it introduces bias, whatever the threshold. For example, if the studies were to cover only combinations that typeface designers already find similar, these combinations would not challenge the model sufficiently. The theoretical typeface model ought to represent the whole spectrum of similarity relationships and thus should be tested with combinations of varying expected similarity, from obvious to little or possibly none.

5.1.1. Enquiry about parts

An approach to similarity that represents characters as collections of features, as mentioned in the introduction and developed in chapter 6: *Modelling character similarity and coherence*, might suggest that the best approach to testing is to ask participants directly to identify similar parts in juxtaposed characters. The following discussion of two basic, possible approaches shows why this is not a good idea.

One approach would be to divide the whole character shapes into parts geometrically (see figure 5.1) and present this division to the participants. This is problematic on two levels. First, it is not always clear where parts border with each other and whether they overlap. Any proposed division might be one of many, and could differ from what the participants would consider natural or helpful. More importantly, assuming that readers in fact use parts to assess similarity between characters limits the data to a particular view of similarity and a particular way of how the parts are used. Avoiding such an assumption helps to collect more generalizable data that can be re-used in various other experiments.

Alternatively, one can provide participants with drawing tools to mark up the characters and ask them to mark areas that seem to contribute to shape similarity in their opinion (see figure 5.2). One can imagine tools like those used to annotate PDF documents. However, particular kinds of tools influence what is selected and what is not. For example, participants provided with a thick virtual marker pen would mark larger areas as opposed to those with finer tools. Even markers that are generally considered freehand tools are far from being unbiased with regard to the content that is being marked up.

Put another way, different kinds of content require different marking tools. Providing a tool would mean that some decisions have already been made regarding the nature of the character parts and the similarity relationships in general. And this is leaving aside the fact that participants might have very different competencies when handling the marking tools.

FIGURE 5.1: Character decomposition into parts is suggested in the wording of the instruction, and in the visual separation of parts that users are asked to select. The figure shows the state before a user's selection.

FIGURE 5.2: Character decomposition into parts is suggested in the wording of the instruction, and by the marking tools available to users. The figure shows the state after a user's marking.

Moreover, some aspects of similarity are likely to be holistic or based on relationships within individual characters. The focus on identifying parts would remove those holistic aspects and affect outcomes of the testing.

Either way, having ready-made decomposition or providing tools to mark parts is nudging participants to think in terms of parts or features while precluding other conceptualizations. However, the nature of how people think about shapes is an essential concept of the approach that is being hypothesized and used for modelling in chapter 6. Therefore, data collected following the two approaches discussed here would not provide a valid verification. This means that to eliminate any bias regarding principles that underpin similarity judgements, characters ought to be presented as wholes without any decomposition suggested graphically or in wording.

5.1.2. Enquiry about similarity

Another major consideration is the way of asking participants about their assessment of similarity. One cannot simply enquire about the mere existence of similarity between two shapes as there always tends to be some degree of similarity. And this applies even more when it comes to similarity between character shapes from high-quality text typefaces.

In their comprehensive overview, Shane T. Mueller and Christoph T. Weidemann (2012) analyse character similarity studies spanning over a century. The majority of the reviewed similarity studies used an approach that involves presenting individual characters one by one and asking participants to name them. Each name given as a response to a particular character presented is recorded. The total count of each pair of presented–named characters (stimulus–response pair) is assumed to be an indication of their confusability[1] (or similarity) and allows for the construction of a confusion matrix or, alternatively, a similarity matrix (for format see table 5.1).

When presenting characters in clear and readable conditions with no time limit for the identification task, only a few errors are generated which prevents making useful conclusions about character similarity (Simpson, Mousikou, Montoya & Defior, 2013). In other words, this way of measuring similarity is not sensitive enough. In order to challenge participants' perception, and to therefore produce more errors, researchers use variations of this approach. These include brief presentation (Bouma, 1971; Mueller & Weidemann, 2012), reducing the letter size or increasing reading distance (Bouma, 1971; Phillips, Johnson & Browne, 1983), presenting the characters in the peripheral visual field (Alexeeva & Konina, 2016; Reich & Bedell, 2000), showing only some parts of the characters (Fiset et al., 2008; Rosa, Perea & Enneson, 2016), or making the viewing conditions worse by means of lower contrast (Geyer, 1977) or tight inter-character spacing (Liu & Arditi, 2001).

However, using overly challenging conditions may be inappropriate to study similarity as characters are viewed in unfamiliar conditions that may prevent careful assessment of the shapes. Moreover, Daniel Fiset et al. (2009) note that low contrast and brief presentation "exacerbate the relative importance of low spatial frequencies".[2] In an attempt to keep the viewing conditions as normal as possible, this work avoids the aforementioned manipulations and tests based on individual character presentation.

Simpson et al. (2013) mention that another issue with these studies is a potential confusion when responding by naming a character based on the

1. Although not identical, confusability is often considered to be inversely related to similarity as more similar letters are easier to confuse.
2. The spatial-frequency channels roughly correspond to different levels of detail in the object observed. High-frequency channels contain information about fine details while low-frequency channels contain information about overall features (Palmer, 1999). See section 6.1 for more details.

phonological similarity of the character names. And naming characters could become even more problematic if participants were unfamiliar with the character names in the script studied.

Another approach used by researchers is soliciting similarity judgements of character pairs by means of subjective ranking. Participants are presented with pairs of characters and asked to rank their similarity on a scale (Boles & Clifford, 1989; Simpson et al., 2013) (see table 5.1). Although, it can be alleviated by statistical methods, participants tend to have a personal approach to how they use the scale and the results may become less reliable. For example, some may be reluctant to give the highest or lowest score, others may prefer to use only the extremes.

	a	b	c	d	e	f	g	h	i
a		2.13	2.50	2.57	3.40	1.06	3.30	1.57	1.16
b	2.13		3.03	5.60	2.27	1.83	3.53	3.70	1.43
c	2.50	3.03		3.57	4.43	1.24	2.47	1.30	1.10
d	2.57	5.60	3.57		2.87	1.43	4.10	2.73	1.73
e	3.40	2.27	4.43	2.87		1.33	2.37	1.23	1.20
f	1.06	1.83	1.24	1.43	1.33		1.40	1.80	4.16
g	3.30	3.53	2.47	4.10	2.37	1.40		1.57	1.53
h	1.57	3.70	1.30	2.73	1.23	1.80	1.57		1.93
i	1.16	1.43	1.10	1.73	1.20	4.16	1.53	1.93	

TABLE 5.1: A portion of a similarity matrix provided by Simpson et al. (2013). The rows and columns are indexed with letters studied. Each cell at the intersections corresponding to two different letters contains the observed similarity measure for these letters as it was deduced from their character-pair experiments. Higher numbers signify stronger similarity relationships. The diagonal values are missing as the similarity of letters to themselves was not tested. Note that the matrix is symmetrical as the order of characters in pairs was randomized and the similarity relationship was assumed to be symmetrical.

More importantly, testing pairs relies on an implicit knowledge of the script. Tversky (1977) showed that similarity judgements are relative, i.e. dependent on the context. Thus, when asking for the similarity of a pair, it is implied that it is a similarity of the two characters within the context of the whole script. For example, the pair `b, h` is relatively more similar in the triplet `b, h, y` than it is in the triplet `b, h, n` where the pair `h, n` is more similar than other pairs in the triplet. When ranking a character pair similarity, participants have to estimate all the relative similarities to other, possibly yet unseen, characters. It does not seem safe to assume that even competent readers have a complete knowledge and recollection of all character shapes belonging to a particular script and

their expected similarities. Not to mention that they may not be familiar with the typeface that is used in the study. Thus, results collected through character pair ranking would have only a limited application to work that is concerned with understanding the similarity relationships that constitute coherence of typefaces.

5.2. The approach taken

In order to examine coherence better, it makes sense to try to obtain more detailed data by considering the relativity and the influence of the context on the similarity judgements systematically. This can be tackled by presenting the characters in groups. Participants could be given a simple task that encourages them to make their judgements based on their subjective notion of similarity.

In order to study the context's influence, the scope has to be kept manageable. With too many contextual characters in the group, it would be difficult to say which of them influenced a participant's response to the task, and how. Therefore, the characters are presented in groups of three characters. The participants are tasked to identify the most different character shape from the other two (the *odd one out*). Assuming similarity judgements are inverse to difference judgements (Tversky, 1977), this is the same as asking them to select the two most similar characters from the triplet. However, the explanation of the task is much simpler as it builds on participants' ability to discriminate the odd one out learned from previous experience with widely used studies and games using a similar approach. As the pilot study showed (see section 5.6), participants feel confident in relying on their intuition when the task is presented in this way. It is also easier to select one character to eliminate than two to keep, both in terms of participants' task and in terms of the technical implementation.

Presenting triplets of characters means that there is always a third character, establishing context for each pair similarity judgement. Different contextual characters for each pair are used during the testing which helps in exploring how the contextual character influences the similarity measure of a particular pair.

Keeping the groups of characters small also makes the task cognitively simpler as it involves fewer similarity judgements than larger groups would. In order to solve the task, participants need to establish their own criteria for similarity for all the pair combinations in the group, i.e. for triplet `a, b, c`, they need to make decisions about the relative degree of similarity between `a, b`, `b, c`, and `a, c` – that is, they have to make three similarity judgements. On the other hand, a quadruplet `a, b, c, d` would require six such judgements: `a, b`, `a, c`, `a, d`, `b, c`, `b, d`, and `c, d`. For a set of five characters, it would be ten, etc. The greater the number of judgements that need to be made, the more complicated and time-consuming each task becomes.

It is worth noting that relative comparisons were introduced by Kelly (1955) in his repertory grid interview technique which is used in the psychology of personal constructs. Interestingly, it also allows collection and mathematical evaluation of the personal constructs. When using Kelly's technique, the participants are typically asked to explain why they find two items similar to each other and thereby different from the third. In this study, participants are not asked to explain their decision and simply select the most different item. Nevertheless, the idea that triplets can provide small, yet sufficient context for participants' decisions is the same.

To my knowledge, this approach has not yet been used to study character similarity. However, relative comparisons or their variants, e.g. "A is closer to B than A is to C", have been used to study similarity of text documents (Schultz & Joachims, 2004), illustrations (Garces, Agarwala, Gutierrez & Hertzmann, 2014), 3D-object models (Lun, Kalogerakis & Sheffer, 2015), or multi-modal similarity (McFee & Lanckriet, 2011). In the last study, three objects were presented, one of them fixed, and the participants were asked to select one object from the other two that is most similar to the fixed one. They had an option to indicate that neither of the two is similar or that they cannot decide. Note that such a task is directional. In the example above the object 'A' is a key object and other objects are compared with it. Directional similarity statements like "the similarity of A to B is..." can lead to different results than the statement "the similarity of B to A is...". This was already discussed by Tversky (1977).

In the approach taken in this research, participants eliminate one character and the order of characters within each triplet is randomized (see section 5.2.1). This effectively removes any directionality, and the pair similarity statements can be considered symmetrical. They are more like "the similarity of A and B is..." where the order of 'A' and 'B' does not matter (see also section 6.1).

The option to "pass" was not adopted in this research in order to obtain results even for the more challenging triplets. The number of individual trials presented in one session is relatively small (see section 5.4.2), hence the danger of participants finding the task daunting and giving up on it is relatively small. An indication of the overall confidence or uncertainty of participants' judgements can still be obtained from the data. Therefore, there is no need to time participants' responses, either.

5.2.1. Triplet sequence and character pair similarity

Triplets are presented to each participant in a sequence. Each sequence contains triplets made of characters from a particular typeface and a script. The order of the triplets in the sequence is randomized across participants in order to eliminate the influence of the order on the similarity judgements.

The order of characters within triplets is also randomized. Consequently, the order of characters can be considered unimportant, and the similarity judgements within the triplets a, b, c or b, a, c are treated as the same.

The similarity is called *specific character pair similarity* when it is studied with a specific contextual character, e.g. the similarity of a, b in the context of c, i.e. in the triplet a, b, c. It is called *generalized character pair similarity* when it is studied with multiple characters used as the third contextual character. For example, similarity of a, b in the context of c or d, as in the triplets a, b, c and a, b, d. The larger the number of contextual characters, the more general the similarity. A measure of more general similarity is useful for making comparisons with other studies or assumptions about character pair similarity in new contexts, e.g. in quadruplets or words.

The generalized character pair similarity measure can be approximated from aggregated triplet responses involving a specific pair. For each triplet a, b, c and response c, the similarity of the pair a, b increases while the similarity of a, c and b, c does not. In order for the comparison of such pair similarity measures to make sense, the set of contextual characters has to be consistent for all of these measures, i.e. the collection of triplets from which the pair measures were deduced has to be the same. For example, it would not make sense to compare a similarity measure of x, y based on triplets x, y, v and x, y, z with a similarity measure of x, v based on triplets x, v, a and x, v, c. Clearly, the results would be skewed based on whether the contextual characters in the triplets were more similar, such as in the triplets for the first pair, or less similar, such as in the triplets for the second pair.

Therefore, in order to keep the context consistent, the triplets in the sequence are an exhaustive collection of all possible combinations of three characters from a predefined, fixed set of characters for each study (see section 5.4.2). The exact method for calculating the generalized character pair similarity measure is further discussed in section 5.13.

5.2.2. Presentation

In order to obtain realistic data, the triplets of characters should be presented in a way that is congruent with readers' daily experience of character shapes, i.e. not obscured or under non-standard lighting conditions.

As discussed in section 2.1, characters are visual representations associated with linguistic units. In order to study characters' visual attributes, the study discourages their linguistic interpretation. This is possible due to the fact that the association between the linguistic units and characters is purely conventional; the units do not determine the appearance of the character shapes in any way and vice versa. An identical shape can represent different sounds in different languages.

Participants who are used to reading characters regularly might not realize the essence of the task at hand: to inspect only character shapes and their appearance only. For example, participants might consider vowels more related to each other than to consonants. That would be incorrect within the scope of this work. Hence, the study has been designed to encourage participants to "see the shapes" as opposed to "read the symbols". This is addressed in the wording of the instructions: "please consider only the shapes, not the meaning of the letters or letter groups". It is reinforced in the presentation of the triplets as a row of three equal *sample squares* with a thick grey border around characters in larger-than-reading size.

When it comes to text typefaces, most typeface designers would at first propose to study them in text sizes (around 9–14 points body size which amounts to a base height[3] of around 0.14–0.22 cm or 0.06–0.09 inches) set in black colour on white background, as this is how they are most likely to be used. This is a reasonable proposal as it allows to assess the overall rhythm, weight, and other typographic attributes; a text can be read easily. On the other hand, these sizes are too small to make judgements about the characters easily without moving closer or squinting. Moreover, for studies conducted online, the low resolution of some computer screens could have a distorting effect on the character shapes. Therefore, and opposed to the initial intuition of typeface designers, it is more practical, and convenient for participants, to show the characters at a size larger than text size, ideally with their base height above 2–3 cm, 0.8–1.2 inches.

Yet, use of very large sizes could be also misleading. As the studied typefaces are intended for text, it is reasonable to assume that fine details do not affect the perception of similarity directly and are irrelevant to the testing. The limits of human vision or technology cause these details to disappear at text sizes. Showing characters at larger sizes might expose these details and have participants examining them in a way that is unrelated to what is perceptible at text size. The pilot study (see section 5.6) showed that such overly detailed examination concerns especially experts, e.g. typeface designers and typographers. It seems to be less of a concern with non-experts who are unaccustomed to inspecting character shapes up close. To address this issue, the participants are instructed to proceed quickly and not to dwell on fine details, and very large character sizes (above 300 points body size, base height around 4.7 cm or 1.8 inches) are avoided.

In the final design used for the studies, the sample squares were 250 pixels wide which is around 5 cm (around 2 inches) tall on average computer screens (see also section 5.5). The Latin and Cyrillic characters were set on the same baseline and scaled so that their base height occupies around 35% of the vertical space of the sample square. The Devanagari characters were

3. For the sake of these approximations the base height is taken to be around 40% of the body height. See figures 3.3, 3.9, and 3.14 for diagrams showing the base height in different scripts and section 3.1 for its definition.

set on the same baseline and scaled so their base height occupies around 60% of the vertical space. This setting allowed enough clearance around the character shapes within the sample squares. For examples of the presentation, see figures 5.3 and 5.4. All character shapes from the studies are shown in appendix 1.

FIGURE 5.3: A character triplet 'b, w, f' from the typeface PT Sans as it was presented in the studies.

FIGURE 5.4: A character triplet 'श, ए, न' [Sha, E, Na] from the typeface Ek Mukta as it was presented in the studies.

Keeping the base height consistent is a customary typographic practice when comparing typefaces (see the last paragraph of section 3.1). Following this practice, all characters in all studies for a particular script were presented in comparable settings. Furthermore, the difference in scaling between Latin, Cyrillic, and Devanagari ensures that characters from each script appeared equally important.

It is possible that the samples, presented on a website, might be displayed smaller or larger depending on a participant's device – especially if it is a mobile device – and any custom screen or browser settings. This should not constitute a major problem as the size of the characters is likely to remain larger than text size (0.49 cm) and smaller than the very large size (10.58 cm) mentioned above. It is unlikely that such differences in scale would affect the results of the study. Reducing text typefaces, even by a factor of two from the default size of the sample, would not make important features disappear. The effect is still much less distorting compared to what occurs in the studies mentioned in section 5.1.2.

5.3. Organisation of the studies

As mentioned above, the studies were conducted online using a custom-built website. Each study consisted of two parts: a questionnaire and triplet testing. There were two kinds of studies for each script: *simple* and *combined* studies (see figure 5.5). The simple studies contained one sequence of triplets made of characters from one typeface and a particular script. The combined studies contained two sequences of triplets, each with triplets made of characters from a different typeface for a particular script. Neither of the two kinds of studies mixed scripts, e.g. by showing Devanagari and Latin characters together. It is therefore possible to refer back to the studies based on the script they tested, that is, Devanagari studies are those which used characters of the Devanagari script.

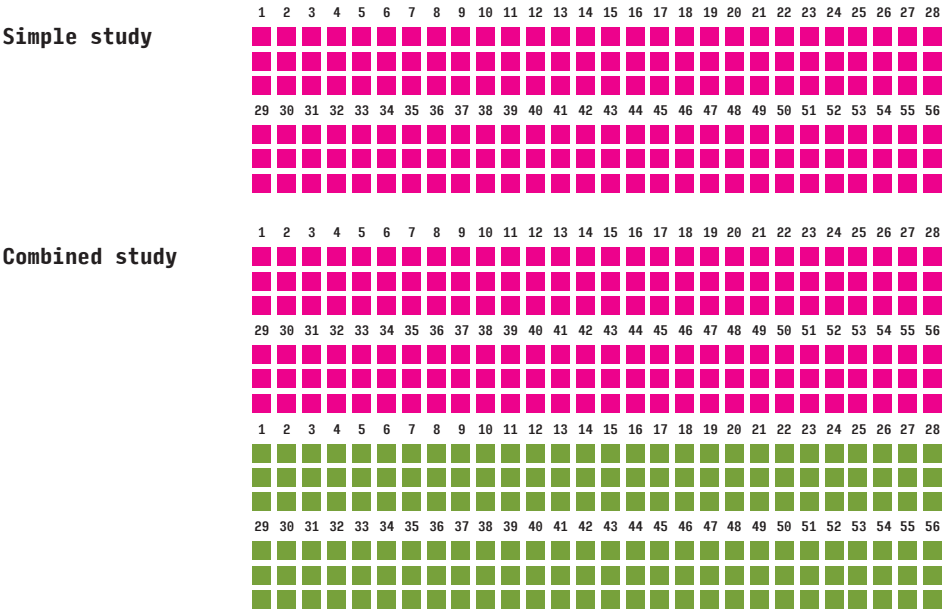


FIGURE 5.5: Simple studies contain one sequence of 56 trials (marked in magenta). Combined studies contain two sequences of 56 trials (marked in magenta and green), each with a different typeface.

5.3.1. Questionnaire

The ability to recognize characters is learned. Hence, there is likely to be an influence of culture and education on the way people conceptualize characters and make judgements about their similarity. In order to record these potential influences, participants completed a questionnaire with questions about:

i. their reading skills in terms of various scripts
ii. their age
iii. their experience as readers
iv. their possible experience as designers

The intention was to collect a reasonable amount of background information about participants in a simple, non-intrusive way. The full questionnaire with all questions and predefined answers is shown in figure 5.6.

There were two questions which attempted to collect information about script-reading skills from participants: one about their native language, and another about languages they can read fluently. The focus of these questions was on "language" rather than on "script" since non-experts often confuse these terms, and more academic terms like "writing system" or "script" can be unclear to them.

Script-reading skills can sometimes be deduced from language skills. This may not always be possible since native speakers of a language are not always fluent readers in the corresponding script. Moreover, some languages, such as Serbian, are written in multiple scripts, and in such a case the deduction from language to script may not always be clear. Nonetheless, language skills provide some indication of the cultural and educational context, and asking about language is still better than asking participants a question they may find confusing.

It is useful to maintain a distinction between scripts learned primarily (native language) and secondarily (fluent languages) as the education might have a different effect in each case.[4]

Notice that the second question was "Which languages can you read fluently?". Unlike the first question, the second question asks specifically about a participant's ability to read. It is also possible that people might not read their native language in its customary script, however unlikely that may sound. The second question gathers additional confirmation of a participant's reading skills.

Readers of different ages might have been educated differently, which might result in different similarity judgements. At the same time, a participant's precise age is not relevant. In order to simplify the process, and to avoid unnecessary intrusion into the participants' privacy, the questionnaire provided age ranges when asking about age.

In order to avoid common issues with self-assessment and categorization in questionnaires, reading experience was inferred through asking an indirect question about frequency of reading.

The question about possible design skills, was, on the other hand, direct. The distinction between readers without any design expertise and graphic designers, typographers, and letter designers is important since, as was already mentioned, professional training can affect the perception of similarity. And lastly, participants could also provide a contact email address to stay informed about the results of the study.

4. It is not the primary objective of this study to find an influence of language proficiency on similarity judgements. In case there is any, the data should only help to hypothesize an emerging pattern.

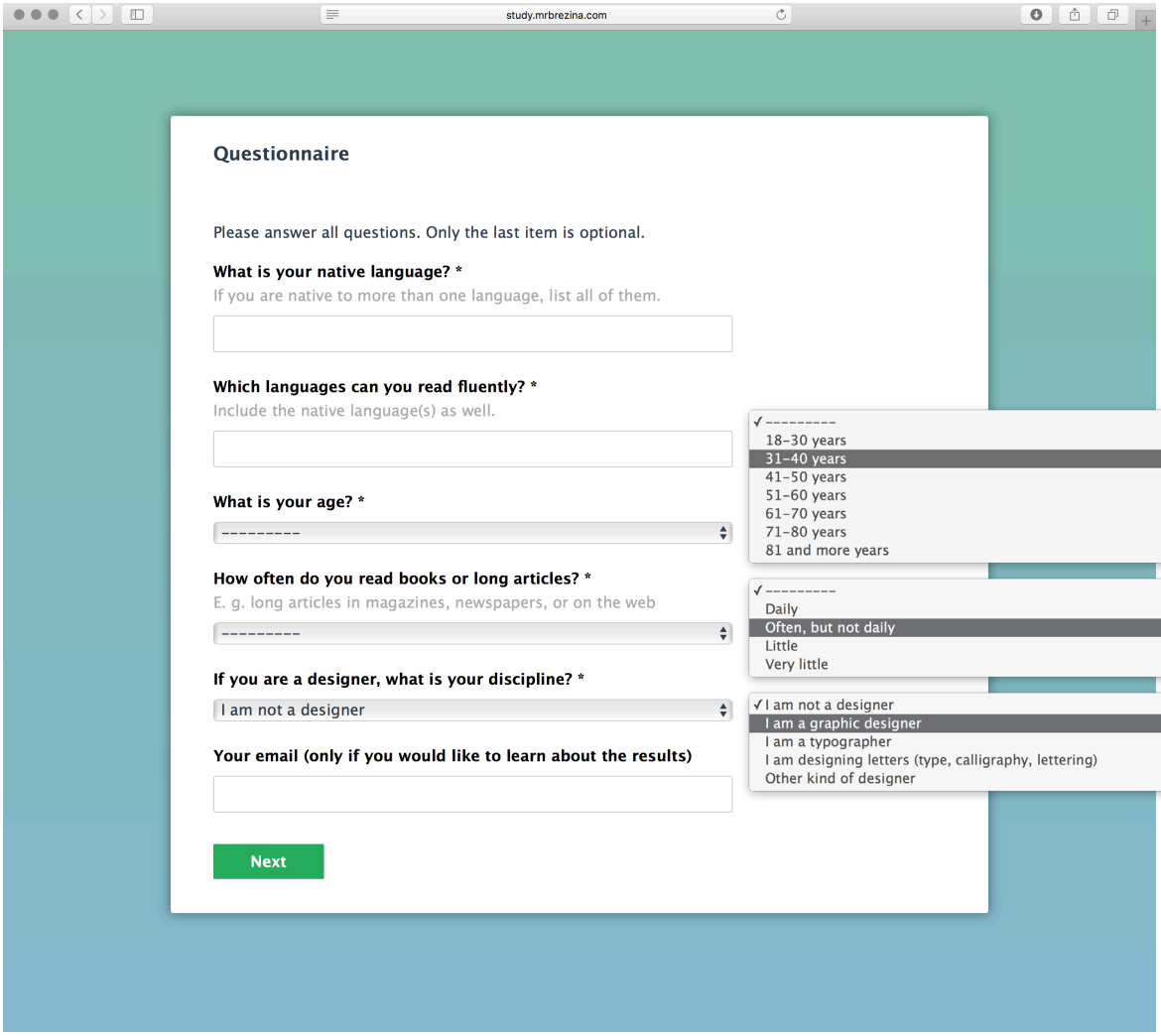


FIGURE 5.6: Questionnaire with answers from the drop-down menus expanded on the right.

5.3.2. Triplet trials

In this part of the study, participants worked through one or more sequences of character triplets. In each step, referred to as a trial, participants saw three character shapes from three different categories, i.e. three characters, each with a different linguistic unit associated with them. They were asked to click on the most different character shape from the other two.

The sequence of triplet trials was introduced by a short instruction page which indirectly discouraged participants from being overly fussy about the shape details. Participants were instructed to proceed quickly and to disregard the meaning of the characters or character groups (see figure 5.7). Additionally, to avoid participants' confusion with more challenging triplets, it was emphasized that there are no right or wrong answers. The individual trials were not timed for technical reasons.

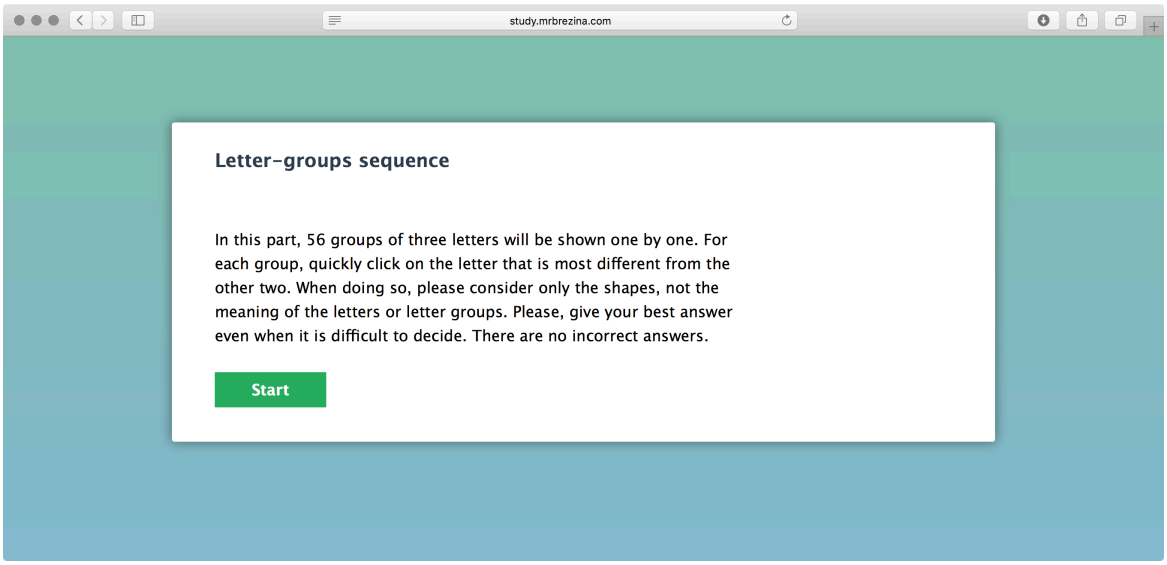


FIGURE 5.7: Instructions for the triplet sequence discouraged participants from considering the meaning of the characters and character groups and encouraged them to proceed quickly.

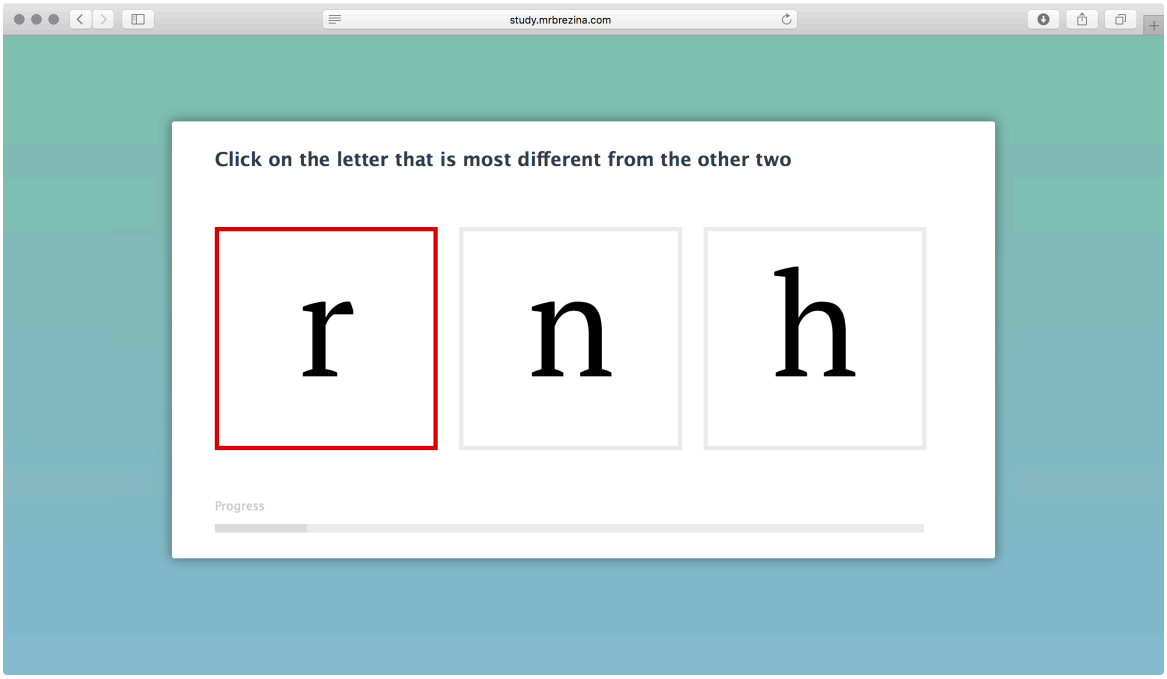


FIGURE 5.8: The presentation of the character triplets in the triplet sequence discouraged participants from reading the triplet as a single word by separating them into individual boxes.

An example of a typical triplet test is shown in figure 5.8. The presentation of the triplets was already discussed in section 5.2.2. The three characters were presented in three equally sized sample squares. When participants aimed their cursor at a sample square, the square's border turned red to signify it was selected for submission. Clicking on the square recorded the answer and showed another triplet from the sequence. Technically, the squares with characters were produced in advance as vector-based images in SVG format (Dahlström et al., 2011) to avoid any problems with font loading and to achieve as good a resolution as possible.

As mentioned previously, triplets were shown to each participant in a random order. The order of characters within a triplet was also randomized. The order of the two triplet sequences in the combined studies was also randomized. This was done to avoid any bias caused by participants transferring their similarity criteria from one triplet or typeface to another. The order of sequences in each combined study was recorded to allow inspection of any patterns resulting from such transfers.[5] The order of triplets or characters within the triplets was not recorded.

Responses for the whole study were saved at the very end. If the participants abandoned the study, the incomplete results were not saved. The progress bar at the bottom of the page gave participants an indication of how many triplets were left to view.

5. Due to the limited amount of the data collected, the possible transfer of criteria between different typefaces in the combined studies was not ultimately explored. The studies were treated as two simple studies.

5.3.3. Language variants

For participants' convenience and to avoid excluding participants who could not read the English instructions, the Cyrillic and Devanagari studies were both translated to one major language using the corresponding script. The studies for the Cyrillic were made available in Russian and English. The studies for Devanagari were made available in Hindi and English. The Latin-script studies only had instructions in English. The translations were made by native speakers with fluent command of the respective languages.

5.4. Scope

Only a limited number of typefaces and triplets were tested, as it would be unfeasible to test all triplet combinations of all characters for all text typefaces for the three scripts studied. The following section explains why it would be unfeasible more thoroughly, and also describes the process of choosing typefaces and characters for the individual studies to ensure they were representative of each script's conventions and visual environment.

5.4.1. Typefaces

Most of the representative typefaces used in the studies were contemporary typefaces that can be considered fairly common and usable for continuous text, as in magazines or books. Many of them have been distributed with major operating or publishing systems, and have become, or are bound to become, familiar to the reading population. Both high-contrast[6] serif and low-contrast sans-serif typefaces were included in the Latin and Cyrillic samples. Similarly, high-contrast as well as low-contrast typefaces were selected for the Devanagari sample. Note that the selection was not attempting to cover all variations outlined by common type classifications. Instead, the focus was on contemporaneity and prevalence of these typefaces in the visual environment of the corresponding scripts. Availability of quality Unicode-based fonts for the corresponding typefaces was also an important practical factor, especially for Devanagari.

Eight typefaces were initially selected for each script. A group of eight typefaces is still feasible to test and, together with the character selection method described below, it allowed for each character from the basic alphabets or syllabary to be represented at least once, often twice.

Note that due to the success of the call for participants for the Latin studies on social media, there was an opportunity to test with more typefaces. This had to be done quickly to make use of the momentum, so additional

6. For the purpose of this work, the term high-contrast refers to Didot-like high-contrast typefaces as well as to typefaces with more moderate contrast. The term low-contrast, on the other hand, refers to low-contrast as well as monolinear typefaces with nearly no contrast at all. For definition of typographic contrast, see section 4.2.

typefaces were selected for convenience from among system fonts. Three of these were from the so-called ClearType Font Collection that comes preinstalled with MS Windows Vista and newer (from 2007 on) operating systems. The characters and triplets tested were reused from the original group of Latin-script typefaces, with only minor modifications.

Typefaces used for the Latin and Cyrillic scripts were: Arial, Century Schoolbook, Courier New, Georgia, PT Serif, PT Sans, Times New Roman, and Verdana. The additional typefaces used for the Latin script were: Calibri, Cambria, Candara, and Futura. Typefaces used for the Devanagari script were: Adobe Devanagari, Devanagari MT, Ek Mukta, ITF Devanagari, Kohinoor Devanagari, Lohit Devanagari, Murty Hindi, and Nirmala UI.

PT Sans and PT Serif were used together in the combined studies for Latin and Cyrillic. ITF Devanagari and Kohinoor Devanagari were used in the combined studies for Devanagari. Note, however, that there were also some simple studies conducted with these typefaces. Regular-style fonts were used to produce the character samples for the studies except in the cases of Futura and ITF Devanagari, where the Medium style was used instead, as the Regular style was not available.

5.4.2. Characters and triplets

As discussed in section 5.2.1, the collection of triplets should be an exhaustive set of all triplet combinations from a limited set of characters. Testing all triplet combinations for all characters in a typeface is clearly unfeasible. Therefore, the studies focused on lowercase characters from the basic English alphabet for the Latin script, lowercase characters from the basic Russian alphabet for the Cyrillic script, and a subset of the basic Hindi syllabary for the Devanagari script (see section 3.4 for more details regarding this choice). These three groups of characters are referred to as *characters studied*.

But even with this limited set of characters, an exhaustive testing of all potential combinations would require a very long study. The number of all triplet combinations[7] is calculated as $\binom{N}{3}$ where N is the number of elements in the set. For a set of 26 elements, this turns out as $\binom{26}{3}=2600$.

Based on the pilot study (see section 5.6), it would take a participant about 10–12 minutes to work through 112 triplets and about 6–7 minutes to work through 56 triplets. At this pace, it would take approximately 4 hours for a single participant to go through all of the 2600 combinations. This is assuming they would be able to remain focused for such a long period of time. It would not be feasible, within the constraints of this work, to test this many combinations with a sufficient number of participants and typefaces.

7. In mathematics, the term combination refers to a selection of a given number of elements from a larger number without regard to their arrangement. The term is used here with the same meaning, hence the corresponding equations.

It seemed more reasonable to keep the length of the study short so the participants would not have a problem staying focused. A long and tedious study could lead to less reliable results and dropouts during the testing. By keeping the number of triplets at $\binom{8}{3}=56$, the number of all unique triplet combinations made of eight characters, the time required for a single participant to complete the study would remain at a reasonable 6–7 minutes. The groups of eight characters for each study are called *study sets*. They varied for each study in order to represent all characters studied as equally as possible.

The key challenge lay in ensuring that the exhaustive collections of triplets made from the study sets contained challenging and diverse similarity relationships – or, in other words, ensuring the triplets would cover a wide spectrum of possible situations.

One way of addressing this would have been to use random sampling of characters. This seemed problematic, as it is not immediately clear whether the triplets generated from a small number of randomly selected characters would include the diverse relationships that may occur in the complete set of combinations. In other words, a random process could not guarantee that the similarity judgements required by the generated 56 triplets would be as diverse as the judgements required by combinations from the total of the 2600 combinations. Therefore, rather than taking a random approach, a systematic stratified sampling with the goal of creating combinations with varying difficulty was used.

Systematic grouping of characters into the attribute groups could potentially introduce bias by pre-selecting compelling combinations of characters to be included in the study sets. This was reduced by using exhaustive combinations of these characters, thus producing a variety of unexpected combinations. Moreover, the characters studied were represented as evenly as possible across the studies for a particular script. The discussion in section 5.11.3 shows that the resulting selection covered a wide range of similarity judgements in terms of difficulty.

Cyrillic	Devanagari	Latin
Halved	**Tailed**	**Arched**
в з э + а в е х ё	र स त + क ख झ द ल श	h m n + a b d f g p q r s
в з э + а в е х ё	र स त + झ द ल श	h m n + a b d f g p q r s
Top bar	**Semirounded**	**Semirounded**
г п т + д л ъ	च ञ व + क ख घ च ज ब	b p q + a b d p q u
г п т + д л ъ	च ञ व + क ख घ च ज ब	b p q + a b d p q u
Vertical	**Hanging**	**Diagonal**
и м н + и й ч ъ	ट ठ द + इ ङ ड ढ ह	k v y + s w x z
и м н + и й ч ъ	ट ठ द + इ ङ ड ढ ह	k v y + s w x z
Leggy	**Orthogonal**	**Narrow**
ж к я	न भ म + ए ग	j i r + l f t
ж к я	न भ म + ए	j i r + l f t
Rounded	**Zigzag-like**	**Rounded**
о с ю + б е ф э ё	इ ड ह + ङ झ	c e o + a b d e g o p q
о с ю + б е ф э ё	इ ड ह + ङ झ	c e o + a b d e g o p q
Descending	**Rounded bottom**	**Halved**
р у ф	ण प फ + थ य ष	s a z + e x
р у ф	ण प फ + थ य ष	s a z + e x
Bottom bar	**Stack**	**Ascending**
ц ш щ + д	उ छ ध + अ घ	l f t + b d h k
ц ш щ + д	उ छ ध + अ घ	l f t + b d h k
Small belly	**Loop**	**Top-opened**
ъ ы ь	ढ थ श + छ	u w x + v y
ъ ы ь	ढ थ श + छ	u w x + v y

TABLE 5.2: Attribute groups for Cyrillic, Devanagari, and Latin. Each attribute group contains three main characters and optionally other additional characters (after the plus sign, not all are shown). The main characters divide the characters studied for each script exclusively. Each attribute group is illustrated by characters from high-contrast and low-contrast typefaces to show potential differences in their designs.

The characters studied were divided into groups based on observed attributes.[8] Each of these *attribute groups* contained three main characters and other additional characters. All characters from a particular group contained a particular visual attribute (or a combination of attributes), and this attribute was assumed to be constant across all of the typefaces studied. The chosen attributes had to be general enough to eliminate the possibility that a particular character would fall out of the attribute group by the specificity of the design of a particular typeface. The main characters from all attribute groups divided the characters studied for a particular script exclusively, i.e. each character appeared as a main character only once, in an attempt to distribute them evenly across the attribute groups.

Having established the attribute groups (see table 5.2), the study sets were composed in order to satisfy the following four requirements:

i. Each set must have three main characters from one attribute group.
ii. Each set must have at least two characters from another attribute group.
iii. Each set must have some characters from yet another attribute group.
iv. All characters studied should be represented among all the study sets and they should be represented as evenly as possible.

Each study set selected this way includes characters with at least three different attributes. Some of the attributes were represented through more characters, thus their combinations include potentially strong as well as weak similarity relationships (see section 5.4.3). Importantly, each study set has three main characters selected so that each attribute group is represented. The other attribute groups represented in a study set (points ii. and iii. above) were selected randomly while maintaining some diversity among the combinations and representing all characters as evenly as possible.

In cases where the characters could not be represented perfectly evenly among the study sets, characters made up of a combination of shapes including an already well-represented character shape were left under-represented, as their shape contributed less to the overall shape diversity. For example, the letter й is made using и in Cyrillic script and therefore й is included in fewer study sets. Additionally, characters that appear less frequently in common texts were left under-represented, e.g. the independent vowels in Devanagari script. The resulting study sets are shown in tables 5.3–5.8.

8. The term attribute is used here instead of the analogous term feature which is used in chapter 6: *Modelling character similarity and coherence*. This is to emphasize that the attributes used here, and the corresponding groups, may be different and typically more loosely defined. While related, their motivation and use is different.

Typeface & style name	Main attribute	Other attributes		Additional
Arial Regular	Descending р у ф	Vertical й м	Rounded с ю	ж
Century Schoolbook Regular	Leggy ж к я	Halved а ё	Top bar г д п	
Courier New Regular	Rounded О С Ю	Bottom bar Д Ц	Leggy К Я	Л
Georgia Regular	Vertical и м н	Rounded б ф	Halved е з	ч
PT Serif Regular*	Halved в з э	Top bar л т	Descending р у	х
PT Sans Regular*	Top bar г п т	Rounded е о э	Bottom bar щ	х
Times New Roman Regular	Bottom bar ц ш щ	Halved а ё	Vertical й ъ	б
Verdana Regular	Small belly ъ ы ь	Halved а в х	Vertical и ч	

TABLE 5.3: Study sets for Cyrillic typefaces with characters grouped into attribute groups. The typefaces and styles specified correspond to those used to create character samples in the studies. Typefaces marked with an asterisk were used in combined studies.

Typeface & style name	Main attribute	Other attributes		Additional
Adobe Devanagari Regular	Tailed त र स	Semirounded ख ब	Stack अ घ	ष
Devanagari MT Regular	Stack उ छ ध	Zigzag-like ड ह	Semirounded च घ	ट
Ek Mukta Regular	Orthogonal न भ म	Semirounded क ज	Loop ढ श	ए
ITF Devanagari Medium*	Semirounded च ञ व	Orthogonal भ म	Tailed झ र	ठ
Kohinoor Devanagari Regular*	Zigzag-like इ ड ह	Orthogonal ग न	Bottom rounded प	ब स
Lohit Devanagari Regular	Bottom rounded ण प फ	Bottom rounded थ य	Tailed क ल	ग
Murty Hindi Regular	Loop ढ थ श	Tailed त द	Bottom rounded ण ष	य
Nirmala UI Regular	Hanging ट ठ द	Stack उ छ ध	Zigzag-like ङ	ल

TABLE 5.4: Study sets for Devanagari typefaces with characters grouped into attribute groups. The typefaces and styles specified correspond to those used to create character samples in the studies. Typefaces marked with an asterisk were used in combined studies.

Typeface & style name	Main attribute	Other attributes		Additional
Arial Regular	Halved asz	Narrow lt	Rounded ep	y
Century Schoolbook Regular	Narrow jir	Rounded opd	Diagonal vx	
Courier New Regular	Rounded ceo	Ascending fhk	Diagonal wy	
Georgia Regular	Diagonal kvy	Narrow ijl	Semirounded bd	
PT Serif Regular*	Arched hmn	Semirounded gq	Diagonal yz	r
PT Sans Regular*	Semirounded bpq	Narrow fi	Diagonal wz	n
Times New Roman Regular	Ascending lft	Arched amn	Arched cu	
Verdana Regular	Top-opened uwx	Arched gos	Ascending dh	
Calibri Regular	Narrow jir	Rounded opd	Diagonal vx	
Cambria Regular	Semirounded bpq	Narrow fi	Arched mn	z
Candara Regular	Diagonal kvy	Narrow ijl	Semirounded bd	
Futura Medium	Arched hmn	Semirounded gq	Diagonal yz	r

TABLE 5.5: Study sets for Latin typefaces with characters grouped into attribute groups. The typefaces and styles specified correspond to those used to create character samples in the studies. Typefaces marked with an asterisk were used in combined studies.

PT Serif	абвгдеёжзийклмнопрстуфхцчшщъыьэюя
PT Sans	абвгдеёжзийклмнопрстуфхцчшщъыьэюя
Arial	абвгдеёжзийклмнопрстуфхцчшщъыьэюя
Century Schoolbook	абвгдеёжзийклмнопрстуфхцчшщъыьэюя
Courier New	абвгдеёжзийклмнопрстуфхцчшщъыьэюя
Georgia	абвгдеёжзийклмнопрстуфхцчшщъыьэюя
Times New Roman	абвгдеёжзийклмнопрстуфхцчшщъыьэюя
Verdana	абвгдеёжзийклмнопрстуфхцчшщъыьэюя

TABLE 5.6: Study sets (in black) for all typefaces used in Cyrillic studies. The black and grey characters correspond to characters studied as they were defined in section 3.4.

ITF Devanagari	अइउएकखगघङचछजझञटठडढणतथदधनपफबभमयरलवशषसह
Kohinoor Devanagari	अइउएकखगघङचछजझञटठडढणतथदधनपफबभमयरलवशषसह
Adobe Devanagari	अइउएकखगघङचछजझञटठडढणतथदधनपफबभमयरलवशषसह
Devanagari MT	अइउएकखगघङचछजझञटठडढणतथदधनपफबभमयरलवशषसह
Ek Mukta	अइउएकखगघङचछजझञटठडढणतथदधनपफबभमयरलवशषसह
Lohit Devanagari	अइउएकखगघङचछजझञटठडढणतथदधनपफबभमयरलवशषसह
Murty Hindi	अइउएकखगघङचछजझञटठडढणतथदधनपफबभमयरलवशषसह
Nirmala UI	अइउएकखगघङचछजझञटठडढणतथदधनपफबभमयरलवशषसह

TABLE 5.7: Study sets (in black) for all typefaces used in Devanagari studies. The black and grey characters correspond to characters studied as they were defined in section 3.4.

PT Serif	abcdefghijklmnopqrstuvwxyz
PT Sans	abcdefghijklmnopqrstuvwxyz
Arial	abcdefghijklmnopqrstuvwxyz
Century Schoolbook	abcdefghijklmnopqrstuvwxyz
Courier New	abcdefghijklmnopqrstuvwxyz
Georgia	abcdefghijklmnopqrstuvwxyz
Times New Roman	abcdefghijklmnopqrstuvwxyz
Verdana	abcdefghijklmnopqrstuvwxyz
Calibri	abcdefghijklmnopqrstuvwxyz
Cambria	abcdefghijklmnopqrstuvwxyz
Candara	abcdefghijklmnopqrstuvwxyz
Futura	abcdefghijklmnopqrstuvwxyz

TABLE 5.8: Study sets (in black) for all typefaces used in Latin studies. The black and grey characters correspond to characters studied as they were defined in section 3.4.

5.4.3. Kinds of triplets included

The collections of triplets generated from study sets following the previously specified method are likely to contain at least one of each of these situations regarding the similarity relationships within the triplet:

i. An all-similar triplet: all characters in the triplet belong to the same attribute group. It is likely that participants will find them all very similar and may find it difficult to establish criteria to decide on the odd one out.
ii. An all-different triplet: each character in the triplet belongs to a different attribute group. It is likely that the participants will find them all different from each other and may find it difficult to establish criteria to decide on the odd one out.
iii. A two-similar triplet: two characters belong to the same attribute group, the third character belongs to another attribute group. It is likely that participants will find it easy to establish criteria to decide on the odd one out as the third character will appear distinct.

These categories are not meant to be exhaustive of all possible situations, but they illustrate some of the challenging situations created by planned mixing of the attribute groups within the study sets.

5.5. Technical implementation

The studies were conducted online. The use of Amazon Mechanical Turk (AMT) (2018) service for crowdsourcing simple human tasks was considered for its convenience (Crump, McDonnell & Gureckis, 2013), but it did not provide enough control over the presentation and the service was unreliable at the time. In the end, a custom website with embedded Javascript program was used instead of AMT. The questionnaire was a typical website form. It was identical for all the studies. The triplet sequences were composed dynamically as they needed to be randomised for each participant, i.e. for each view of the website.

For each typeface, there was a predefined list of character triplets and character samples as vector images in SVG format. When the website loaded into a browser, it randomly selected a typeface, loaded a list of triplets for the typeface and created a randomised sequence of triplets for the testing using the corresponding images.

Participants' answers were submitted to a service for collecting data from online forms (Furious Collective, 2014; Getform, 2015). The responses were always recorded for a triplet in a lexical order, e.g. `a, b, c` not `b, a, c`. This way, the responses from different participants could be matched and compared regardless of the order of the characters. Collected data were downloaded from the services as spreadsheet documents for further processing.

Unfortunately, controlling the absolute physical dimension of the shapes on different devices was impossible using common, contemporary web technologies, hence the actual dimensions of shapes might have varied for different participants and their devices (see section 5.2.2). However, in an attempt to obtain at least some indication of the kind of devices the participants used to complete the study, the website collected this information through traffic analysis tools. This data was collected globally, independently of any information about participants and their responses, for technical reasons.

Having minimized the unwanted effects of screen resolution and small size (see section 5.2.2), the incomplete control over viewing conditions could be seen as a benefit. Collecting similarity judgements in diverse conditions achieves a more realistic representation of human perception of similarity than a study with fixed conditions would.

The statistics provided below were calculated using Python programming language (Python Software Foundation, 2018) with SciPy (Eric, Travis, Pearu & others, 2001), Pandas (McKinney, 2010), Scikit-learn (Pedregosa et al., 2011),

matplot (Hunter, 2007), and RPy2 (Gautier & rpy2 contributors, 2008) libraries. Fisher's exact tests have been calculated using the statistical package R (R Core team, 2013).

5.6. Pilot study

The study was piloted using 23 triplets of characters from a single typeface. The website was a prototype, somewhat different in design from the website finally used. There were six participants who were informally interviewed afterwards. It took them 2–3 minutes to work through all triplets. The pilot study showed that participants tend to agree on the more obvious triplet trials. They did not have a problem understanding the assignment and using the website. Some participants with design knowledge focused on fine details. Some thought the task was to spot a letter from a different typeface within each triplet. This ultimately led to rephrasing of the instructions to discourage participants from focusing on details.

5.7. Participants

5.7.1. Recruitment

The participants were invited by convenience sampling, social networks, and via websites dedicated to recruiting study participants. Three services were used: *Twitter* (2018), *Call For Participants* (Kruusimagi, Terrell & Ratzinger, 2018) and the *Reddit's* community *SampleSize* that is used to post calls for participants (Huffman & Ohanian, 2005). The only constraint was that the participants needed to be adults which was clearly stated in the introduction to the study. There was no financial incentive for taking part in the studies.

Paid Twitter advertising (Twitter, 2018) was used to target the call for participants to the right audience for the Cyrillic and Devanagari studies, i.e. to people who can speak the relevant languages or live in corresponding countries for each script. Participants were not prevented from taking part in studies for a script they cannot read; however, this was not encouraged. Note that the script fluency can be checked additionally using the data collected through the questionnaire form (see section 5.10).

5.7.2. Conversion rate and dropout

The combined studies were conducted online over the course of the year 2016 while the main portion of the studies, the simple studies, were conducted between May and September 2017. According to the website traffic analytics (Heap, 2013) used, 3428 people visited the studies homepage. Of these, 2506 started filling in the questionnaire, 1755 started the triplet sequence, and

1475 submitted the results. The breakdown of these numbers per script is in table 5.9. Note that the total numbers from the analytics do not correspond to the actual number of participants who successfully submitted their results as some of them may have their browser security settings set to prevent such tracking. Thus, the following numbers have to be considered estimates. The conversion rate, or the portion of participants who completed the study to those who viewed the website, is 43.03%. An estimated 15.95% of participants dropped out of the study after they started the triplet sequence part.

Step	Cyrillic	Devanagari	Latin	All
Viewed study	1133	1254	1041	3428
Started questionnaire	852	794	860	2506
Started triplet sequence	580	478	697	1755
Finished	475	389	611	1475
Conversion rate	41.92%	31.02%	58.69%	43.03%
Triplet sequence dropout	18.10%	18.62%	12.34%	15.95%

TABLE 5.9: The website traffic analytics showing the numbers of tracked participants for particular steps of the study together with conversion rates and dropouts in percents. Counts and percentages are shown for each script and for all scripts collectively (column All).

5.7.3. Devices used

Device types used by participants as reported by the website traffic analytics are shown in table 5.10. Again, the numbers are estimates for technical reasons.

Device type	Cyrillic	Devanagari	Latin	All
Desktop	400 (43.15%)	492 (54.01%)	489 (51.04%)	49.39%
Mobile	360 (38.83%)	399 (43.80%)	403 (42.07%)	41.56%
Tablet	160 (17.26%)	15 (1.65%)	63 (6.58%)	8.51%
Unknown	7 (0.76%)	5 (0.55%)	3 (0.31%)	0.54%

TABLE 5.10: The website traffic analytics depending on the type of device tracked participants used. Counts and percentages are shown for each script and for all scripts collectively (column All).

5.7.4. Comments from participants

Some participants expressed their personal experience with the studies either on social networks or through an email. Some participants on social networks did not appreciate the demographic aspects of the questionnaire and considered them too intrusive (this could explain the dropout after viewing the questionnaire, see table 5.9).

One participant struggled with the missing definition of similarity to the point where they had to give up the study as they considered themselves not informed enough to provide useful responses. Other participants keenly observed that they had changed strategies for their decisions depending on a particular triplet.

As expected, some noted that certain triplets were hard to resolve and caused fatigue and insecurity about what the right response was. Participants also felt that some combinations of characters occurred more than once. This is perhaps the effect of triplets having very similar characters in them. For example, the triplets `h, o, m` vs. `h, o, n`, when seen with a few triplets in between them, can create the impression of seeing the identical triplet twice.

5.8. Pre-processing

Participants' responses required some pre-processing after collection to make the data set easier to use in subsequent analyses.

Firstly, combined-study sessions that contained two triplet sequences were split into two sessions with one sequence each, both completed by the same participant. Sessions with the first triplet sequence were included among the simple-study sessions. The sessions with the second triplet sequence were included as well. They were marked as second so they could be analysed separately later. For example, one could analyse the influence the order of the two typefaces could have had on similarity judgements.

Secondly, conducting the studies in three different languages also necessitated additional, semi-manual processing of the responses to the first two questions about language skills where the participants could enter free-form text. The language names had to be translated from the language participants used to their standard English names, their spelling corrected, and normalized. Naturally, participants using the Russian version of the study entered the languages in Cyrillic in Russian, sometimes in other languages too, e.g. in Ukrainian, and similarly for the Hindi version of the Devanagari study. Some participants even entered the names of the languages in the languages themselves, e.g. Telugu in the Telugu language and script. All versions of the language names were normalized to language names as used in Ethnologue (2018) and ISO 639-3 (ISO, 2007); e.g. Sambalpuria was converted to Sambalpuri, Marwadi to Marwari, etc. Where participants differentiated between speaking Chinese Cantonese or Mandarin, or reading simplified as opposed to traditional Chinese, this detail was preserved. Some participants used the name of a script instead of a language, or vague terms such as "bahasa", "creole", "hokkien", or "frisian". These were also preserved in the processed data.

For further analysis, it is useful to identify two groups of participants. The first, *native readers*, are participants who considered themselves native in a language associated with the script studied. An assumption is being made that those participants were also literate in the script associated with the language. Given the studies were conducted online and with the script explicitly mentioned during the call for participants, this is a safe assumption. The second group, *fluent readers*, is a group of participants who considered themselves fluent in reading a language associated with the script studied.

The association between the scripts and languages was dealt with using sets of native languages and scripts, which were mapped to each script based on whether the script is commonly used to write the language (see table 5.11). The script names were also included in these sets, as participants mixed them into their responses, too. Note that only languages and scripts that participants actually mentioned in the questionnaire were included, and that some languages were included for more scripts as they can be written in multiple scripts. Additionally, other scripts and languages were included in this table to evidence the diversity of the participants.

If participants mentioned one of the language names from table 5.11 in the questionnaire field “What is your native language?” (see section 5.3.1 and figure 5.6), they were identified as native readers of the associated script from the table. Accordingly, if they mentioned one of the language names in the field “Which languages can you read fluently?”, they were identified as fluent readers of the associated script from the table.

Lastly, one of the participants left both fields empty. Therefore, their responses could not be considered reliable and were removed from the data prior to further processing.

5.9. Data set overview

The total 1721 participants consist of 509 participants in the Cyrillic studies, 432 participants in the Devanagari studies, and 780 participants in the Latin studies. There were 1787 sessions with 56 triplet trials with character shapes produced with a single typeface. This corresponds to total 100,072 trial responses made by all participants. See table 5.13 for the total numbers of participants for each typeface. Unless stated otherwise, for this analysis, the combined studies are considered as two simple studies. Therefore, the total number of sessions is higher than the total number of participants.

Script	Associated languages
Cyrillic	Bashkir, Belarusian, Bosnian, Bulgarian, Kazakh, Macedonian, Russian, Serbian, Tatar, Ukrainian, Uzbek, Yakut
Devanagari	Bhojpuri, Devanagari, Garhwali, Hindi, Kashmiri, Konkani, Maithili, Marathi, Marwari, Nepali, Newari, Rajasthani, Sanskrit, Sindhi
Latin	Afrikaans, Bahasa, Bosnian, Catalan, Creole, Croatian, Czech, Danish, Dutch, English, Estonian, Faroese, Finnish, French, Frisian, Galician, German, German (Swiss), Hungarian, Icelandic, Indonesian, Irish, Italian, Latin, Latvian, Lithuanian, Malay, Norwegian, Polish, Portuguese, Portuguese (Brazilian), Scottish Gaelic, Serbian, Slovak, Slovenian, Spanish, Spanish (Mexican), Swedish, Tatar, Turkish, Uzbek, Vietnamese, Welsh
Other	Arabic, Armenian, Assamese, Bengali, Chinese, Chinese (Cantonese), Chinese (Mandarin), Chinese (Simplified), Georgian, Greek (Classical), Greek (Modern), Gujarati, Hebrew, Hokkien, Japanese, Javanese, Kachchi, Kannada, Korean, Malayalam, Oriya, Panjabi/Gurmukhi, Persian, Sambalpuri, Syriac, Tagalog, Tamil, Telugu, Urdu

TABLE 5.11: Mapping between the scripts and languages based on whether the script is commonly used to write the language. Other scripts and languages that participants filled in are included in the last row.

5.10. Participants' background data

In statistical terms, the participants constitute a population sample. This sample may be used to make inferences about the whole population. An overview of participants' backgrounds, provided below, helps to assess their diversity and representativeness in this regard. In this context, the whole population refers to all users of a particular script with varying levels of fluency.

The reports and charts are based on participants' self-assessments in the study questionnaires. For the sake of simplicity, they are broken down by script, i.e. the data from studies with different typefaces for the same script are kept together. Given that randomization procedures were put in place, and that all typefaces are relatively familiar and readable, it seems reasonable to aggregate the responses in this way.

The participants came from diverse linguistic backgrounds. They reported 96 languages, out of which they considered 79 to be native languages, and 88 languages they could read fluently. The proportion of native vs. non-native and fluent vs. not-fluent readers for each script studied is shown in table 5.12. Note that there is a significant proportion of non-native participants in the Devanagari studies. Part of the reason may be that there were Indian participants who cannot be considered either native or fluent in Devanagari according to the mapping described in section 5.8, although it is likely they are quite familiar with seeing Devanagari in the Indian visual environment.

Script	Participants	Native	Non-native	Fluent	Non-fluent
Cyrillic	509	469	40	479	30
Devanagari	432	259	173	356	76
Latin	780	739	41	778	2

TABLE 5.12: The total numbers of participants, native/non-native readers, and fluent/non-fluent readers.

The variety of languages for each script is shown in table 5.14. Notably, many of the participants were fluent in English, which is to be expected from a study promoted on social networks where English is the lingua franca.

Script	Typeface	Participants	SS Sessions	CS Sessions
Cyrillic	Arial	69	69	
	Century Schoolbook	62	62	
	Courier New	67	67	
	Georgia	75	75	
	PT Sans	53	53	9
	PT Serif	50	50	10
	Times New Roman	73	73	
	Verdana	60	60	
Devanagari	Adobe Devanagari	50	50	
	Devanagari MT	55	55	
	Ek Mukta	58	58	
	ITF Devanagari	48	48	9
	Kohinoor Devanagari	45	45	12
	Lohit Devanagari	63	63	
	Murty Hindi	62	62	
	Nirmala UI	51	51	
Latin	Arial	67	67	
	Calibri	77	77	
	Cambria	74	74	
	Candara	61	61	
	Century Schoolbook	59	59	
	Courier New	85	85	
	Futura	67	67	
	Georgia	64	64	
	PT Sans	50	50	15
	PT Serif	52	52	11
	Times New Roman	61	61	
	Verdana	63	63	

TABLE 5.13: The total numbers of participants, sessions from the simple studies (SS Sessions), and secondary sessions from the combined studies (CS Sessions, see sections 5.3 and 5.4) for each typeface studied.

Script	Language	Native readers	Fluent readers
Cyrillic	Belarusian	5.30 %	9.40 %
	English	3.90 %	63.50 %
	Russian	83.70 %	89.60 %
	Ukrainian	16.30 %	28.10 %
	Other languages	7.70 %	33.00 %
Devanagari	English	14.40 %	94.40 %
	Gujarati	4.90 %	8.10 %
	Hindi	41.20 %	78.00 %
	Kannada	4.20 %	3.20 %
	Malayalam	4.20 %	3.50 %
	Marathi	13.00 %	17.80 %
	Nepali	5.10 %	5.60 %
	Tamil	7.20 %	6.00 %
	Telugu	5.10 %	4.60 %
	Other languages	17.80 %	42.60 %
Latin	Czech	6.30 %	6.40 %
	Dutch	10.10 %	11.50 %
	English	41.30 %	95.80 %
	French	7.10 %	18.10 %
	German	15.60 %	22.40 %
	Italian	1.50 %	4.20 %
	Polish	2.40 %	2.20 %
	Portuguese	2.80 %	4.60 %
	Spanish	5.00 %	10.80 %
	Other languages	15.50 %	28.10 %

TABLE 5.14: Languages with at least ten native readers for studies for each script. The percentages provided show the total number of participants who considered themselves either native or fluent readers. Languages with fewer than ten native readers-participants are included collectively as “Other languages”.

There is a clear tendency towards younger participants in the studies (see figure 5.9), although this is less marked for the Devanagari studies. Again, this is possibly an effect of the study being conducted online. Most participants considered themselves competent readers, reading daily or often (see figure 5.10 for the distribution of participants reading habits). And most participants were non-designers (see figure 5.11), except for the Latin-script studies, where the call for participants started trending among designers which led many of them to participate.

The analyses in section 5.15 will compare native vs. non-native readers and experts vs. non-experts to assess whether participants from these different groups responded differently.

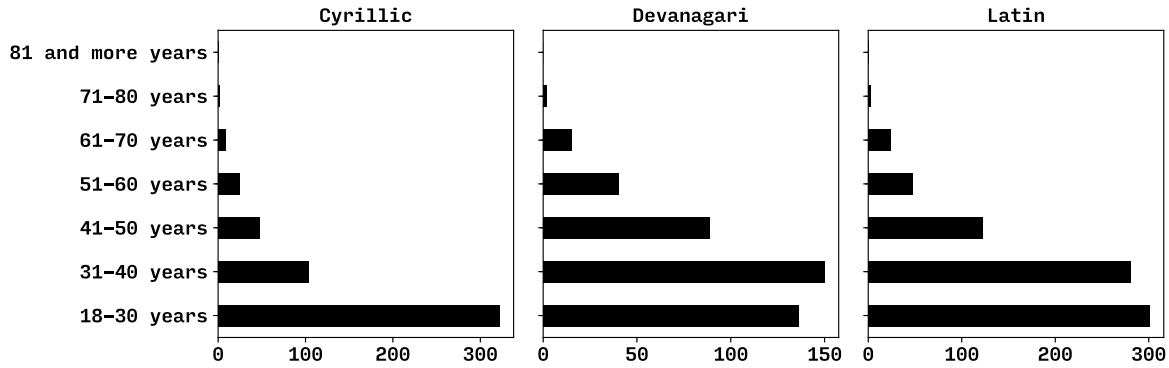


FIGURE 5.9: Charts showing age bands for participants in the Cyrillic, Devanagari, and Latin studies.

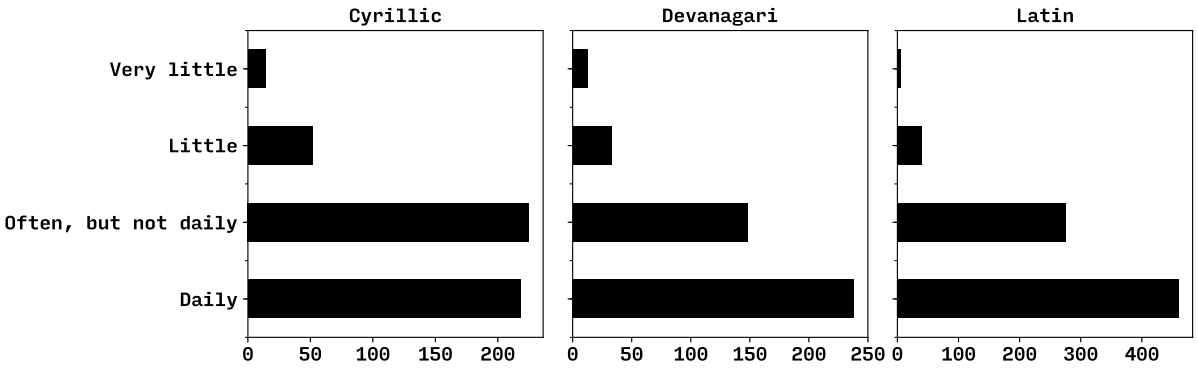


FIGURE 5.10: Charts showing how often participants read in the Cyrillic, Devanagari, and Latin studies.

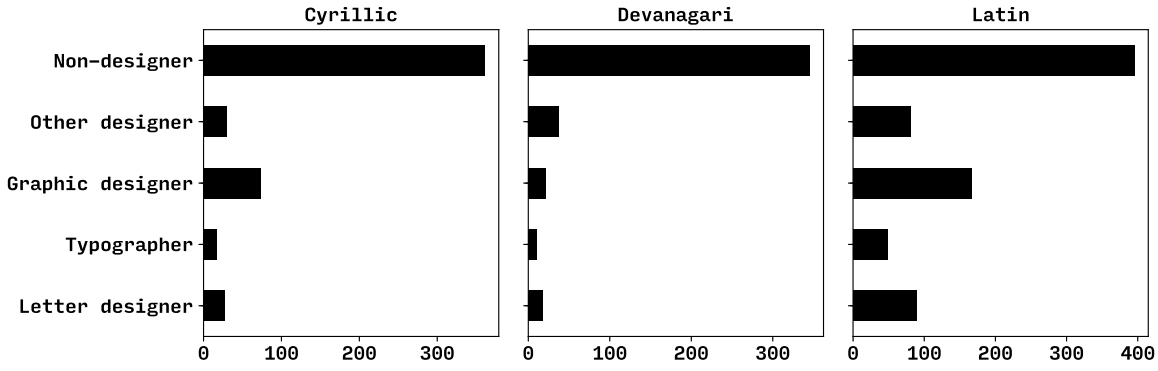


FIGURE 5.11: Charts showing what kinds of designers participated in the Cyrillic, Devanagari, and Latin studies. The majority of participants were non-designers.

5.11. Response data

5.11.1. Density of the data

Having reduced the scope of all combinations (see section 5.4.2 above), it is worth assessing the extent to which the selected triplets, and the pairs they contain, cover the whole spectrum of combinations, and how much data was collected with respect to triplets and pairs of characters. The counts for triplets and pairs that were included for each script are presented in table 5.15 (unique) and 5.16 (recurring). The tables show that some triplets and many of the pairs occurred multiple times in different studies, which provides an opportunity to compare the results among different typefaces even though this is not the main objective of the study. The higher rate of repetitions is caused by the additional studies in Latin, which shared the study sets with the initial Latin studies (see section 5.4.1).

Note that two triplets presented in two different typefaces are considered as two separate triplets in all the calculations below.

Script	Unique triplets (% of all)	Unique pairs (% of all)
Cyrillic	446 (8.17 %)	210 (39.77 %)
Devanagari	443 (5.70 %)	209 (31.38 %)
Latin	457 (17.58 %)	196 (60.31 %)

TABLE 5.15: Numbers of unique triplets and pairs studied per script regardless of which typeface they were presented in, and the proportion of all possible combinations expressed as percentage.

Script	Recurring triplets	Recurring pairs
Cyrillic	448	1344
Devanagari	448	1344
Latin	672	2016

TABLE 5.16: Numbers of recurring triplets and pairs studied per script, regardless of which typeface they were presented in.

There are 5456 triplet combinations and 528 pair combinations that can be made from 33 characters (Cyrillic), 7770 triplet combinations and 666 pair combinations from 37 characters (Devanagari), and 2600 combinations and 325 pair combinations from 26 characters (Latin). Clearly, the triplets included in the studies (see table 5.15) are only a small portion of all the combinations possible: 8.17% for Cyrillic, 5.7% for Devanagari, and 17.58% for Latin.

It is also worth looking at what pair combinations are represented, and how much data there is for unique pairs, in terms of the number of studies in which they appeared. Comparing unique triplets in table 5.15 with recurring triplets in table 5.16 shows that very few triplets repeat in the data for Cyrillic and Devanagari, while there are many more repetitions among the Latin triplets. Analogically, pairs repeat about six times on average in the Cyrillic and Devanagari data and about ten times on average in the Latin data. The dot plots in figure 5.12 show that the pairs included in the studies are distributed evenly among all of the potential combinations.

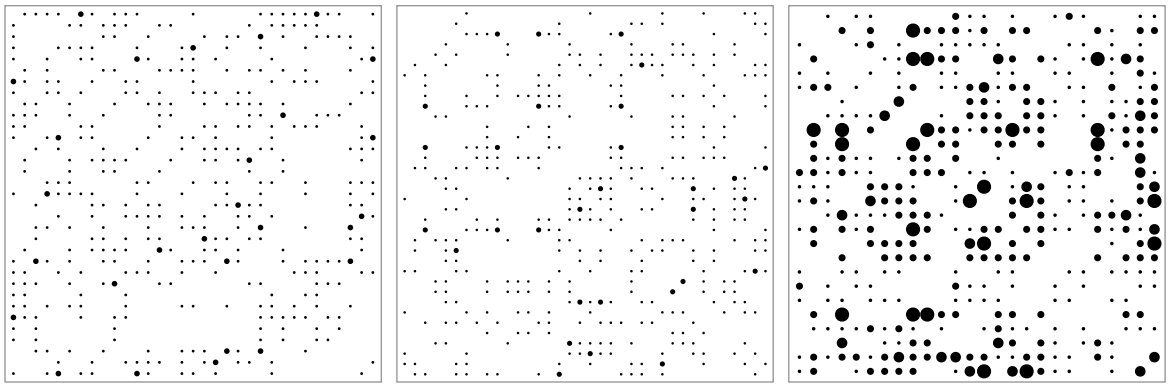


FIGURE 5.12: The plots for Cyrillic (left), Devanagari (middle), and Latin (right) follow the convention of the similarity matrices (see table 5.1), i.e. the plots have all characters studied from the respective script along the horizontal and vertical axes (not-shown) in the alphabetic/syllabic order. Each plotted dot marks whether their combination was represented in the studies. The size of the dot indicates how many times the combination recurred overall. In Cyrillic and Devanagari, pairs were repeated either 6 or 12 times. In Latin, they were repeated 6, 12, 18, or 24 times.

5.11.2.Response frequencies

All responses to triplet trials can be aggregated across participants to obtain cumulative response frequencies as an indication of the overall opinion. The frequencies are normalized by the total number of responses for a particular combination of a triplet and a typeface to make them relative. This makes the frequencies comparable to each other even when the numbers of participants that responded to each trial differ. For example, if for a triplet `a, b, c`, 11 participants voted for `a`, 2 for `b`, and 15 for `c`, the cumulative frequencies would be $n_a = 11$, $n_b = 2$, $n_c = 15$ while the relative frequencies would be $f_a = 0.39$, $f_b = 0.07$, $f_c = 0.54$, or more succinctly $f_{abc} = (0.39, 0.07, 0.54)$. The relative frequency 1.0 corresponds to 100% agreement of all participants, 0.5 means that 50% of the participants selected a particular response, etc. For each trial, the character that gets most of the votes – the response with the highest frequency – is *the most popular response* or *the overall odd one out*.

It is important to note that majority does not mean a response with more than 50% of votes (frequency 0.5 or higher). In a triplet trial, the most popular response – response with majority of votes – might not get more than half of all votes. For example, this is the case when the relative frequencies for a triplet a, b, c are $f_{abc} = (0.4, 0.3, 0.3)$.

For the sake of clarity, for the rest of this work, cumulative frequencies are called *counts* and are not normalized, and all *frequencies* have been normalized, i.e. they are relative.

5.11.3. Participants' agreement and triplet difficulty

An important metric used to describe the data is participants' level of agreement on a particular triplet trial. Clearly, it would not make much sense to model human perception if people provided inconclusive or random responses; one could use a random generator instead. There needs to be a clear indication of some agreement among the participants to be able to purposefully model predictions as discussed in chapter 6.

It is useful to look at two extreme cases first. A triplet trial where participants agreed completely would have frequencies $f_{abc} = (0.0, 0.0, 1.0)$ for example, i.e. one response received 100% votes and the other two did not get any votes at all. On the other hand, a trial with the lowest agreement would have frequencies close to $f_{abc} = (0.33, 0.33, 0.33)$ where 0.33 is a periodic number (exactly one third), i.e. the three responses each received exactly one third of participants' votes. The frequencies for the most popular responses in these two extreme cases correspond to the theoretical maximum (1.0) and minimum (0.33) of participants' agreement, respectively.

As expected from the construction of triplet combinations and from participants' comments on the study, there were triplets of varying agreement: those where most participants agreed, and those where there was less or no consensus at all. An analysis of the distribution of the frequencies of the most popular responses from the trials can be used to gain insights about participants' agreement. Table 5.17 shows minima, quartiles, medians, and maxima of this distribution for each script and for all scripts together; figure 5.13 shows corresponding box plots.[9] Taking the overall participants' agreement on a particular triplet as an indication of the difficulty of the judgements, one can distinguish trials with different levels of difficulty. The lower the agreement among participants, the more challenging the similarity judgements for the trial were. And vice versa, the higher the agreement, the easier the judgements.

9. Box plots are a more transparent alternative to histograms. See figure 5.13 for their interpretation.

Script	Min	1st	Median	3rd	Max	STD
Cyrillic	0.35	0.57	0.69	0.82	0.99	0.15
Devanagari	0.34	0.50	0.61	0.74	0.94	0.15
Latin	0.34	0.57	0.74	0.87	1.00	0.18
All	0.34	0.54	0.68	0.82	1.00	0.17

TABLE 5.17: The minima (Min), quartiles (1st, 3rd), medians, maxima (Max) and standard deviations (STD) describing the distribution of the most popular response frequencies, for each individual script and for all scripts together.

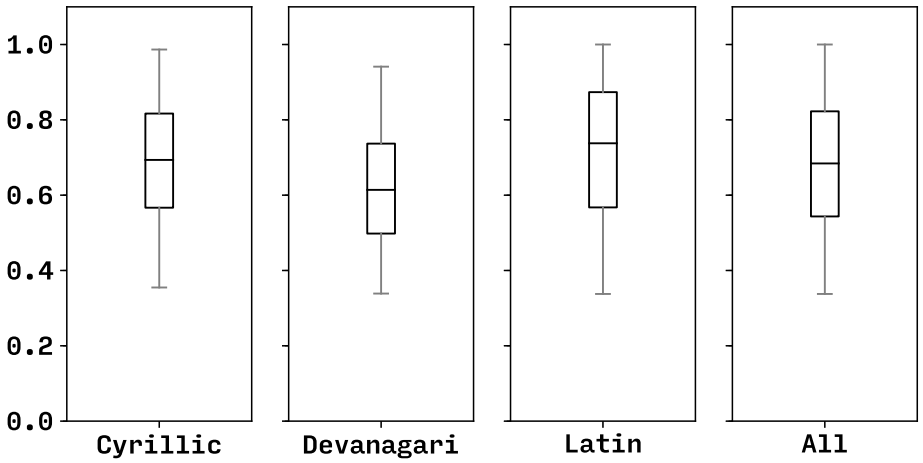


FIGURE 5.13: Box plots showing distribution of the most popular response frequencies for Cyrillic, Devanagari, Latin, and all studies. The boxes represent the two central quartiles, i.e. 50 % of the data, in each case. The central horizontal lines crossing the boxes represent the medians. The other two quartiles are represented by the two whiskers while their outside caps represent the minimum and maximum.

The bottom rows in tables 5.17 and 5.18 and the rightmost box plot in figure 5.13 describe the overall distribution for all the studies. The minimum is near 0.33 and the maximum is near 1.0. The median and mean are at 0.68, which is a little above the middle between the minimum and maximum frequencies in the plots. The even spread between the theoretical minimum and maximum means that the triplets included in the testing covered a wide spectrum of similarity relationships in terms of difficulty (as discussed in 5.4.3). Trials from the first quartile are called *difficult*; their frequency for the most popular response is below 0.54. Trials from the third quartile are called *easy*; their frequency for the most popular response is above 0.82.

Script	Mean	SEM
Cyrillic	0.69	0.007
Devanagari	0.62	0.007
Latin	0.72	0.007
All	0.68	0.004

TABLE 5.18: The means and standard errors of the means (SEM) of the most popular response frequencies – individually and for all scripts together.

The distributions have been tested for normality using a hypothesis test (D'Agostino & Pearson, 1973). The test's null hypothesis that the distribution of the most popular responses comes from a normal distribution could be rejected as the resulting p-value was smaller than $\alpha = 0.01$ for the scripts individually as well as all together. The histograms in figure 5.14 illustrate that the distributions depart from the normal distribution. Generally, the difficult trials are slightly less frequent than if the distribution was normal while the easy trials are slightly more frequent than if the distribution was normal.

Looking at all the figures for the individual scripts reveals that there was more agreement between the participants in the Latin and Cyrillic studies than in the Devanagari studies, where the mean is lower. There were more difficult trials in the Devanagari studies, while the Latin studies contained a larger quantity of easy trials. At the same time, there is a larger variety among the frequencies in the Latin study compared to Cyrillic and Devanagari studies, i.e. the spread between 1st and 3rd quartile in the relevant box plot is larger (see figure 5.13).

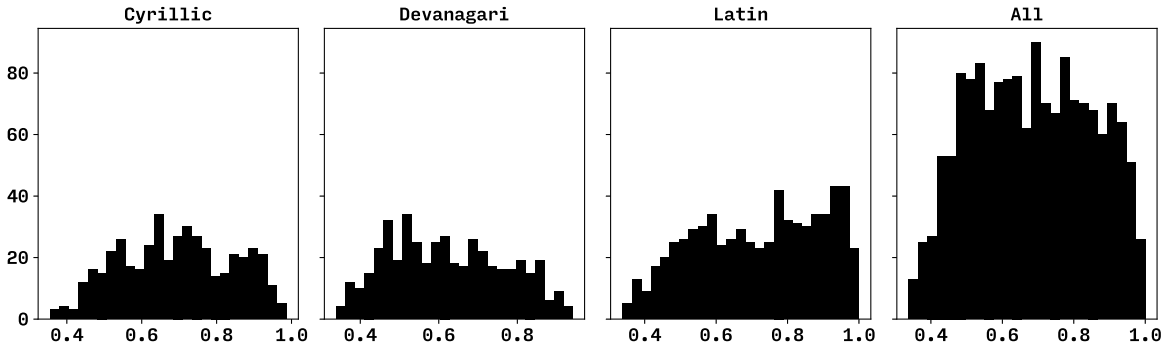


FIGURE 5.14: The histograms for the distribution of the most popular response frequencies – individually and for all scripts together. The quantity is plotted on the vertical axis, the frequency is on the horizontal axis.

5.12. Specific character pair similarity measure

The three response frequencies associated with each triplet can be interpreted as measures of similarity within triplets. As discussed in section 5.2, marking a character as the odd one out is a statement about the similarity of the other two characters, the complementary pair. The response frequency for a particular character can be taken as a measure of the specific similarity of the complementary pair in the context of this character (see section 5.2.1 for definition of specific character pair similarity). The specific similarity measure of the pair a, b in the context of the character c is written as: $s_c(a,b)$. This is assuming the same typeface is used and specified explicitly. For a trial with a triplet a, b, c, the response frequency for c is $s_c(a,b)$, the response frequency for a is $s_a(b,c)$, and the response frequency for b is $s_b(a,c)$. Tables 5.20–5.22 show response frequencies for selected triplets. See table 5.19 for an explanation of how to interpret the frequencies from these tables. The rest of the response frequencies are included in appendix 2.

It is important to note that when a participant makes the decision to vote for the odd one out, for example a in the triplet a, b, c, they make a direct statement about the complementary pair – i.e. b, c, and its similarity – but not about the other two pairs – i.e. a, b and a, c, and their relative similarities even though, presumably, their similarity was judged as part of the process of determining the odd one out.

character		count	freq.	confidence interval
e	Э	41	0.66	(0.53, 0.78)
ghe	Г	12	0.19	(0.1, 0.31)
ha	Х	9	0.15	(0.07, 0.26)

$s_e(ghe, ha)$

TABLE 5.19: The response frequency for a character from a triplet gives the specific similarity measure of the other two characters in the context of this character. The characters marked in magenta are the main characters; the character marked in green is contextual. In the character column, character names are on the left and small versions of the samples as they appeared in the studies are on the right. The overall odd one out is marked with a grey background.

character		count	freq.	confidence interval
ef	ф	7	0.10	(0.04, 0.2)
em	м	61	0.88	(0.78, 0.95)
er	р	1	0.01	(0.0, 0.08)

character		count	freq.	confidence interval
em	м	53	0.77	(0.65, 0.86)
er	р	13	0.19	(0.1, 0.3)
es	с	3	0.04	(0.01, 0.12)

character		count	freq.	confidence interval
a	а	48	0.77	(0.65, 0.87)
de	д	4	0.06	(0.02, 0.16)
ghe	г	10	0.16	(0.08, 0.28)

character		count	freq.	confidence interval
de	д	4	0.06	(0.02, 0.16)
ghe	г	5	0.08	(0.03, 0.18)
io	ё	53	0.85	(0.74, 0.93)

TABLE 5.20: Response counts, response frequencies, and confidence intervals for selected triplets from the Cyrillic studies with typefaces Arial (top) and Century Schoolbook (bottom).

character		count	freq.	confidence interval
A	अ	27	0.54	(0.39, 0.68)
Ba	ब	20	0.40	(0.26, 0.55)
Gha	घ	3	0.06	(0.01, 0.17)

character		count	freq.	confidence interval
Ba	ब	11	0.22	(0.12, 0.36)
Gha	घ	20	0.40	(0.26, 0.55)
Ka	ख	19	0.38	(0.25, 0.53)

character		count	freq.	confidence interval
Ca	च	25	0.45	(0.32, 0.59)
Cha	छ	21	0.38	(0.25, 0.52)
Dda	ड	9	0.16	(0.08, 0.29)

character		count	freq.	confidence interval
Cha	छ	6	0.11	(0.04, 0.22)
Dda	ड	42	0.76	(0.63, 0.87)
Dha	ध	7	0.13	(0.05, 0.24)

TABLE 5.21: Response counts, response frequencies, and confidence intervals for selected triplets from the Devanagari studies with typefaces Adobe Devanagari (top) and Devanagari MT (bottom).

character		count	freq.	confidence interval
a	a	1	0.01	(0.0, 0.08)
e	e	1	0.01	(0.0, 0.08)
l	l	65	0.97	(0.9, 1.0)

character		count	freq.	confidence interval
e	e	11	0.16	(0.08, 0.27)
l	l	51	0.76	(0.64, 0.86)
p	p	5	0.07	(0.02, 0.17)

character		count	freq.	confidence interval
d	d	14	0.24	(0.14, 0.37)
i	i	9	0.15	(0.07, 0.27)
x	x	36	0.61	(0.47, 0.73)

character		count	freq.	confidence interval
j	j	21	0.36	(0.24, 0.49)
o	o	15	0.25	(0.15, 0.38)
x	x	23	0.39	(0.27, 0.53)

TABLE 5.22: Response counts, response frequencies, and confidence intervals for selected triplets from the Latin studies with typefaces Arial (top) and Century Schoolbook (bottom).

In order to provide an indication of similarity judgements the general population would make, there are binomial confidence intervals provided for each response (see tables 5.20–5.22). In statistical terms, there is a 95% chance that each of these intervals contains the response frequency for a particular character in a particular triplet in a particular typeface. In other words, there is a 95% chance that these intervals contain the specific similarity measure of the complementary pair in the context of the corresponding character.

The binomial confidence intervals provided in the tables are calculated using Clopper-Pearson interval method (Clopper & Pearson, 1934) that is based on Beta distribution. The method is exact, not an approximation; therefore, it works with a small number of participants. Moreover, it is relatively conservative compared to other methods, i.e. the provided intervals are sometimes wider than they have to be.

5.13. Generalized character pair similarity

The response frequencies for each typeface study can be aggregated into similarity matrices that are typically used in cognitive psychology (see table 5.1). As discussed in section 5.12, the response frequencies for a particular character can be taken as a specific similarity measure of the complementary pair. As each pair appears in multiple triplets, the similarity of a particular pair can be generalized to some extent. The *generalized pair similarity measure* (see section 5.2.1 for a definition of generalized character pair similarity) is calculated as a mean of the specific pair similarity measures for the particular pair in each triplet where the pair occurs, regardless of the contextual character.

It is important to note that although the contextual characters are disregarded during the process, the generalized similarity measures should still be considered to have a context made of all contextual characters from the aggregated triplets, i.e. all the other characters from the study set for a particular typeface.

This limited scope should be always kept in mind, as the measures could have a misleading effect when generalized. For instance, when studying the similarity of `v, z` in the context of `o`, `p`, or `b`, the pair may seem more similar than when studied in the context of `w`, `x`, or `y`. On the other hand, although incomplete, the characters included in the study sets for each typeface were selected for the sets to be diverse. Thus, the measures can be used to identify general tendencies.

Note that a pair similarity measure that would be completely general could only be calculated if a particular script had a fixed number of characters and all combinations of the pair with each of the scripts' characters were tested in the studies.

The generalized similarity measures for the typefaces studied are presented in 8×8 similarity half-matrices with all eight characters from each typeface study on the axes. The axes are the same for the vertical and horizontal axes and include the names of all characters as well as their rendering in the corresponding typeface. The value on the intersection for any two characters is their generalized character pair similarity measure. The values for pairs `a, b` and `b, a` are identical following the general assumption that the order of pairs and triplets do not matter; hence, half of a complete matrix is sufficient. The diagonal entries are missing in all the matrices, as pairs of two identical characters never occurred in the studies. Example half-matrices for each script are in tables 5.23–5.25. The rest of the typeface similarity half-matrices can be found in appendix 3.

a	а	-						
ghe	г	0.16	-					
de	д	0.20	0.66	-				
io	ё	0.56	0.08	0.08	-			
zhe	ж	0.25	0.18	0.21	0.15	-		
ka	к	0.27	0.34	0.29	0.10	0.80	-	
pe	п	0.12	0.83	0.79	0.08	0.20	0.32	-
ya	я	0.65	0.21	0.26	0.25	0.49	0.60	0.20
characters		а	г	д	ё	ж	к	п	я
		a	ghe	de	io	zhe	ka	pe	ya

TABLE 5.23: Character pair similarity matrix based on the Latin study with the typeface Century Schoolbook.

A	अ	-						
Ka	ख	0.24	-					
Gha	घ	0.35	0.28	-				
Ta	त	0.18	0.24	0.32	-			
Ba	ब	0.15	0.29	0.40	0.34	-		
Ra	र	0.21	0.51	0.22	0.50	0.16	-	
Ssa	ष	0.14	0.18	0.48	0.33	0.83	0.15	-
Sa	स	0.36	0.66	0.32	0.42	0.20	0.55	0.31
characters		अ	ख	घ	त	ब	र	ष	स
		A	Ka	Gha	Ta	Ba	Ra	Ssa	Sa

TABLE 5.24: Character pair similarity matrix based on the Devanagari study with the typeface Adobe Devanagari.

a	a	-						
e	e	0.89	-					
l	l	0.04	0.06	-				
p	p	0.47	0.48	0.21	-			
s	s	0.71	0.71	0.11	0.27	-		
t	t	0.09	0.08	0.79	0.21	0.08	-	
y	y	0.18	0.19	0.24	0.59	0.18	0.40	-
z	z	0.35	0.37	0.25	0.17	0.65	0.18	0.40
characters		a	e	l	p	s	t	y
		a	e	l	p	s	t	y

TABLE 5.25: Character pair similarity matrix based on the Latin study with the typeface Arial.

The similarity half-matrices for a particular script could be obtained by aggregating all the specific similarity measures across all typefaces for the script in the same way as it was done for typefaces. These matrices would have all characters studied on the axes' indices and they would be incomplete in the sense that some character pairs did not appear in any of the studies. Nonetheless, such matrices are not presented here as they would be misleading. The individual generalized pair similarity measures would have had different contexts, depending on the typeface studies the pairs appeared in. On the other hand, the graphical or tabular representation would suggest that the context is identical for all the pairs and corresponds to all characters studied.

The character pair similarity measures can be visualized as geometrical distances in a two-dimensional plane. Firstly, each similarity half-matrix is turned into a complete symmetrical matrix and then converted to a distance matrix by calculating the difference between its values and 1.0. Secondly, the missing diagonal values are set to 0.0 to represent that two identical characters do not have any distance between themselves. The values in the distance matrix correspond to distances between the characters in the row and column indices; the smaller the value the closer the two characters are. An eight-dimensional geometric space is then constructed; each dimension corresponding to a distance from a particular character. Characters are placed within this space by taking each row from the distance matrix as their coordinates in this space.

The eight-dimensional space is then reduced to a two-dimensional space using a method called locally linear embedding (LLE) (Roweis & Saul, 2000) or specifically its modified variant (Zhang & Wang, 2007).[10] The characters' shapes are plotted at their new two-dimensional coordinates to produce the cluster plots in figures 5.15–5.17. The rest of the cluster plots can be found in appendix 4. Note, however, that the dimensional reduction is a drastic simplification of the similarity relationships and the resulting plots should be viewed with caution.

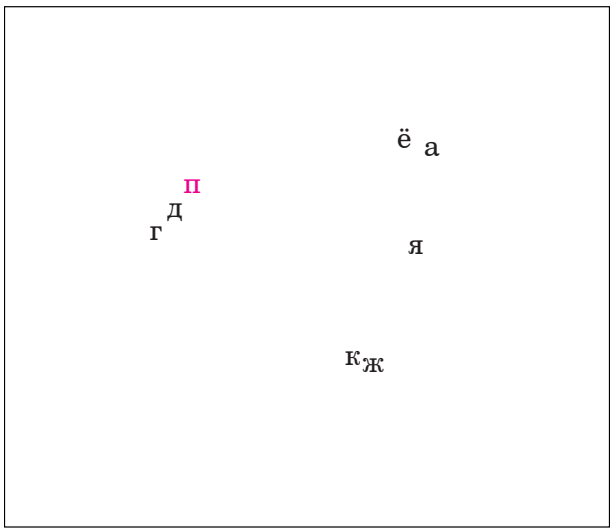


FIGURE 5.15: Plot with character clusters based on the generalized pair similarity measures derived from the Cyrillic study with the typeface Century Schoolbook. The proximity of the characters corresponds to their similarity, simplified by the projection. For transliteration, see the corresponding similarity matrix in table 5.23. The character in magenta has been shifted slightly from its default position to avoid its collision with other characters in the plot.

10. One could also use multidimensional scaling (MDS) as discussed in section 6.1 or other techniques, but LLE seemed best at keeping character pairs with higher similarity measures closer together thus providing better explanatory value.

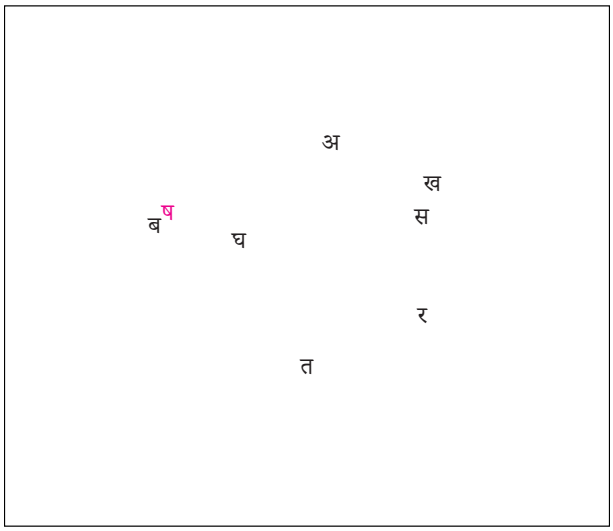


FIGURE 5.16: Plot with character clusters based on the generalized pair similarity measures derived from the Devanagari study with the typeface Adobe Devanagari. The proximity of the characters corresponds to their similarity, simplified by the projection. For transliteration, see the corresponding similarity matrix in table 5.24. The character in magenta has been shifted slightly from its default position to avoid its collision with other characters in the plot.

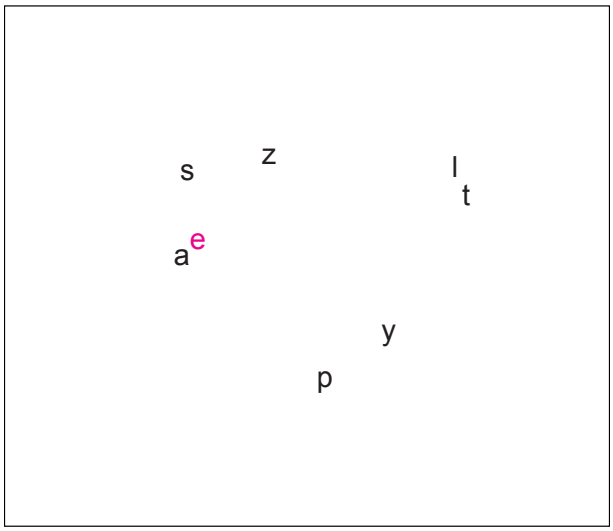


FIGURE 5.17: Plot with character clusters based on the generalized pair similarity measures derived from the Latin study with the typeface Arial. The proximity of the characters corresponds to their similarity, simplified by the projection. The character in magenta has been shifted slightly from its default position to avoid its collision with other characters in the plot.

The distance matrix can be also used to produce dendrograms such as those in figure 5.18. However, dendrograms are an even greater simplification of the similarity relationship than cluster plots. This can be easily illustrated by comparing the plot and dendrogram visualizing the similarity of the Latin characters from the typeface Arial (compare the plot in figure 5.17 with the left dendrogram in figure 5.18). In the plot, the character ‘p’ is just about as close to ‘e’ as it is to ‘t’ for example. One would expect that as ‘e, p’ exhibit a certain roundness in their shapes and ‘p, t’ have a straight vertical component. On the other hand, the ‘p’ is much closer to ‘t’ than it is to ‘e’ in the dendrogram. It has been grouped with other characters based on them having vertical components and the rounded characters have been grouped in another section of the dendrogram.

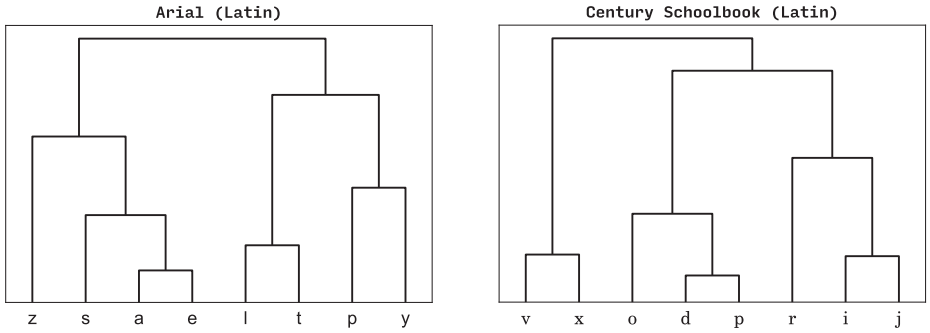


FIGURE 5.18: Dendrograms based on the generalized pair similarity measures derived from the Latin studies with typefaces Arial (left) and Century Schoolbook (right). Pairs of characters are connected with branches. The shorter the length of the branch, the greater the generalized character pair similarity. For example, in the left dendrogram, ‘a’ and ‘e’ are considered most similar.

5.14. Character distinctiveness and perceivability

If the data can be aggregated pair-wise, they can be also aggregated character-wise. This time, each response frequency for a particular character gets averaged across all triplets within a single typeface (see table 5.26 and appendix 5 for the complete set for all the typefaces). The resulting arithmetic mean measures how often the character is considered the odd one out and can be used loosely as a measure of the character’s *distinctiveness* from all other characters in the study set. Again, the result is relevant only in the context of the characters from a particular typeface study. The reason for not presenting these aggregations for a particular script is the same as described in the section 5.13.

For studies where the viewing conditions are comfortable with no additional challenges such as small size, filtering, or brief exposure, the distinctiveness measure may be related to *perceivability* as defined by Mueller and Weidemann (2012): “a theoretical construct affecting the probability the observer forms a veridical percept from the stimulus, independent of response factors”.

character		score
zhe	ж	0.44
short i	й	0.46
em	м	0.31
er	р	0.21
es	с	0.30
u	у	0.38
ef	ф	0.31
yu	ю	0.26

character		score
A	अ	0.54
Ka	ख	0.31
Gha	घ	0.32
Ta	त	0.33
Ba	ब	0.32
Ra	र	0.34
Ssa	ष	0.31
Sa	स	0.19

character		score
a	a	0.22
e	e	0.20
l	l	0.51
p	p	0.32
s	s	0.25
t	t	0.48
y	y	0.38
z	z	0.32

TABLE 5.26: Character distinctiveness measures for (from left to right) the Cyrillic study with typeface Arial, the Devanagari study with typeface Adobe Devanagari, and the Latin study with typeface Arial.

5.15. Comparison of responses

5.15.1. Comparison of repeated triplets between typefaces

The main focus of this work is on studying coherence and similarity perception in different scripts, hence it is also of interest whether finer differences in typefaces’ designs have an influence on similarity perception and whether the approach to the studies is sensitive enough to detect these influences. However, a detailed exploration of the design’s effect on similarity was not the main objective of the studies. The typefaces studied were selected to provide an unbiased testing ground, not to challenge the similarity judgements by their design. However, as mentioned above, some triplets recurred in two or more typeface studies (in section 5.11.1). This allows comparison of participants’ responses to these triplets represented in different typefaces.

The testing is done on individual triplets. The response counts for the recurring triplets in two trials, each with a different typeface, are juxtaposed, with one set of counts for one typeface in one column and another set of counts for the other typeface in another column. The resulting contingency table is then compared using Fisher's exact test of independence (Fisher, 1922). Usually, Pearson's chi-squared test (Pearson, 1900) or G-tests (Sokal & Rohlf, 1994) would be used to assess how well two distributions fit. In this case, however, some of the response counts could be small or zero, thus the requirements for using these tests are not satisfied. Also, Barnard's test (Barnard, 1945), often used instead of Fisher's test, is not applicable here as the contingency tables for the comparison are larger than 2×2.

Crudely put, Fisher's exact test sees whether there is an association (contingency) between the two sets of counts and the fact that they are from two different typefaces. The null hypothesis of this test is that there is no significant difference between the first and the second set of counts, i.e. no association. The alternative hypothesis is that there is a significant difference between them, i.e. there is an association between responses and a typeface. Fisher's exact test is used to calculate a p-value. Where the p-value is smaller than $\alpha = 0.05$, the null hypothesis is rejected and the alternative hypothesis is accepted, as there is sufficient evidence that the difference between the two sets of counts is significant. Where the p-value is larger than $\alpha = 0.05$, the null hypothesis is not rejected as there is not sufficient evidence of a difference between the two sets of counts. In other words, the responses for a trial in one typeface could not be distinguished from the responses for a trial with the other typeface in this test.

Note however, that although it is possible to calculate the p-value, one still has to bear in mind that the number of participants in both sets should be large enough to be representative. This is not a problem when comparing different typefaces, since all participants' responses are used.

See table 5.27 for juxtaposed responses for some of the recurring triplets (the complete set is included in appendix 6). Where there is a significant difference in participants' responses, it is possible to hypothesize why this happened. For example, the responses for triplet 'g, q, y' in typefaces Futura and PT Serif differ significantly. For Futura, the most popular response was clearly 'y', possibly due to the fact that the construction of 'y' uses diagonals while the other two characters are rounded and thus more closely related. For PT Serif, the 'g' received a larger number of votes, possibly because it looked less related to 'q' due to its more complex, so-called double-storey construction (see table 5.27 for the actual rendering of the character shapes discussed). Notice that the different construction of 'g' did not have a strong effect in the triplet 'g, h, m'. Possibly because the pair 'h, m' looks strongly related in both of these typefaces.

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.79	0.89	g	g
h	h	0.04	0.05	h	h
m	m	0.16	0.06	m	m

p-value = 0.154 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.09	0.37	g	g
q	q	0.01	0.13	q	q
y	y	0.90	0.51	y	y

p-value < 0.001 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.00	0.00	b	b
d	d	0.02	0.00	d	d
i	i	0.98	1.00	i	i

p-value = 1.000 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
j	j	0.25	0.62	j	j
k	k	0.69	0.36	k	k
l	l	0.07	0.02	l	l

p-value < 0.001 (typeface has an effect)

TABLE 5.27: Juxtaposed response frequencies and comparison p-values for selected recurring triplets in typefaces Futura and PT Serif (top) and Candara and Georgia (bottom). The selections were made to show variation in responses. The complete set is included in appendix 6. For convenience, relative frequencies are shown instead of cumulative frequencies (counts) that were used to calculate the p-value using Fisher's exact test. The overall odd one out is marked with a grey background.

There are other, more subtle cases where the responses differ, e.g. triplet 'j, k, l' in Candara and Georgia where the two letters 'j' differ in the treatment of their descenders.

Some of the differences might have also been caused by the differences between the study sets. Seeing different characters in the triplet sequence prior to a particular trial could have cultivated different decision strategies.

Considering the total number of recurring triplets, the cases where the typeface had an effect (the p-value was below 0.05) were not very frequent. This is due to the fact that all the typefaces in the studies belong to a similar genre (typefaces for continuous reading) and are therefore quite similar. The pairs of typefaces with most differences are Candara and Georgia, and Futura and PT Serif, which are also typefaces from two distinct categories: sans-serif and serif, respectively, for each pair.

Naturally, these results cannot be used to make general conclusions about the differences between these typefaces, but they are evidence of the sensitivity of this form of testing to the subtleties of typeface designs, and its potential for making this kind of design comparison. Note that, although the comparisons in table 5.27 and appendix 6 juxtapose mostly serif and sans-serif typefaces, the intention was not to compare serif and sans-serif, specifically. These combinations emerged from the data.

5.15.2. Agreement between groups of participants

The sections 5.15.3 and 5.15.4 look at pairs of groups of participants and compare the distributions of their most popular response frequencies. In order to make these comparisons, there first needs to be some indication of an overall agreement between the groups on what constitutes the most popular response for each trial. It is insufficient to compare only their response frequencies. Two trials can have the same most popular response frequency, yet the actual responses may differ. For example, if for a triplet a, b, c one group responded with frequencies $f_{abc} = (0.8, 0.1, 0.1)$ and the other group responded with frequencies $f_{abc} = (0.1, 0.1, 0.8)$, the most popular response frequencies would be identical (0.8), but the responses would differ.

Typically, Fisher's exact test could be used to compare the trials' response frequencies as was done in section 5.15.1. Unfortunately, when divided into two groups, there is often an insufficient number of participants and thus there are only a few triplets with the number of participants large enough to be considered representative.

An alternative solution is to look at the agreement on the most popular responses and compare them directly for each trial. In other words, one can calculate how often the two groups of participants agree on the overall odd one out. This can be represented as a percentage describing the proportion of trials where both groups agreed and the total number of trials compared.

Nonetheless, the percentage does not account for participants providing the correct response by chance. Therefore, Cohen's kappa statistic (Cohen, 1960) is provided as well, since it addresses this issue. Cohen's kappa measures inter-rater agreement for series of items classified into mutually exclusive categories. The categories in this case indicate which of the three characters in the trial has been selected as the overall odd one out. The interpretation of Cohen's kappa suggested by a "commonly cited scale" (Viera & Garrett, 2005) is provided in table 5.28.

Value	Interpretation
< 0.0	Less then a chance agreement
0.01–0.20	Slight agreement
0.21–0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substantial agreement
0.81–0.99	Almost perfect agreement
1.00	Perfect agreement

TABLE 5.28: A common scale used to interpret Cohen's kappa statistic (adapted from: Viera and Garrett, 2005).

In order to maintain some representativeness of the results, the percentage and Cohen's kappa are reported only for trials that were completed by at least five participants from each of the compared groups.

5.15.3. Comparison between experts and non-experts

The third research question set out in section 1.1: "Does the coherence perceived by readers differ from the coherence intended by designers?" is related to the question of whether experts perceive similarity differently from non-experts.

Having obtained the responses together with participants' background information, it is possible to look at the relationship between the participants' design skills and their similarity judgements. There are two kinds of experts: general designers, e.g. graphic designers, typographers, and other designers; and letter designers[11], e.g. lettering artists, typeface designers, and calligraphers. The first category would typically use letters in their work with some consideration, but would not regularly produce characters. On the other hand, the latter produce characters regularly and thus may be more conscious of their shapes and relationships. Therefore, the comparisons below are for three categories: non-designers, designers, and letter designers (see table 5.29).

11. In the context of this work and the terminology used, the term character designer would have been more appropriate, but an exception has been made here for the sake of clarity.

Questionnaire response	Design categories
I am not a designer	Non-designers
I am a graphic designer	Designers
I am a typographer	Designers
Other kind of designer	Designers
I am designing letters (type, calligraphy, lettering)	Letter designers

TABLE 5.29: Mapping between participants' self-assessment in the questionnaire of the study (see section 5.3.1) and the categories used in the comparisons below.

The percentages and Cohen's kappa statistic in table 5.30 show substantial and almost perfect agreement regarding the most popular response between participants divided into different categories based on their design skills. One can thus compare the distributions of the most popular response frequencies across these groups and observe whether there is an association between participants' backgrounds and changes in the distribution. Statistics describing the distributions are in table 5.31, while figure 5.19 shows the corresponding box plots.

Script	Design categories compared	Trials	Percentage	Kappa
Cyrillic	**Non-designers (369) vs. designers (126)**	448	87.28%	0.809
	Non-designers (369) vs. letter designers (33)	112	85.71%	0.786
	Designers (126) vs. letter designers (33)	112	85.71%	0.786
Devanagari	**Non-designers (356) vs. designers (73)**	392	76.02%	0.640
	Non-designers (356) vs. letter designers (24)	112	81.25%	0.719
	Designers (73) vs. letter designers (24)	112	75.89%	0.638
Latin	**Non-designers (411) vs. designers (304)**	672	88.39%	0.826
	Non-designers (411) vs. letter designers (91)	504	80.56%	0.708
	Designers (304) vs. letter designers (91)	504	85.52%	0.783

TABLE 5.30: The percentages and Cohen's kappa describing agreement between different design categories of participants. Numbers of participants in each category are included in parentheses. Number of trials used for the calculation is in the column "Trials".

Script	Design category	Count	Mean	SEM
Cyrillic	Non-designers	369	0.69	0.007
	Designers	126	0.72	0.008
	Letter designers	33	0.76	0.010
Devanagari	Non-designers	356	0.62	0.007
	Designers	73	0.64	0.008
	Letter designers	24	0.82	0.011
Latin	Non-designers	411	0.72	0.007
	Designers	304	0.72	0.007
	Letter designers	91	0.78	0.007

TABLE 5.31: The counts of studies and means and standard errors of the means (SEM) describing the distribution of the most popular response frequencies for trial responses by non-designers, designers and letter designers for the three scripts studied.

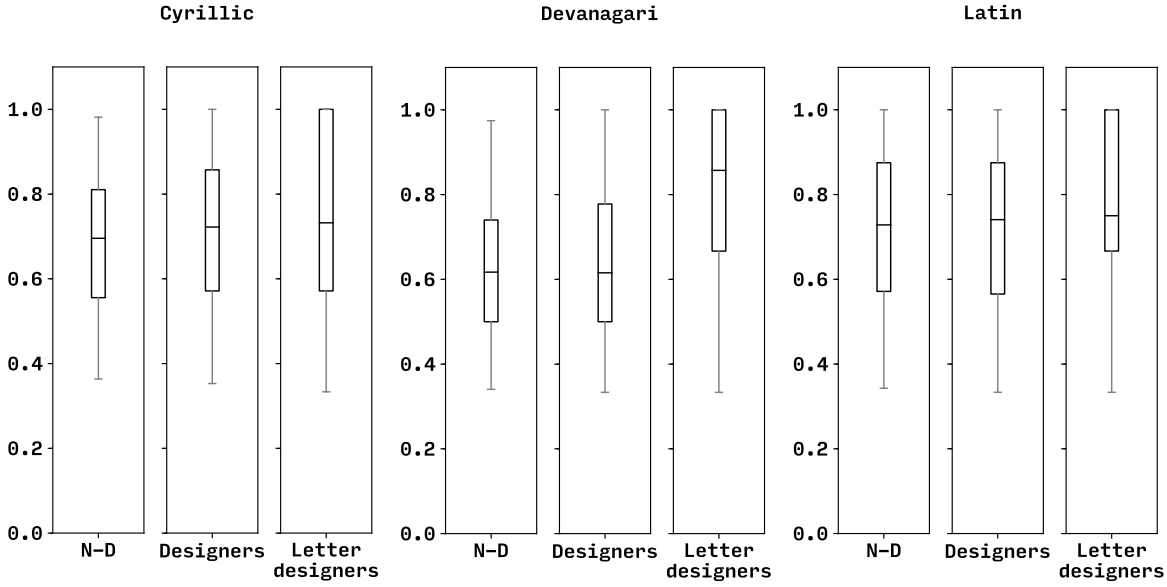


FIGURE 5.19: Box plots showing distribution of the most popular response frequencies for non-designers (N-D), designers, and letter designers for Cyrillic, Devanagari, and Latin. See figure 5.13 for interpretation of box plots.

Table 5.32 shows p-values from a standard t-test (Kmenta, 1986) of the null hypothesis that the means of two distributions are equal. The t-test is used to compare each group with each of the other two groups. These comparisons are done separately for each script. The intended $\alpha = 0.05$ was adjusted using the Bonferroni correction[12] (Dunn, 1961) to $\alpha = 0.05 / 14$. With the adjusted $\alpha = 0.0036$, the p-values indicate that the null hypothesis of no difference between the means holds only between designers and non-designers. The null hypothesis can be rejected for all other cases. Note that the t-test is only an approximation as the distributions are not normal.

The statistics and the box plots show that the distributions of the most popular responses for designers and non-designers are generally quite similar and the results of the t-tests confirm this. The triplet trials were easier for letter designers overall as their agreement on individual trials is significantly higher compared to either designers or non-designers. This can be attributed to their expertise. Note, however, that the box plots for letter designers include the difficult trials as well. Phrased differently, letter designers show greater agreement on trials that are generally easier, but there are still trials that are difficult for them and there they tend to disagree. Again, note that the distributions are not normal and standard errors can be only taken as approximations.

Script	Design categories compared	p-value
Cyrillic	Non-designers vs. designers	0.019
	Non-designers vs. letter designers	< 0.001
	Designers vs. letter designers	0.002
Devanagari	Non-designers vs. designers	0.061
	Non-designers vs. letter designers	< 0.001
	Designers vs. letter designers	< 0.001
Latin	Non-designers vs. designers	0.833
	Non-designers vs. letter designers	< 0.001
	Designers vs. letter designers	< 0.001

TABLE 5.32: The p-values from t-tests comparing the means of the distributions of the most popular response frequencies for categories of participants depending on their design skills.

12. In the sections 5.15.3 and 5.15.4, 14 different comparisons of the means are being tested using the t-test on the same data (see p-values in tables 5.32, 5.34, and 5.35). To account for the increased likelihood that the null hypothesis would be rejected as rare events are more likely to occur during repeated testing, the p-value is divided by 14.

The amount of data for the letter designers is quite limited, especially for Cyrillic and Devanagari (see column “Count” in table 5.31). This could also explain a greater variance of the distributions for Cyrillic and Devanagari where there are fewer letter designers. More data would be required to comment with greater confidence; however, the current data seems to support the hypothesis that letter designers have clearer opinions about similarity judgements perhaps due to professional training.

5.15.4. Comparison between native and non-native, and fluent and non-fluent participants

Another question is whether competent readers perceive similarity differently from less competent readers of a particular script. The technique used for this data exploration is the same as in the previous section 5.15.3.

One might expect that native and fluent readers (see section 5.8), given their experience and familiarity with the script, would have developed more consistent strategies for judging similarity and the trials would be easier for them. The data do not support this hypothesis. The statistics describing agreement regarding most popular responses between the different groups are in table 5.33. The statistics describing the distributions of the most popular responses for these groups of participants are in tables 5.34 (native vs. non-native) and 5.35 (fluent vs. non-fluent). The box plots are in figures 5.20 and 5.21, respectively. These suggest, contrary to the initial expectation, that the contrasting groups generally agree substantially or almost perfectly with each other. Moreover, the reported means indicate that the trials were generally easier for non-native and non-fluent participants. There is not enough data for non-fluent Latin readers, therefore the statistics and box plot for the corresponding comparison are not shown. A slightly higher variance for non-native and non-fluent participants can be seen in the standard errors.

Script	Groups compared	Trials	Percentage	Kappa
Cyrillic	Native (480) vs. non-native (48)	168	89.29%	0.839
	Fluent (495) vs. non-fluent (33)	56	80.36%	0.705
Devanagari	Native (271) vs. non-native (182)	448	81.25%	0.719
	Fluent (372) vs. non-fluent (81)	448	78.12%	0.672
Latin	Native (765) vs. non-native (45)	112	91.96%	0.879

TABLE 5.33: The percentages and Cohen’s kappa describing agreement between native and non-native and between fluent and non-fluent participants. Numbers of participants in each category are included in parentheses. Number of trials used for the calculation is in the column “Trials”. The data for the non-fluent participants in the Latin-script studies are too small to make a meaningful comparison.

Script	Native	Count	Mean	SEM	p-value
Cyrillic	**Native**	480	0.69	0.007	< 0.001
	Non-native	48	0.77	0.009	
Devanagari	**Native**	271	0.62	0.007	0.001
	Non-native	182	0.65	0.008	
Latin	**Native**	765	0.72	0.007	< 0.001
	Non-native	41	0.82	0.008	

TABLE 5.34: The counts of numbers of sessions, and means and standard errors of the means (SEM) describing the distribution of the most popular response frequencies for trial responses by native and non-native participants for the three scripts studied. The p-values are for t-tests comparing the means of the distributions of the most popular response frequencies between the groups of native and non-native participants of the same script.

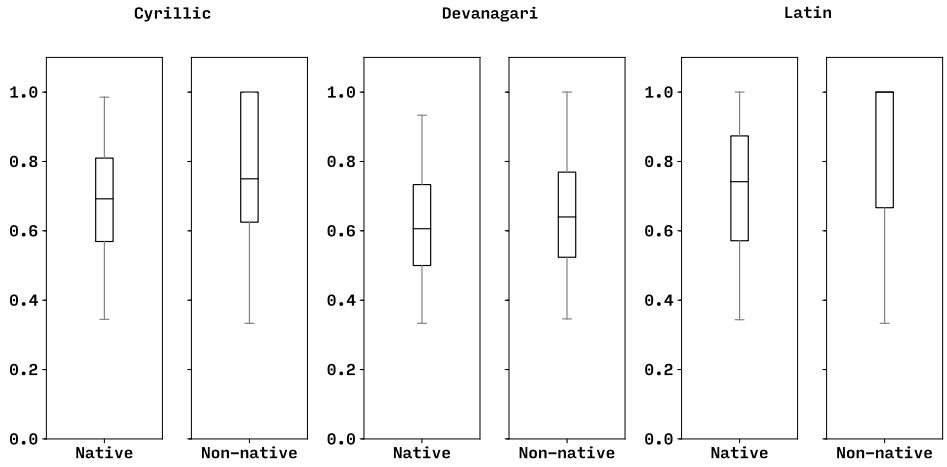


FIGURE 5.20: Box plots showing distribution of the most popular response frequencies for native and non-native participants in the corresponding script for Cyrillic, Devanagari, Latin. See figure 5.13 for interpretation of box plots.

The t-test can be used to compare the distributions of the most popular response frequencies between the groups of native and non-native participants and between the fluent and non-fluent participants. These comparisons are done separately for each script. For the adjusted $\alpha = 0.0036$, the p-values from tables 5.34 (native vs. non-native) and 5.35 (fluent vs. non-fluent) indicate that the null hypothesis of no difference between the means does not hold. There is a significant difference between the means of the corresponding groups.

The numbers of non-native and non-fluent readers are relatively small, and therefore the following comments should be taken as an indication of possible tendencies rather than generalizable conclusions. Also note that none of this suggests that any of the groups perceive the similarities better than others.

One possible explanation for the lower difficulty of individual trials among non-native and non-fluent readers could be that they are mostly experts, but table 5.36 shows that this is not the case. Possibly, native and fluent readers considered more complex relationships or might have been influenced by other, non-presented typefaces, writing conventions, or linguistic or phonetic values of the characters. Non-native and non-fluent readers may have treated character shapes in a more straightforward way due to their limited previous exposure to the corresponding script. Related considerations have been also discussed by Mary C. Dyson, Keith Tam, Clare Leake, and Brian Kwok (2015). Alternatively, being less familiar with the presented script, the non-native and non-fluent participants might spend more time assessing the shapes, and not jump to conclusions as easily as the more familiar, and possibly more confident, native and fluent participants.

Script	Fluent	Count	Mean	SEM	p-value
Cyrillic	**Fluent**	495	0.69	0.007	< 0.001
	Non-fluent	33	0.77	0.009	
Devanagari	**Fluent**	372	0.62	0.007	< 0.001
	Non-fluent	81	0.68	0.008	

TABLE 5.35: The counts of numbers of sessions, and means and standard errors of the means (SEM) describing the distribution of the most popular response frequencies for trial responses by fluent and non-fluent participants for two of the three scripts studied. The data for the non-fluent participants in the Latin-script studies are too small to make a meaningful comparison. The p-values are for t-tests comparing the means of the distributions of the most popular response frequencies between the groups of fluent and non-fluent participants of the same script.

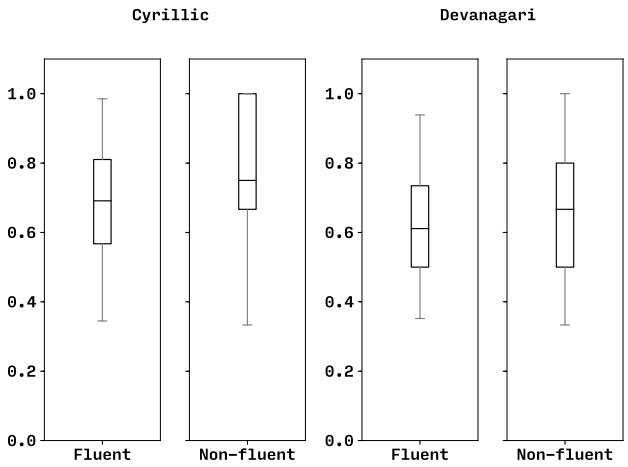


FIGURE 5.21: Box plots showing distribution of the most popular response frequencies for fluent and non-fluent participants in the corresponding script for Cyrillic and Devanagari. The data for the non-fluent participants in the Latin-script studies are too small to make a meaningful comparison. See figure 5.13 for interpretation of box plots.

Script	Design skill	Native	Non-native	Fluent	Non-fluent
Cyrillic	**Graphic designer**	74	3	76	1
	Letter designer	28	5	28	5
	Non-designer	334	35	347	22
	Other designer	30	1	30	1
	Typographer	14	4	14	4
Devanagari	**Graphic designer**	5	16	8	13
	Letter designer	11	13	14	10
	Non-designer	231	125	313	43
	Other designer	22	18	32	8
	Typographer	2	10	5	7
Latin	**Graphic designer**	160	10	170	0
	Letter designer	81	10	90	1
	Non-designer	400	11	411	0
	Other designer	81	4	85	0
	Typographer	43	6	48	1

TABLE 5.36: Numbers of participants with different design skills and nativeness/fluency for each script studied.

5.16. Visual vs. linguistic interpretation

The potential interference between visually- and linguistically-motivated criteria mentioned in section 5.2 is an important consideration. The participants were given clear instructions and were nudged by the design of the study towards assessing the visual aspects of the characters. However, it might have been difficult for them to separate the linguistic aspects completely as most of the time they are used to seeing characters when reading or writing.

A study using participants with such diverse linguistic backgrounds alleviates this problem to a large extent. Because if a language had an effect, it would be highly unlikely that the effect would be the same for all the different languages and participants. However, if there was a significant effect, it would show as a difference between the native and non-native or fluent and non-fluent participants. Statistics in table 5.33 show that there is a substantial agreement regarding the most popular responses between these contrasted groups of participants, illustrating there is no or minimal effect.

Chapter 6 will discuss how well expert knowledge of visual character similarity can explain participants' responses. Here, it should suffice to show a few examples to provide additional evidence that linguistically-motivated criteria are unlikely to be influential enough to be of concern. Trials where participants could have decided based on linguistic aspects, but chose differently, can provide such evidence.

Discussion of what constitutes a linguistically-motivated criterion is not the objective of this research. Therefore, only two simple criteria participants were likely to have used are discussed. These are judged with respect to the three major languages among the languages participants reported as their native language: Russian, Hindi, and English. Firstly, the distinction between vowel and consonantal characters such that participants would always keep vowel characters and consonantal characters together. Table 5.37 shows that this is not the case. Secondly, the connection between the non-aspirated, e.g. 'प' [Pa], and aspirated forms, e.g. 'फ' [Pha], of syllables in Hindi suggests that participants would always keep them together. Table 5.38 shows that this is not the case. The tables illustrate that these linguistically-motivated judgements did not override visually-motivated judgements that may have been involved. To show whether and how big a linguistic influence there was, is perhaps a topic for further research or a study with a single linguistic background.

character		count	freq.	confidence interval
a	a	8	0.13	(0.06, 0.25)
ha	x	49	0.82	(0.7, 0.9)
ve	в	3	0.05	(0.01, 0.14)

character		count	freq.	confidence interval
ha	x	6	0.10	(0.04, 0.21)
u	y	15	0.25	(0.15, 0.38)
ze	з	39	0.65	(0.52, 0.77)

character		count	freq.	confidence interval
Ddha	ढ	31	0.53	(0.4, 0.67)
E	ए	17	0.29	(0.18, 0.43)
Ma	म	10	0.17	(0.09, 0.29)

character		count	freq.	confidence interval
A	अ	14	0.28	(0.16, 0.42)
Gha	घ	14	0.28	(0.16, 0.42)
Ra	र	22	0.44	(0.3, 0.59)

character		count	freq.	confidence interval
a	a	14	0.21	(0.12, 0.33)
l	l	51	0.76	(0.64, 0.86)
p	p	2	0.03	(0.0, 0.1)

character		count	freq.	confidence interval
f	f	24	0.32	(0.22, 0.44)
i	i	19	0.26	(0.16, 0.37)
n	n	31	0.42	(0.31, 0.54)

TABLE 5.37: Trials where the majority of participants could have marked the only vowel character as the odd one out (marked in green), but did not. This is judged with respect to Russian, Hindi, and English. The overall odd one out is marked with a grey background.

character		count	freq.	confidence interval
Cha	छ	10	0.20	(0.1, 0.33)
Da	द	33	0.65	(0.5, 0.78)
Dha	ध	8	0.16	(0.07, 0.29)

character		count	freq.	confidence interval
Ka	क	27	0.43	(0.3, 0.56)
Pa	प	32	0.51	(0.38, 0.64)
Pha	फ	4	0.06	(0.02, 0.15)

TABLE 5.38: Trials where the majority of participants marked characters for the non-aspirated forms of syllables 'द, प' [Da, Pa] as the odd one out while their respective aspirated forms 'ध, फ' [Pa, Pha] were also present in the trial triplets. This is judged with respect to Hindi. The overall odd one out is marked with a grey background.

5.17. Possible methodological improvements

Concluding the studies provides an opportunity to look back and judge the appropriateness of some of the methodological decisions.

The studies presented participants with a new, albeit seemingly simple, task which required some familiarization. Thus, the responses to the first few trials in the studies could be considered less reliable. Through randomization of the order of the trials for each participant, any systematic effect on the results is avoided. However, to avoid some of these less reliable responses, a short practice session might have been included at the beginning of each study allowing participants to familiarize themselves with the task. Note that removing the first few responses of each participant was not an option as all responses from each session had to be retained for aggregation (see sections 5.2.1 and 5.13).

Another possible improvement might be in relation to the sampling. Having found out that it is possible to collect a large number of responses using this kind of study online, it might be useful to select the characters for the study sets randomly and not solely rely on systematic stratified sampling. Systematic sampling, although justified, seems in retrospect overly complicated and might obfuscate a potential unwanted bias in the selection of the triplets. On the other hand, simple random sampling might not be representative of the whole population of character triplets for each script in terms of difficulty. Thus, it would be ideal to complement the studies based on systematic stratified sampling with a study based on simple random sampling.

Timing of participants' responses was not implemented for technical reasons, but it could confirm the difficulty of individual trials.

And lastly, the combined studies, i.e. the pairs of two individual simple studies in one session (see section 5.3), were analyzed as two individual studies completed by the same participant. They could be used to study a transfer of participants' criteria from one typeface to another. For example, participants could consider two characters in one typeface more related than suggested by their shape because the two characters were more related in the previous typeface they have seen. This would require collecting more data, though.

5.18. The use of the data

The cluster plots (see figures 5.15–5.17 and appendix 4) and the comparisons between typefaces (see table 5.27 and appendix 6) illustrate that participants understood the task they were given in the studies. They predominantly judged visual similarity of character shapes and not, for example, their linguistic or phonetic similarity (see section 5.16). The analysis of their agreement on individual trials shows that the trials from the studies provided a diverse testing ground in terms of difficulty and that there is some indication of an overall opinion on what constitutes a correct response.

The variety of participants' backgrounds and typefaces studied serves primarily to provide a generalized testing ground, not to study differences between these groups. The comparisons in section 5.15 should be viewed as an attempt to explore potential divergencies in the data before it is used, rather than an ultimate goal. The intention is to use the diverse data collectively. While there were differences between the different groups of participants and depending on the typeface used, none of these were large enough to justify isolating a particular group of participants. Namely, the letter designers who participated in the study are relatively few to affect the data significantly. Keeping the data together will provide more generalizable results than would a study with a single typeface and a specific group of participants.

However, when making inferences about the general readers' population, one should bear in mind participants' background and the way the data was collected and assess whether they matter to the inferences drawn. For example, the participants of the studies are mostly under 50 years of age. This does not necessarily mean that the findings of this study do not apply to older readers, but it might require some caution in extrapolating to a wider population. The fact that no Bantu readers took part in the studies does not mean that they perceive similarity differently. However, if there was evidence of different educational or visual processes in their population that might affect their similarity judgements, a different study should be conducted.

In the next chapter, the collected observations of readers' similarity judgements are used to train a computational model. It might seem useful to identify and remove outliers to emphasize dominant opinions among participants. Such an elimination constitutes a principal problem as there are no right or wrong answers when it comes to similarity judgements. Even minority idiosyncratic opinions can inform one about criteria readers used when judging similarity.

6. Modelling character similarity and coherence

This chapter discusses the construction and assessment of a newly proposed theoretical model of character similarity and coherence. The research objective is to query whether coherence intended by experts differs from coherence as perceived by readers. In order to do this, the model is based on expert knowledge (for overview see chapter 4: *Approaches to character conceptualization*). The data from the studies with readers (see chapter 5: *Studies with readers*) represents coherence as perceived by readers. Following from the discussion of the research questions in the introduction (see section 1.1), the model is conceived as independent of any design principles specific to a particular script or technology.[1]

The data collected is limited to similarity judgements on a particular set of character combinations that were included in the studies. On the other hand, the model can potentially predict responses for any character combination as long as the character shapes can be represented in the model and the character shapes were produced using the same typeface.

The model generates *predicted response frequencies* (or *predictions*) for participants' responses in the studies. In order to describe general readers' response behaviour, the predictions are compared with aggregated participants' responses from the studies, i.e. the *observed response frequencies* (or *observations*). Note that the response frequencies for each trial are relative and normalized to add up to 1.0 (see section 5.11.2). Thus, they can be used as probability estimates. Unlike some models used by other researchers, for example, in cognitive science, the model used in this work does not predict responses of individual participants. Instead, it predicts probabilities of all participants' responses. The model is given a triplet[2] from a particular trial as an input and generates predictions of frequencies for each of the three responses. It predicts the probability that a participant would consider the first character in the triplet as the odd one out, and then the same for the second and the third character.

The model contains a set of inner parameters which can be tuned in such a way that the model's predictions are close to the observations. In machine

1. Note that reproduction technology does not only include machinery but also chirographic tools, since different scripts may have different conventions based on chirographic traditions which may employ different tools. The model accounts for some of these, but does not rely on them.
2. In practice, a triplet of character names and a typeface name are passed as an input to the model to obtain a response prediction.

learning, this process is called *supervised training*. It means that the model is trained by example. It can see the intended output, i.e. the observations, during the training and alters its inner parameters automatically to achieve as close a fit of its own predictions to the observations as possible.

The model's assessment indicates to what extent the model can, based on expert knowledge, account for the observations, i.e. participants' perception of coherence in the studies.

6.1. Modelling approaches

A good model of some observed phenomena does not only make predictions that are close to the observations (goodness of fit). Following from the scientific principle of parsimony that things are usually connected or behave in the simplest or most economical way (Thorburn, 1918), a good model should also make the predictions in the simplest way possible. It should not use an unnecessary number of parameters or overly complicated formulas if the same can be achieved by simpler means. Moreover, a good model should also provide results that are clear to interpret and that provide insights about the phenomena under study. In this case, the insights ought to relate to cognitive processes involved in similarity judgements and to typeface design practice. Indeed, a model that can predict and provide some insights is superior to a model that can merely predict.

Besides the empirical approaches discussed in section 5.1.2, there have been various theoretical approaches to modelling and calculating the measure of similarity among character shapes. The following is a brief overview of the more common ones.

Lucien A. Legros and John C. Grant, in their seminal work *Typographical printing-surfaces* (Legros & Grant, 1916), showed an approach where pairs of characters were superimposed to identify their shared, overlapping area. This area was related to the total area of both character shapes to calculate a measure of their similarity (see figure 6.1).

The approach uses only one way to position the overlay and thus one resulting shape of the shared area, which begs the question of whether there could sometimes be a better way to overlay the characters. The method gives intuitive results when the two overlaid shapes are quite similar. However, if the two characters are similar because of a feature that is placed in a relatively different location in each character this method of comparison is insufficient or even counter-intuitive (see the overlaid pairs of characters 'u, n' and 'i, l' in figure 6.2). Moreover, it cannot explain similarities that are not based on exact shape similarity. For example, the letter shapes of a and e are similar because they have a stroke crossing them in the middle, but they are quite different in terms of geometry.

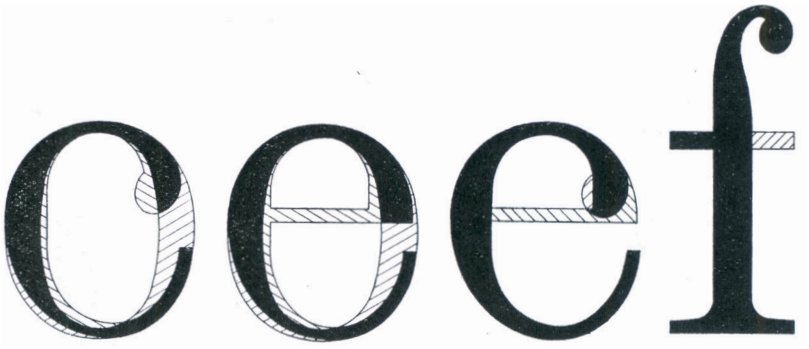


FIGURE 6.1: Legros and Grant superimposed character shapes to study their similarity (illustration from: Legros & Grant, 1916, p. 162).

FIGURE 6.2: Comparison of more distinct shapes using Legros and Grants' approach is less congruent with perceived similarity. On the left, the characters 'n' and 'u' are superimposed and the arched part which is shared by both letters does not contribute to their similarity in this approach. On the right, the comparison of the characters 'l' and 'i' superimposes only a portion of the top serif (illustration from: Legros & Grant, 1916, p. 162).

The approach described by Legros and Grant can be viewed as a variation of a *template approach* that uses templates made of receptors to identify shapes and assess similarity (see figure 6.3). The more receptors are activated by a perceived shape, the more similar it is to the shape the template represents. The template approach corresponds well to our current understanding of low-level vision processes in the brain related to identification of edges and lines, but is unable to explain high-level processes such as shapes' invariance[3] to affine transformations (translation, rotation, reflection, or scale). Hierarchical templates can be used to address some of the weaknesses of the simple template approach, but they are not a definite solution and increase the complexity of the models significantly (Palmer, 1999).

3. Irvin Rock (1973; as discussed in Palmer, 1999) showed that in some cases these transformations indeed affect the perception of shape identity. However, as shown by Marianne Wiser (1981) these issues seem to affect amorphous shapes more than those with "good" intrinsic qualities.

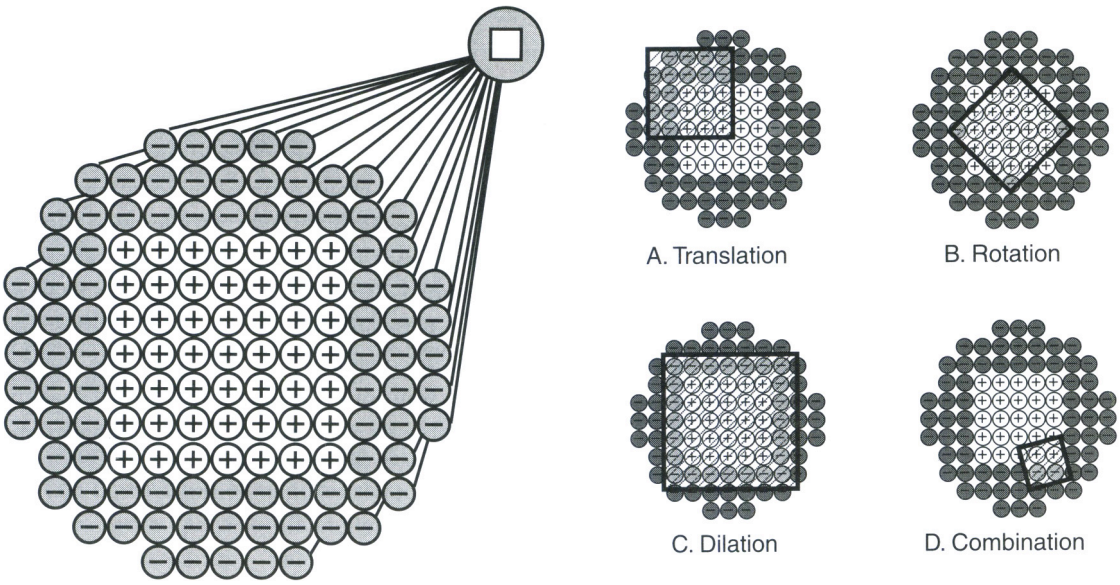


FIGURE 6.3: A square shape represented as a template made of receptors (left). The images A, B, C, and D on the right show the problems with common spatial transformations. People perceive the shaded shape as a square despite the fact that it overlaps with the template poorly. Illustrations from: Palmer, 1999, pp. 378 & 380.

Another approach is based on a *Spatial frequency theory* which represents an image as a set of two-dimensional patterns, called sinusoidal gratings, that differ in their spatial frequency. The approach assumes that a perceived shape is processed using a Fourier-like transformation in the visual system in order to obtain the gratings. These can be used to produce spatial frequency channels of the image (see figure 6.4). Spatial frequency channels are hypothetical mechanisms in the visual system that are selectively tuned to a limited range of spatial frequencies. The channels contain information about the corresponding level of detail in the shape. High-frequency channels contain information about fine details while low-frequency channels contain information about overall features (Palmer, 1999).

The spatial frequency approach has been successfully used to explain effects in letter identification (Majaj et al., 2002) or to interpret finer aspects of typeface design practice (Ahrens & Mugikura, 2014). Nonetheless, the typographic expertise described in chapter 4 that deals with establishing coherence among characters is not primarily concerned with low-level vision challenges such as edge and line detection or a particular level of detail. For this reason, neither the template approach, nor the spatial frequency approach seem suitable for modelling character similarity in this research.

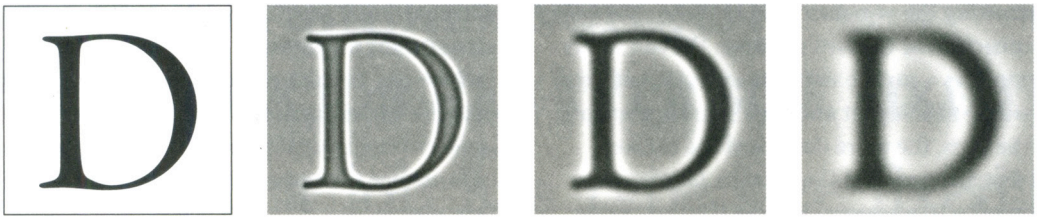


FIGURE 6.4: Latin capital letter 'D' (left) has been analyzed into three frequency channels. These are illustrated in the three images to the right: high-frequency (first), middle-frequency (second), and low-frequency (third). Illustration from: Ahrens & Mugikura, 2014, p. 29.

On the other hand, a *feature-based approach* interprets objects as collections of symbolic features (see figure 6.5). These features are usually assigned to the object by researchers based on their direct interpretation of the objects or mathematical analysis of empirically collected data. When assessing the similarity of any two objects, their features are compared. Typically, the features common to the two objects contribute towards their similarity while the features that are distinct do not. By interpreting objects as collections of features, each with its own salience, the approach provides insights regarding the underlying processes involved in object similarity judgements that are seen as feature-matching processes. Unlike the template approach which treats objects as wholes and only matches features that are in corresponding relative locations, the feature-based approach allows matching of features in different relative locations.

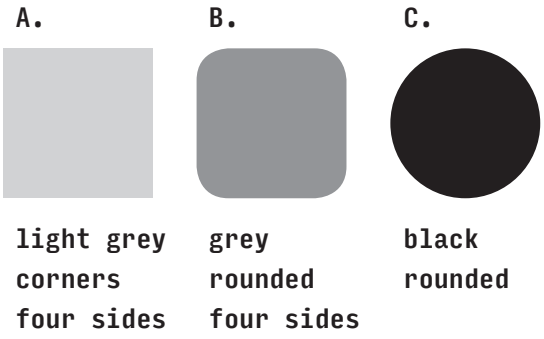


FIGURE 6.5: Three objects (A, B, C) described using several features without any concern for the features' saliences. Naturally, there could be various feature-based representations of these three objects.

A *geometric model* is a specific version of the feature-based approach where objects are represented as points in a multidimensional coordinate space and their dissimilarity measures correspond to metric distances in this space. Each coordinate axis corresponds to one of the features an object can have, the position on the axis corresponds to the feature's salience (see figure 6.6).

If an object has a feature, its representative point is moved along this particular axis. The geometric model assumes that dissimilarity and similarity are mutually inverse and that they satisfy three basic metric axioms of minimality, symmetry, and triangle inequality. These axioms are further explained here only in terms of similarity as the cases for dissimilarity are typically inverse. In the following, A, B, C denote three different objects and s(A, B) denotes a measure of similarity between two objects A and B. The minimality[4] axiom states that an object is most similar to itself: s(A, B) < s(A, A). The symmetry axiom states that the ordering of two objects does not affect their similarity measure: s(A, B) = s(B, A). And lastly, the triangular inequality axiom states that the similarity measure of any two objects A and B cannot be smaller than their similarity measures to a third object C combined: (s(A, C) + s(C, B)) ≤ s(A, B).

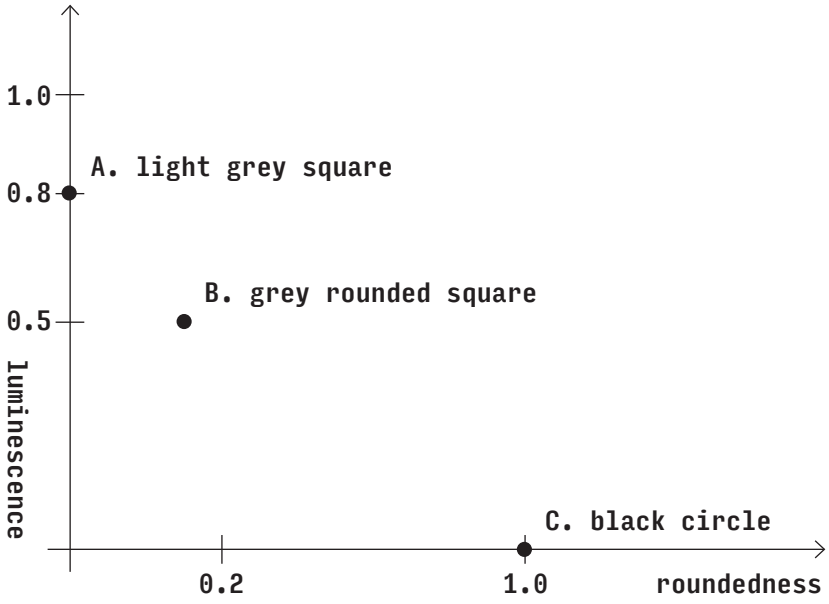


FIGURE 6.6: A simple example of a geometric approach shows the objects from figure 6.5 plotted in a two-dimensional space defined by two axes based on two features: "luminescence" and "roundedness". The position of an object on each axis depends on the salience of the corresponding feature in this object.

Interpreting similarity as a metric is an elegant and intuitive solution allowing convenient calculations. For example, it makes it possible to find minimal and maximal similarity measures or study clusters of objects grouped by their similarity. However, Amos Tversky (1977) showed that similarity as perceived by humans does not always conform to these three axioms. Some objects are more frequently confused with others than identified correctly, thus causing a violation of the minimality axiom. Tversky (1977) pointed out that

4. The name of this axiom comes from the minimality of the distance, i.e. maximal similarity.

similarity statements are directional, thus breaking the axiom of symmetry. Note that both of these objections depend on the way the experiments were conducted. If an experiment required comparison of objects instead of identification, the minimality axiom might not be violated. Also, if the experiment did not require directional similarity statements, such as the form of the trials in this research, the axiom of similarity would hold as well.[5]

To disprove triangular inequality, Tversky gives an elegant example: "Jamaica is similar to Cuba (because of geographical proximity); Cuba is similar to Russia (because of their political affinity); but Jamaica and Russia are not similar at all.". One can easily construct an analogical example for letters: a k and a d are similar (because of their ascenders), the d and an o are similar (because of their rounded bowls), but the k and the o are not similar at all.

Thus the geometric approach may not always be ideal as a model of perceived similarity. However, one of the main advantages of the geometric approach is that multidimensional scaling (MDS) (Torgerson, 1952; Shepard, 1962a; Shepard, 1962b; Kruskal, 1964) can be used to obtain two- or three-dimensional interpretations of the multidimensional space describing objects' similarity. This is helpful when one needs to get basic, if approximate, intuitions about the relative similarities of objects studied.

A variation of the geometric model, popular in the 1970s and 1980s, is based on a *choice theory* by R. Duncan Luce (1963; 1977). At the time, the choice theory seemed to have produced highly accurate predictions (Townsend, 1971; Holbrook, 1975). The theory assumes two ratio scales. The first scale is interpreted as a similarity metric described above and is determined by sensory influences. The other scale represents the bias associated with the response system (which does not depend on sensory effect). The response to a similarity judgement is determined by a combination of these two scales. Although the two scales are assumed to be independent, Gideon Keren and Stan Baggen (1981) showed this is not always the case. Moreover, they also pointed out that the choice model incorporates a relatively high number of parameters: "350 parameters have to be estimated in order to predict 676 cells of the confusion matrix", thus questioning the model's parsimony (Keren & Baggen, 1981, p. 237).

By way of addressing the discrepancies between the perceived similarity and the metric axioms of the geometric model, Tversky (1977) proposed his *contrast model* based on a class of matching functions. For any two objects, a matching function assesses the intersection of their shared features and the features assigned to one object, but not to the other. Different matching functions weight these assessments differently and produce various measures of similarity (see figure 6.7). Some of these measures preserve some of the metric axioms, namely the axioms of minimality and symmetry.

5. This is further supported by the fact that the order of characters in trials has been randomized which balances out any effects of asymmetry in similarity judgements.

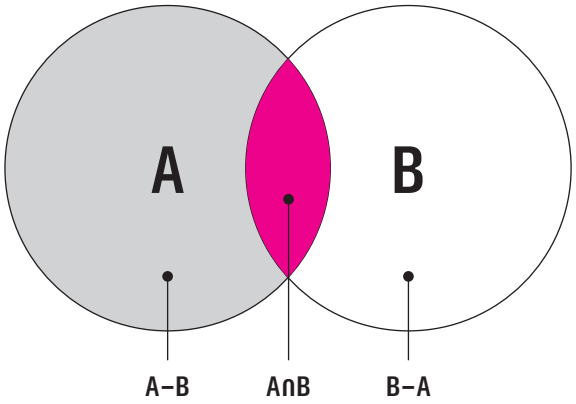


FIGURE 6.7: The Venn diagram shows the set of features in the concepts A (the area of the circle on the left) and B (the area of the circle on the right). Moreover it shows the set of features shared by concepts A and B (the area marked in magenta), the set of features in A, but not in B (the grey area), and the set of features in B, but not in A (the white area in the circle on the right). Tversky's matching functions combine the latter three areas to produce various measures of similarity.

The contrast model also permits study of the influence of context on the similarity judgements. Tversky (1977) suggests that changing context or frame of reference changes the classificatory significance of features, which in turn affects feature salience. In this research, the features representing a contextual character are involved in the calculation of the similarity measure of the characters that are being compared.

Conveniently, the feature-based approach also allows for model training based on observation data collected in experiments. Contemporary researchers have successfully used the approach in their studies (Keren & Baggen, 1981; Garces, Agarwala, Gutierrez & Hertzmann, 2014; Lun, Kalogerakis & Sheffer, 2015), although the formulas used for combination of features may differ from Tversky's contrast model.

Moreover, the feature-based models lend themselves well to the study of coherence seen as a result of designers' creative considerations. Many designers consciously distribute common features across objects in a series and may suppress features that are distinct (see sections 2.6 and 4.4).

Supported by current theories of the human visual system, some researchers have suggested hierarchical models of object representation and recognition where objects are represented as tree-like structures with features of varying levels of generality (e. g. Palmer, 1977; Marr & Nishihara, 1978; Riesenhuber & Poggio, 1999; Palmer, 1999). However, it might be unnecessary to adapt this approach to the modelling of similarity. Even though, objects are represented hierarchically, their similarity may be judged in a non-hierarchical way. It is unclear whether context affects the way the salience values of the more general features are derived from the more specific features connected to

them. If this is not the case, the relative saliences within each object are constant and describing the similarity using non-hierarchical models is sufficient. The proposed model uses specific as well as more general features, but it treats them as equals when predicting similarity judgements (see introduction to section 6.3).

Hierarchical approaches with multiple layers inspired experiments with deep learning that were conducted during the course of this research. Both neural networks using multi-layer perceptrons as well as convolutional neural networks were tested. However, the experimental neural-network models did not predict better than the model proposed below. Moreover, they provided only a limited explanation of the underlying processes. Their parameters might be interpretable by a human as, for example, one can visualize the layers of a convolutional neural network trained on images of objects to reveal specific parts used to identify these objects. However, these images of parts are not easily connected back directly to ready-made hypotheses. Having a human interpret every image and evaluate which feature it represents and to what extent this feature might have been used in similarity judgements would be extremely tedious and potentially subjective. Even though neural networks provide an interesting direction for the analysis of similarity and possibly for discovery of features as well, they could not be used to test a priori hypotheses derived from expert knowledge to address the research questions. Therefore, they are not discussed further here.

In this context, it is perhaps worth noting that the trials from the studies in chapter 5 may seem like a classification task – something neural networks are known to tackle extremely well – but they are not. Instead, the trials are a comparison task which seems to require a different, possibly new, kind of neural network architecture.

The last approach that is mentioned here interprets similarity as a transformation distance (Hahn, Chater & Richardson, 2003). When an object A can be transformed to an object B, the measure of their similarity can be determined from the number and complexity of transformations required to change object A to B. It can be seen as an extension of the feature-based approach where features based on component shapes can also be transformed. Such a hybrid approach could be based on expert knowledge and would help to incorporate the tranformation-based conceptualizations discussed in section 4.3. However, similarity of character shapes seems to be predominantly based on the introduction or removal of diverse features. Their transformations seem to be of lesser importance. Thus, their similarity can be explained well by a feature-based approach. Also, transformations cannot explain more general features such as those discussed in section 6.3.3. Nonetheless, viewing similarity as transformation promises to be an interesting area of exploration for the future.

6.2. The proposed model of coherence

The approach used in this work builds on the contrast model by Tversky (1977) while implementing some modifications in how the similarity measure is calculated. The proposed model consists of three main components:

i. a character concepts database,
ii. typeface mappings, and
iii. a predictive method.

The concepts database contains *character concepts* that can be used to represent character shapes. Note that character concepts in the model are also character concepts as they were defined in section 2.2. The *typeface mappings* define which character concepts should be used for character shapes produced using a particular typeface studied. And lastly, a *predictive method* which, given a triplet of character names and a typeface name, calculates predicted response frequencies as if these were used in a study trial (see section 5.2 for description).

Character concepts in the model's database are made general in order to represent character shapes made using different typefaces. For example, the same concept is used to represent the character shapes for 'b' made with the typefaces Calibri as well as Futura (see figure 6.8).[6] On the other hand, the model can contain multiple concepts for the same character category. For example, different concepts are used to represent an 'a' in typefaces Calibri and Futura (see figure 6.8). Using a single concept for each character category for all the typefaces studied would be too restrictive; the view that complete character categories can be defined using a single shape-like prototype, was challenged in section 2.2. As mentioned above, the correspondence between character shapes and concepts is stored in typeface mappings.

To find out what and how many concepts are required, character shapes made from all the selected typefaces for a particular script are compared visually. Character shapes that can be potentially represented with the same concept are put together according to the conceptualization process described in section 6.3. A *representative character shape* for each of these character shape groups is then selected and used to produce a character concept that can in turn represent all the grouped character shapes.

6. Character concepts are only used to represent character shapes for a single script. Although, it would be theoretically possible to use the same concept, for example, for Latin character 'a' as well as for Cyrillic character 'a' which tend to look identical in both scripts.

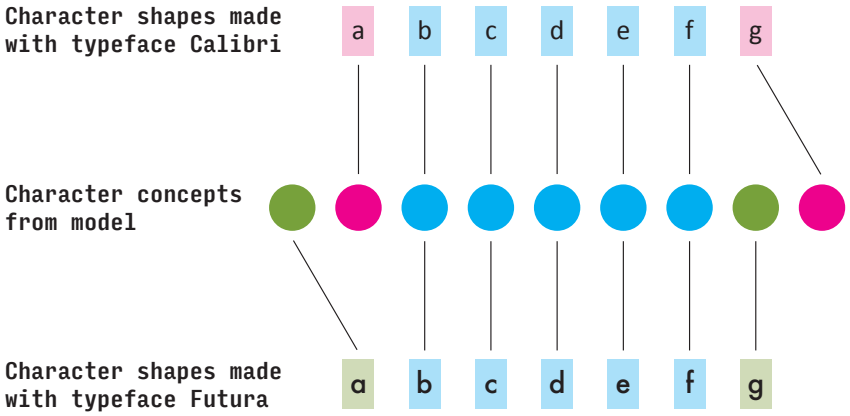


FIGURE 6.8: Character shapes produced using different typefaces (top and bottom row) can be described with character concepts from the model (middle row). Within the same category, different character shapes can use the same concept (marked in blue) or a different one (marked in green and magenta).

The typefaces and corresponding character shapes are separated into two groups for each script for the sake of convenience:

- **Latin sans-serif** group contains typefaces: Arial, Candara, Calibri, Futura, PT Sans, Verdana.
- **Latin serif:** Cambria, Courier New, Century Schoolbook, Georgia, PT Serif, Times New Roman.
- **Cyrillic sans-serif:** Arial, PT Sans, Verdana.
- **Cyrillic serif:** Courier New, Century Schoolbook, Georgia, PT Serif, Times New Roman.
- **Devanagari high-contrast:** Adobe Devanagari, Devanagari MT, ITF Devanagari, Murty Hindi.
- **Devanagari low-contrast:** Ek Mukta, Kohinoor Devanagari, Lohit Devanagari, Nirmala UI.

Figures 6.9–6.11 show comparisons of the character shapes for all the typefaces and scripts studied. The columns correspond to individual character categories, each row to a different typeface. Within each column, character shapes that can be represented with the same concept have the same background colour. For the model to work with the data from the studies, only the concepts for character shapes that appeared in the studies need to be implemented. Those are framed in the figures. However, for the model to predict responses for any combination of characters from each of the typefaces, all the various concepts needed for the character shapes in the charts would need to be implemented. The model includes 190 different concepts: 46 for the Cyrillic serif group, 33 for the Cyrillic sans-serif group, 33 for the Latin serif group, 28 for the Latin sans-serif group, and 50 for Devanagari. The Devanagari concepts are often used for typefaces from both high-contrast and low-contrast groups.

The representative character shapes are predominantly selected from a single, default typeface to make sure the resulting concepts have comparable proportions. The concepts for the serif groups are mostly represented using character shapes from the typeface Georgia. The typeface Arial is used for the sans-serif groups. The typeface Murty Hindi is used for the majority of the Devanagari concepts. The exceptions, i.e. the concepts that cannot be represented by the default typefaces, are represented using character shapes from other typefaces. Those character shapes are rescaled to have the same base height to make sure the concepts are as comparable as possible. In some cases, the ascenders and descenders in the representative character shapes from non-default typefaces are stretched a little to make their proportions in relation to the base height identical to that of the default typeface. These modified representative characters are only used to obtain some of the features automatically and in the concept diagrams in figure 6.14 and appendix 7.

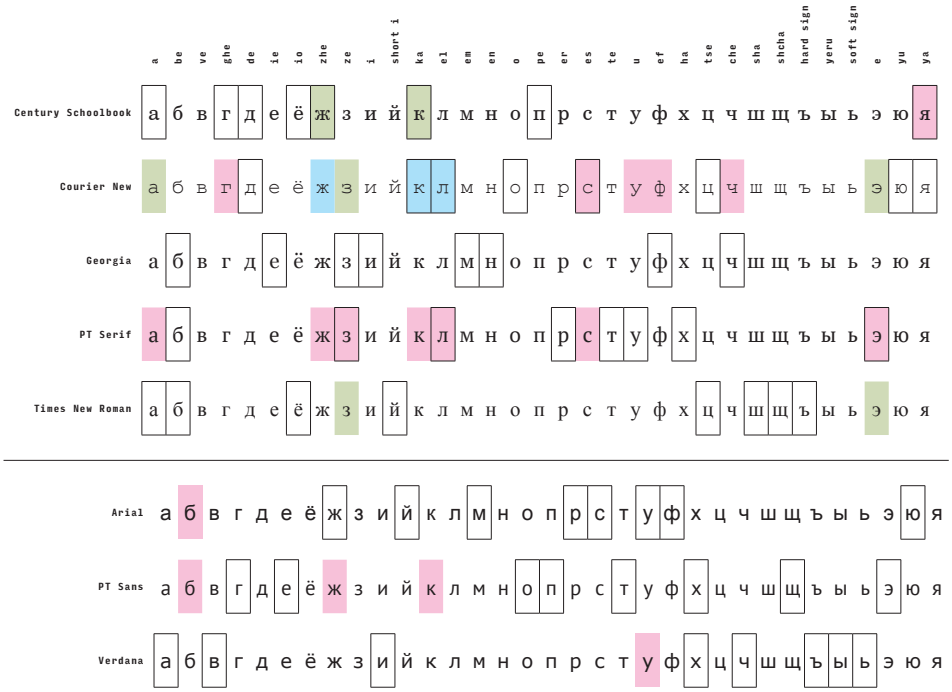


FIGURE 6.9: Comparison of character shapes for the Cyrillic typefaces studied. The character shapes are divided into two groups: serif (top), sans-serif (bottom). Character shapes from the same category (in the same column) with the same background colour can use the same character concept. Character shapes in frames were used in the studies.

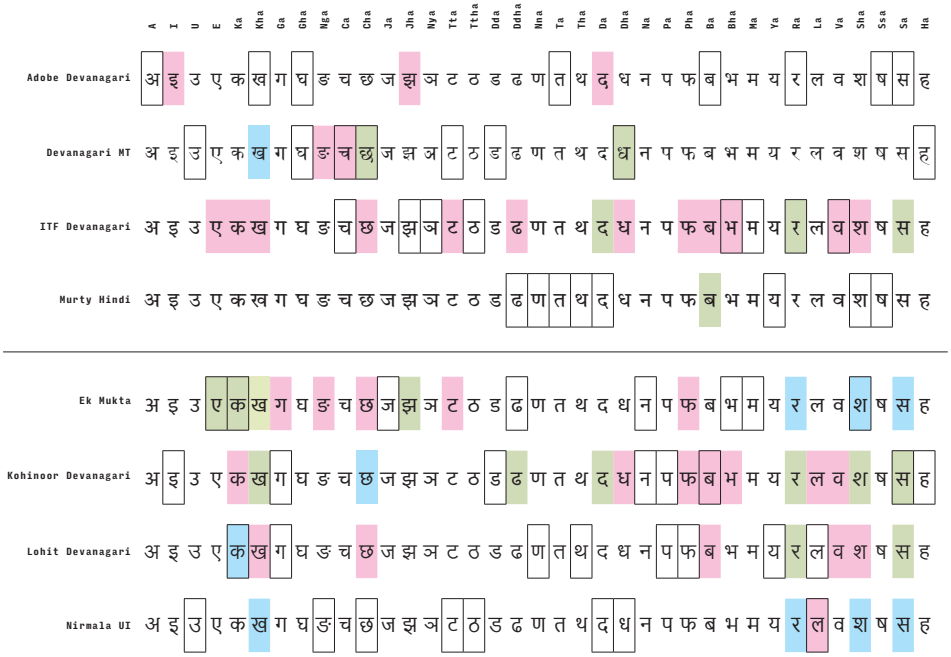


FIGURE 6.10: Comparison of character shapes for the Devanagari typefaces studied. The character shapes are divided into two groups: those with a high contrast (top) and a low contrast (bottom). Character shapes from the same category (in the same column) with the same background colour can use the same character concept. Character shapes in frames were used in the studies.

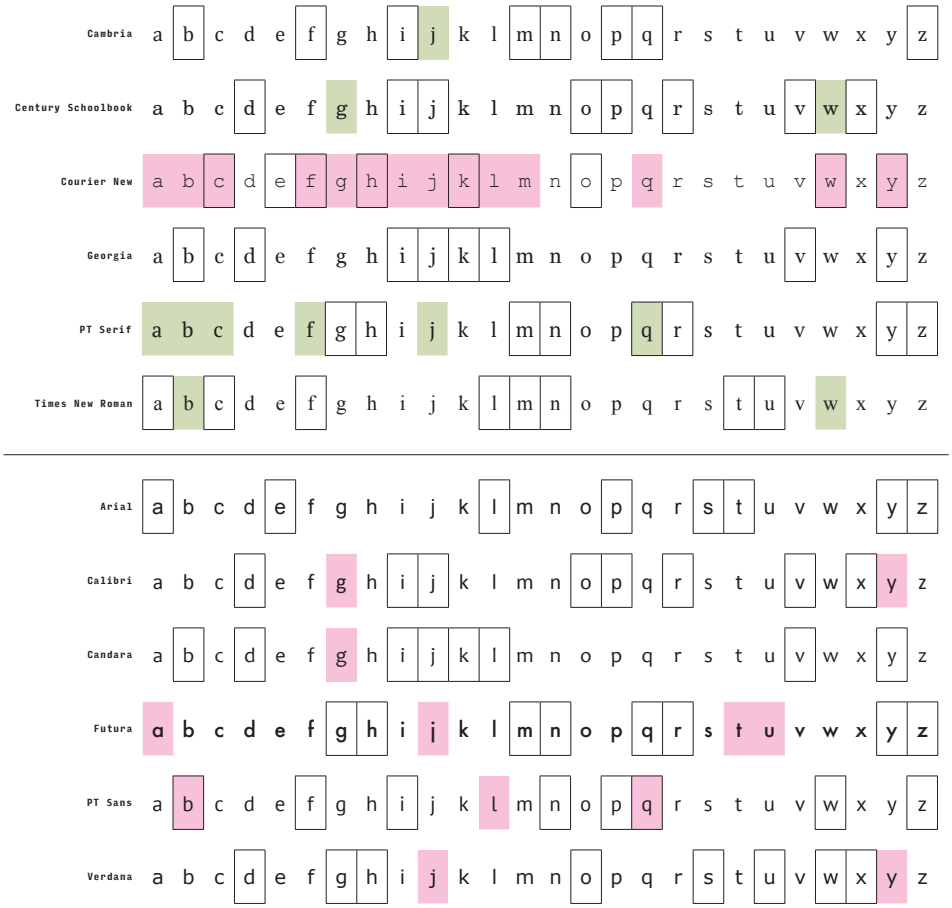


FIGURE 6.11: Comparison of character shapes for the Latin typefaces studied. The character shapes are divided into two groups: serif (top), sans-serif (bottom). Character shapes from the same category (in the same column) with the same background colour can use the same character concept. Character shapes in frames were used in the studies.

6.3. Character conceptualization

Character concepts in the model are collections of features. The nature of features used in a model depends on the objectives and context of the particular research. Since the focus of this research is on typeface design, the features that describe characters are visual and not, for example, linguistic. Additionally, the objective is to describe similarities between characters and not the complete appearance of characters. Each feature in the model represents a potential visual analogy[7] that can be drawn between two or more character shapes. This analogy has to be perceptible and distinctive enough to be considered in similarity judgements. Similarity of the concepts is viewed as a collection of elementary visual analogies represented by the features.

7. The term *analogy* is used to refer to elementary similarity-identity for the sake of brevity.

Deriving features from the character shapes can be a challenging process of interpretation. The concepts are to be used in predicting similarity, thus they need to represent as many analogies between the character shapes as possible. On the other hand, they also need to be kept simple to maintain a good level of parsimony of the model, to keep it manageable, and to allow as clear an interpretation as possible. Also, the more features a model has, the more data is required to train it. Thus, keeping the model simple ensures it can be tested with less data (see section 6.7.1 for evaluation of the size of the data set). Finally, the fewer features that are used for a character concept, the more general the concept is. Each feature can be seen as a gatekeeper that permits character shapes to be represented by a particular concept. If the concept contains too many features, fewer character shapes comply with it and it becomes more specific.

Such simplification is made particularly difficult by the fact that different people would give different importance to features. What is considered an important aspect of similarity for one person could be unimportant to another person. The results of the studies from chapter 5 illustrate that readers often disagree in their similarity judgements. Clearly, assumptions need to be made regarding what readers perceive, but it is crucial to keep the interpretation as objective as possible and to document the subjective judgements of the person conducting the conceptualization. Moreover, for the work to become transferable to other typefaces and scripts, the process of character conceptualization needs to be as deterministic as possible. Theoretically, one could come up with an indefinite number of ways to represent characters in terms of features (Decock & Douven, 2011). It would be difficult to apply the results of this research beyond the scope of this work if different researchers derived significantly different character concepts from identical character shapes using the same methodology. The following sections discuss the conceptualization to make the process coherent, well documented, and easy to reproduce. The tradeoffs made between the simplicity of a model and the complexity of perception are made explicit where deemed necessary.

In other studies of perceived similarity, it is common to explore the observations by way of data analysis first. Analysed trends in the data are then explained in terms of features. However, descriptions of such features are often the authors' own interpretations. It is rarely the case that they would be described in terms of established typographic practice. Contrary to this, the features used in the model in this research are based on expert knowledge described in chapter 4. However, note that characters are studied as shapes viewed by non-expert readers. Interpretations based on additional insights regarding construction processes, historical developments, or learned conventions are avoided. Those could include, for instance, how a particular shape was written or drawn in the past, or how a particular shape looks in other common typefaces.

Different kinds of features and methods for their discovery are overviewed briefly here. A more detailed discussion can be found in sections 6.3.2 and 6.3.3.

The first method of discovery uses character shapes and their sub-shapes as a close reference. The resulting features are called *parts* and can be considered relatively specific. Parts can be either the more general *strokes* or the more specific *non-strokes*. Linking parts closely with the geometrically-defined character shapes and their sub-shapes helps to keep the process more deterministic and objective. The process of analysing character shapes to define the parts for the corresponding character concepts is called *parts discovery* and is described in section 6.3.2 below.

The second method adds a more general kind of feature, called *roles*[8], by focusing on part relationships and overall character qualities. The roles are relatively general and admittedly, the process of their discovery can be more subjective on the side of the researcher. Roles can be either the more general *character qualities* or the more specific *part composites*. The *roles discovery* process is described in section 6.3.3 below.

To conclude, there are two kinds of features used in the model: roles and parts. Each of these kinds has also two sub-kinds. This creates a hierarchy of features going from more general to more specific (see figure 6.12). As discussed in section 6.1, all of these features are treated as equal by the model, but not necessarily as equally salient.

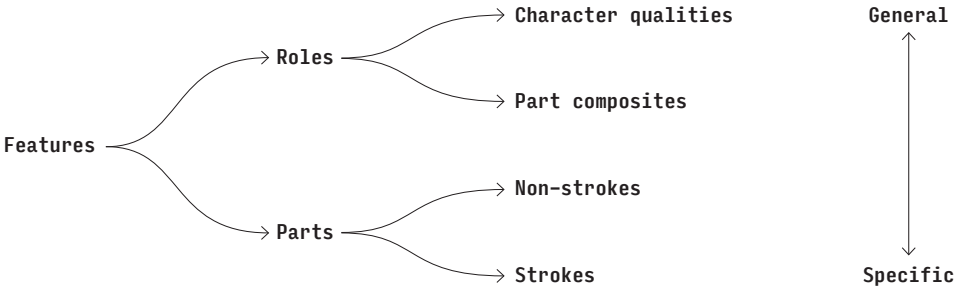


FIGURE 6.12: Hierarchy of features used in the model depending on how specific or general they are.

8. The meaning of the term role is different from that of Hofstadter (see section 4.5.2), although the distinction between parts and roles builds on his work and the work of the Fluid Analogies Research Group (see also section 2.2).

6.3.1. Salience

Not all the features contribute towards (or against) similarity in the same way. Each feature has a different *salience*. According to Tversky (1977), salience is determined by two kinds of factors. The first kind relates to an increase in intensity or signal-to-noise ratio, e.g. the size or brightness of an object. The second, diagnostic kind refers to the classificatory significance of features, e.g. for most classifications, the size of an animal is more significant than the colour of its eyes.

In the model, feature saliences are represented as zero or a positive decimal number. The two factors discussed above are represented as two kinds of saliences: *local saliences* describe the salience of features within a character, and *global saliences* describe the salience of each feature relative to other features in the model. Thus, when trying to make sense of which feature is relatively more important, one has to combine both kinds of salience for a particular feature and character.

The local saliences are set manually based partly on the spatial prominence of the corresponding features. This is explained below in sections 6.3.2 and 6.3.3. The local saliences are fixed in the model. On the other hand, the global saliences are variable in the model. They are initially set randomly and automatically altered by the model training routine described in section 6.5.

6.3.2. Parts

The view taken during the parts discovery is a view of perception, not exact measurement. In this view, two shapes can be different geometrically, but if they are likely to look alike to an average reader, they are considered identical. One could think of this *perceived identity* as a geometric identity with tolerance. This tolerance depends on the reader's vision, attention to detail, their past experience with letterforms, and also on reading conditions, size, and other qualities of the shape. Notably, it may also depend on the mode by which shapes are interpreted (see below). In this particular case, character shapes are analysed as they occurred in the studies, and with little attention to fine detail as per the instructions given to participants (see section 5.2.2).

The process of parts discovery is mostly manual and it is based on character shape decomposition. A part corresponds to one or more perceivably identical character sub-shapes that cannot be further decomposed into sub-shapes associated with other parts. In other words, parts are categories of primitive sub-shapes that are likely to look alike to average readers. Identical parts that appear in two or more character concepts represent shape-based analogies that can be made between character shapes. This approach is supported by theory of recognition-by-components (Palmer, 1977; Biederman, 1987).

As discussed in chapter 4, and leaving aside possible transformations for the moment, elemental shapes can be conceptualized either as strokes or, using their contours, as non-strokes. It has to be emphasized that this distinction is notional. In principle, any shape can be thought of as written (stroke) or not (non-stroke) (see the introduction to chapter 4). From a typeface design perspective, sub-shapes are interpreted as strokes based on convention and convenience (see sections 3.1, 4.2, and 4.4). With consideration to perception, one can notice that what matters is how strokes and non-strokes relate to other strokes and non-strokes, respectively.

It is an important assumption of this work that the perceived identity works differently for strokes and non-strokes. Strokes are considered identical if their midlines are perceivably identical. To make the process of identification simpler and deterministic, the midlines are categorized as vertical, horizontal, diagonal, rounded, and headline[9]. If midlines of two strokes are categorized in the same way, the strokes are considered identical. Note that the pen shape is disregarded here as its effect is considered mostly constant, and thus normative, across character shapes produced using a single high-quality typeface (see section 4.2). The objective is merely to compare strokes, not to re-create them. The length of the midline is represented as a local salience of the corresponding stroke part. On the other hand, non-strokes are related in a more straightforward way. Two non-strokes are considered identical if their overall shapes are perceivably identical. As a result of the simplified midline categorisation, strokes relate to each other in a more general way than non-strokes.

The position of strokes and the position and orientation of non-strokes are disregarded at this point. Another assumption being made is that the affine transformations of translation, rotation, and reflection do not alter sub-shapes in a way that would make them difficult to identify as the same part or change their salience (see section 4.3 and figure 4.10).[10] Parts' positions or orientations that are considered relevant can be implemented by means of roles (see section 6.3.3).

During the parts discovery process, character shapes are decomposed into individual sub-shapes. Sub-shapes that correspond to strokes from simple handwriting models are tentatively interpreted as strokes. These stroke-shapes are separated from each other at junctions, i.e. places where two or more strokes meet, at their thickest and thinnest points, and at points where their

9. Headline stroke is a special category of stroke used to distinguish the headline in Devanagari which usually joins neighbouring characters in their top part (see section 3.3 and figure 3.14).

10. As already mentioned in section 6.1 (in footnote 3), this is not always the case. However, the sub-shapes used in high-quality typefaces for continuous reading seem to be well-formed enough for this not to be an issue.

midline categories change. Distinctive stroke endings are also separated from the strokes. In case of doubt, the simpler division is preferred to one which would lead to more sub-shapes. Visual representations of the concepts in the model are shown in the concept diagrams in figure 6.14 and appendix 7.

Each particular kind of stroke becomes a part[11], thus there is a **vertical**, **horizontal**, **diagonal**, **sweep** (a rounded stroke), and **headline** (see table 6.1 and figure 6.13). In the concept diagrams, the strokes are marked using their approximate midlines which have been produced manually. However, the categorisation of the midlines is automated.

Sub-shapes that are not easily interpreted as strokes, i.e. their stroke representation is too complex considering their relative size, are interpreted as non-strokes. These parts typically include different kinds of stroke endings (or *terminals*), for example: the bilateral horizontal ending of a stroke becomes a **serif**, the unilateral serif becomes a **horizontal half serif**, and the stroke ending with a ball-like shape becomes a **ball**. Other parts often appearing near stroke junctions are also interpreted as non-strokes, e.g. a **knot** or **spur** (see table 6.1 and figure 6.13). Simple cut-off terminals are disregarded as non-salient in the typefaces studied. Some of the diacritical marks, such as the **dot** in the Latin character `i`, are also represented as non-stroke parts. In the concept diagrams, the non-stroke parts are marked at their position using different symbols to keep the diagrams visually simple.

The local salience of the same non-stroke parts can be constant as their scaling or elongation in different characters are designed to be almost imperceptible in high-quality text typefaces. It is set arbitrarily to 1.0. As mentioned above, the local salience of strokes is based on the relative length of their approximate midlines stored in character concepts. The length of each midline is made relative to the base height[12] of the representative character shape. This makes the result independent of scale and allows comparison of strokes across concepts. Character shapes from the same category produced using different typefaces can have different relative stroke lengths. However, the differences were small for the typefaces studied and did not justify creating separate concepts to record each of these length differences precisely. And lastly, although in calligraphy strokes can overlap with one stroke covering a portion of another, in the model only the midlines for the visible portions of the strokes are included (this is visible in the diagram for ‘k’ in figure 6.14 where the midline for the top **diagonal** does not join the midline for the **vertical** as a calligrapher might expect).

11. The names of features are marked in bold for the sake of clarity.

12. See section 3.1 for definition of the base height.

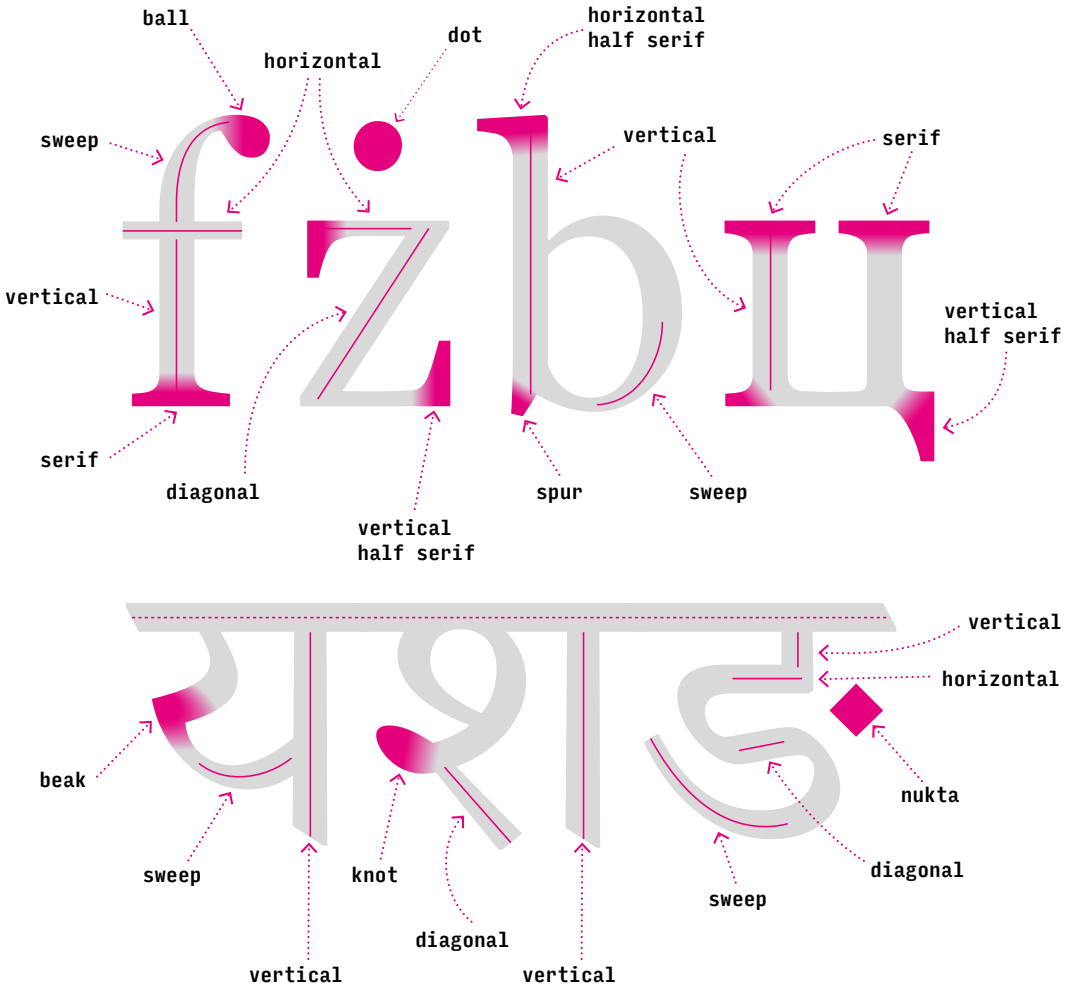


FIGURE 6.13: Sub-shapes corresponding to parts. The non-strokes are marked with magenta shading. Strokes are marked with their midline in magenta. Not all of the strokes from the corresponding concepts are marked here.

Part name	Type	Description
vertical	Stroke	straight vertical stroke
horizontal	Stroke	straight horizontal stroke
diagonal	Stroke	straight diagonal stroke
headline	Stroke	top horizontal stroke joining neighbouring characters in Devanagari
sweep	Stroke	a rounded stroke with modulation going from thin to thick only
serif	Non-stroke	horizontal bilateral orthogonal stroke ending
horizontal half serif	Non-stroke	horizontal unilateral orthogonal stroke ending
vertical half serif	Non-stroke	vertical unilateral orthogonal stroke ending
ball	Non-stroke	bulbous stroke ending
spur	Non-stroke	triangular protruded corner where two strokes meet
knot	Non-stroke	emphasized stroke ending or junction where the direction changes abruptly
beak	Non-stroke	sharp corner where two strokes meet
dot	Non-stroke	dot above the character
nukta	Non-stroke	dot next to or below a Devanagari character

TABLE 6.1: Table describing parts used in character concepts for Latin, Cyrillic, and Devanagari. See also figure 6.13.

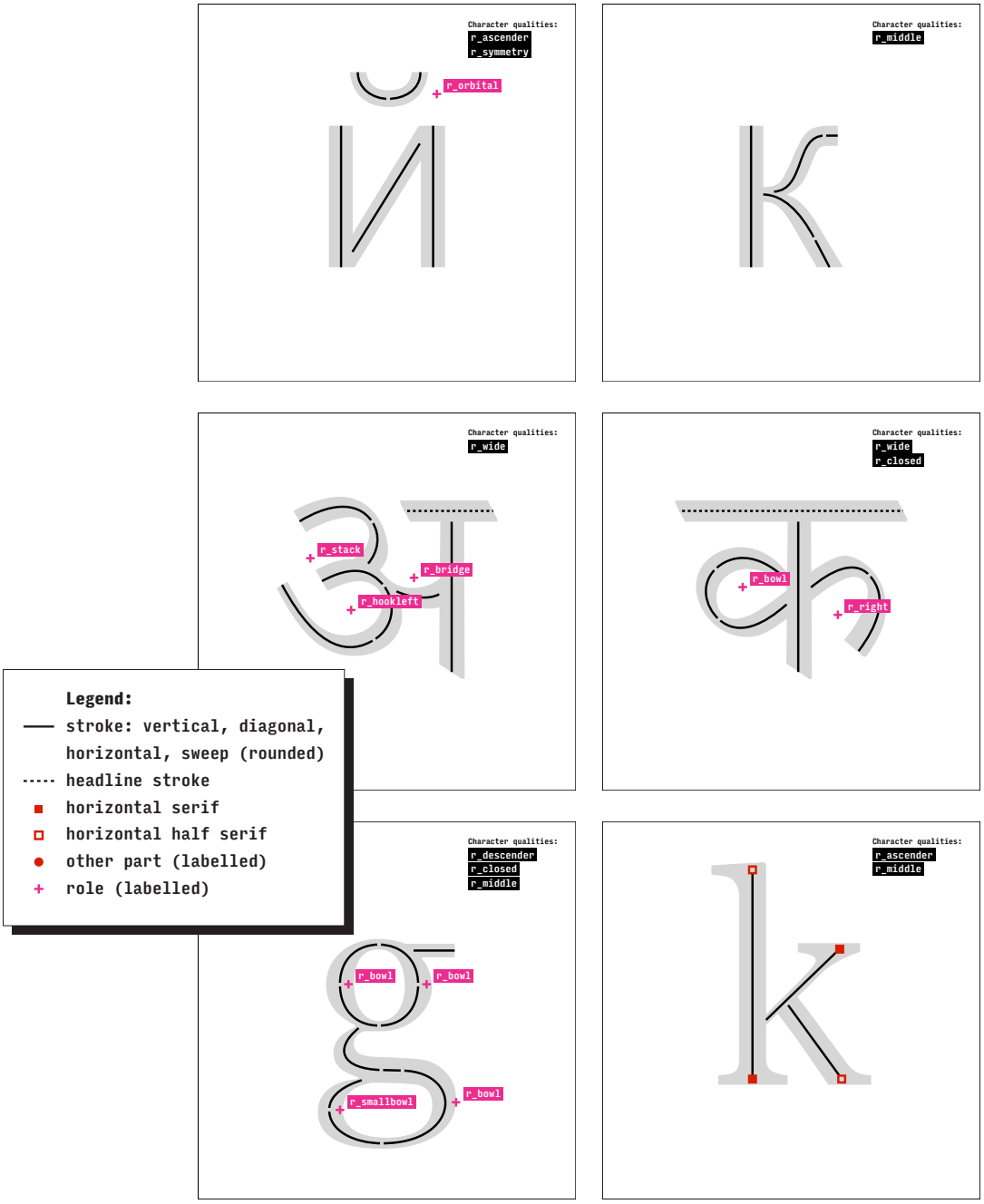


FIGURE 6.14: Some of the concept diagrams from appendix 7 (from top to bottom, left to right): concepts for sans-serif Cyrillic 'й' [short i] and 'к' [ka], high-contrast Devanagari 'अ' [A] and 'क' [Ka] and serif Latin 'g', and 'k'. The stroke parts are marked with their approximate midlines. The non-stroke parts are marked with red symbols, each symbol representing a different part. The diagrams also include roles (see section 6.3.3) that are marked with a cross in magenta with a label.

Conceptualizing characters using parts can lead to tradeoffs between analogy-making and distinction-making. There are some observable differences within the parts discovered this way; the **sweep** part represents perhaps the largest simplification in this case and merits discussion. If there were multiple different **sweeps**, some analogies would not be covered. For example, in figure 6.15, if the **sweep** in 'n' was treated as different from the **sweeps** in 'b, d', it would be more precise, but the analogical roundness of the strokes in these characters would not be represented in the model. In the current solution, the **sweeps** in 'b, d, n' are treated as identical, which is not precise, but this solution maintains the overall analogy of rounded strokes. Some of these kinds of tradeoffs are addressed by roles discussed in the next section.

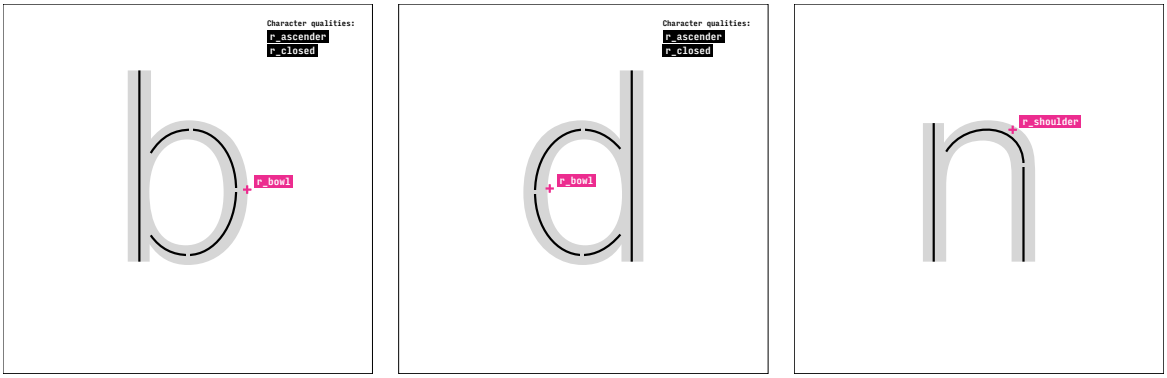


FIGURE 6.15: Concept diagrams for sans-serif Latin 'b, d, n' illustrate that there are two kinds of visual analogies based on the **sweeps** in these characters. On the one hand the **sweep** parts in 'b' and 'd' seem very close compared to the **sweep** in 'n', but there is also a sense of roundness in the **sweeps** in the three characters. Instead of treating the **sweeps** in 'b, d' and 'n' as different parts, they are treated as the same part and an additional role **r_bowl** is added to 'b' and 'd' to express their closer relationship (they are forming a bowl in both characters).

Although it has not been done in this work, parts can be derived not only from the positive (printed), but also from the negative (non-printed) sub-shapes of character shapes, e.g. their counters. Moreover, boundaries of parts can be fuzzy in that they may not have all of their edges defined clearly, e.g. strokes connected to other strokes, or non-printed spaces within character shapes connected to the outside area of the character shape as in open counters (see figure 4.3). Such a fuzzy definition of a shape is not a problem, as the intention is to create concepts that allow the modeling of perceived similarities, rather than create precise descriptions of character shapes.

The names of the features, parts as well as roles, such as **spur** or **serif** may be familiar as they often come from common type-anatomy nomenclatures (Menhart, 1957; Gaskell, 1976; McLean, 1992; Baines & Haslam, 2005; Pecina & Březina, 2008; Naik 1971; Dalvi 2009). Some of these are presented in section 4.4. However, their meaning is not intended to be purely descriptive.

They are convenient labels used for easier reference. Conclusions regarding shape similarity should not be made based on the similarity of the feature names, e.g. one should not assume that parts **serif** and **vertical half serif** represent similar shapes.

If the model were not used to hypothesize about expert knowledge and its ability to predict readers' similarity judgements, the character conceptualization could be automated. The perceived identity could be, for example, defined as a geometric identity with tolerance and used to find identical sub-shapes across all conceptualized characters. Such a process could lead to a different set of parts than those based on the expert knowledge presented in table 6.1, but it would speed up the conceptualization significantly and make the analysis more transferable. Note, however, that the people's different preferences in weighting mentioned in the introduction to section 6.3 would still remain an issue. Only this time, the interpretation would be expressed through an algorithm. One such process, an iterative decomposition into parts, is briefly described below, but it was not implemented in this research. An approach like this was applied on 3D-object models by Lun et al. (2015) where the models were automatically decomposed into individual 3D-part models which were then used to calculate a measure of similarity between the whole 3D-object models.

At the beginning of the process, the representative characters for each concept studied are put together; each of them as a whole shape. When two or more shapes contain a perceivably identical sub-shape, they are split to produce the shared sub-shape and the remaining sub-shapes that are left after removal of the shared sub-shape from the original shape. This splitting is repeated until no further sub-shapes can be found; that is, all resulting shapes are primitive or further division would lead to too small, imperceptible shapes.

At this point, the shape decomposition process is complete, and the set is consolidated. All identical shapes are replaced by a single representative shape. As a result, the consolidated set contains only unique shapes. Each shape is given a name and acts as a potential character part (see figure 6.16). And finally, the character concepts are composed using only the parts from the consolidated set that correspond to sub-shapes of the original character shape.

This iterative decomposition has a significant disadvantage. If new character shapes were to be added to the studied set, the whole set would need to be re-analysed. This is called an "extension effect" by Tversky (1977). The addition of new shapes may cause previously unique sub-shapes to become relevant as parts through their repetition in the newly added shapes. In the proposed model and discovery methods, this is alleviated by reliance on expert knowledge that is independent of the shapes that are being conceptualized. Although introduction of new shapes may still require new parts or roles, the existing concepts are considerably more stable than if iterative decomposition was used.

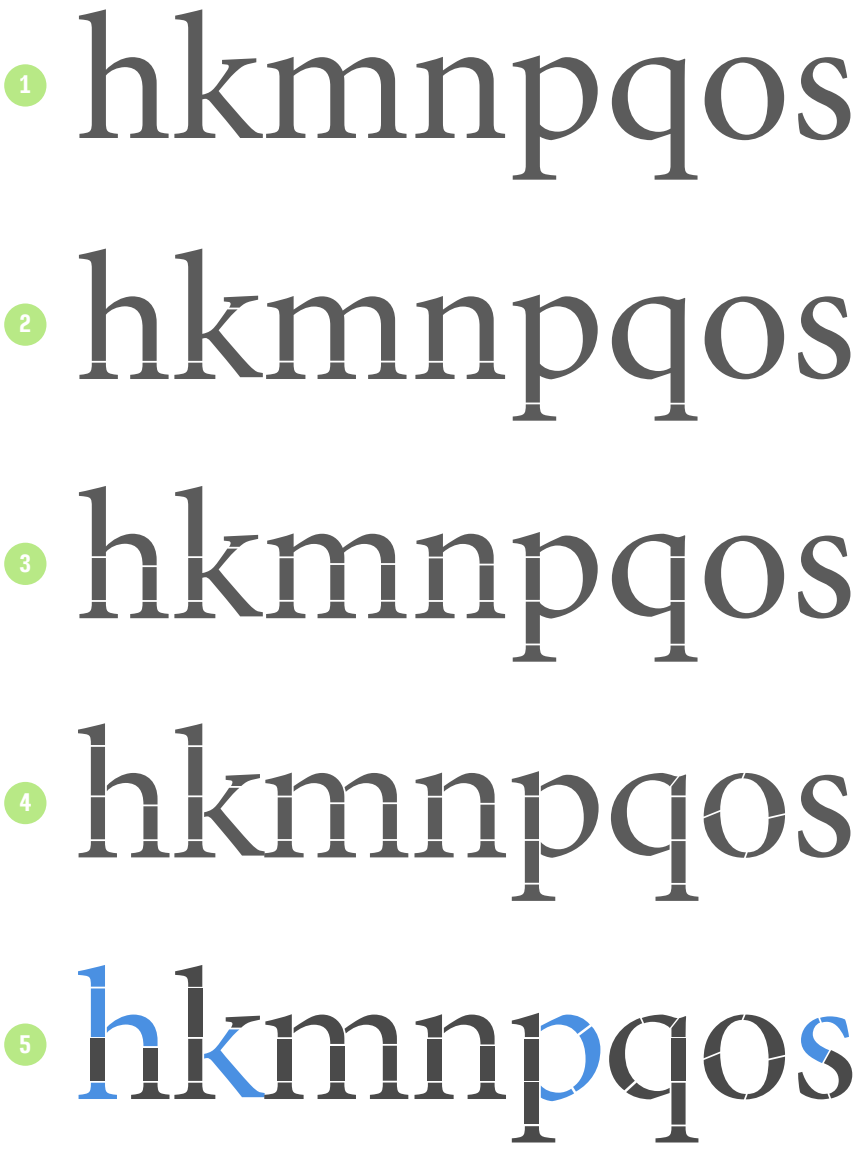


FIGURE 6.16: The decomposition process on a few selected characters with iterations going from top to bottom. The consolidated set of parts is marked in blue.

6.3.3. Roles

As mentioned above, the shape decomposition used to discover parts also removes some observable holistic qualities associated with combinations of parts or with whole character shapes. These could include mutual relationships of parts as well as overall relationships between character shapes, i.e. their overall qualities. Parts, or their corresponding sub-shapes to be precise, could join to form a larger composite that repeats across different character shapes. Thus, although selectively ignored during the part decomposition, their relative position and orientation could become relevant in some situations as well. Moreover, the relationships of the whole characters could be related to position or their relative size as well. For example, characters could be categorized as ascending above others or as wider than others. These kinds of qualities might influence readers' similarity judgements and therefore should be represented in the model.

Theories suggesting the existence of more general features based on relations between parts or even the general features themselves have been proposed by researchers in different fields (Palmer, 1978; Medin, Goldstone & Gentner, 1990). Proposals for more general features were made in relation to character recognition (Blesser et al., 1973) and in cognitive science and typeface design (see section 4.5.2). Common typographic attributes can be seen as general features as well (see section 4.5.1).

The general features are implemented in the proposed model as roles. It is useful to think about them in terms of grouping. Once there is a group of two or more characters with a common visual trait that is not based on a single part and that can be used by participants in their similarity judgements, a role should be introduced. Naturally, there is no need to introduce roles that would be equally salient in all characters studied or that are present in only a single character.

For example, in a trial with the triplet `b, d, p`, some participants might make their judgement based on the the fact that `b, d` ascend above the base height area, while `p` descends below.[13] Other participants might focus on the fact that `b, p` have a bowl made of sweeps on the right side of their main verticals while `d` has the bowl on the left side. Neither of these judgements can be explained by the mere presence of parts. Their explanation requires an introduction of additional roles. Roles **r_ascender** and **r_descender** can be used to express that `b, d` ascend above, and `p` descends below the base height area. The role **r_leftbowl** could be used in the concept for `d` to express the sweeps forming the bowl are positioned to the left of the main vertical unlike the sweeps in `b, p` which are oriented to the right. Note that the potential role **r_rightbowl** is not needed, as the orientation to the right is considered

13. In other words, `b, d` ascend above and `p` descends below characters like `m, n, x`.

normative in Latin and Cyrillic. In Devanagari, the orientation to the left is normative, thus there is a use for role **r_right**. A role or multiple roles are introduced only where there is an observable deviation from the norm defined by the majority of characters.

Role names start with “r_” to make them distinct from parts. Each role has a name and a description. The descriptions are used as criteria to assign them to concepts. If a character shape or a group of parts satisfy the role description, the role is assigned to the corresponding concept. Table 6.2 and figure 6.17 describe roles based on part composites; table 6.3 and figure 6.18 describe roles based on character qualities. These roles were used in the concept diagrams in figure 6.14 and appendix 7. Although theoretically possible, not all of the roles have to be used for all the scripts. For example, there is no need for **r_orbital** in Latin if the only orbital sign appearing in the set of character shapes studied was a dot accent and it was already represented as a part.

FIGURE 6.17: Diagram showing roles based on part composites from table 6.2.

Role name	Description	Cyrillic	Devanagari	Latin
r_bowl	multiple sweeps enclosing a counter	●	●	●
r_shoulder	a sweep and a vertical forming a shoulder-like shape	●		●
r_orbital	one or more parts forming an orbital sign	●		
r_smallbowl	sweeps forming a small bowl (about half the height of the r_bowl)	●		●
r_bigbowl	larger kind of r_bowl		●	
r_stack	two bowl-like sub-shapes stacked vertically		●	
r_loop	sweeps forming a small, closed or nearly closed loop		●	
r_hookdown	crescent-like shape with the stroke ending aiming downwards		●	
r_hookleft	crescent-like shape with the stroke ending aiming to the left		●	
r_angularneck	angular combination of a short vertical and horizontal or sweep connecting to the headline		●	
r_tail	stroke starting in a knot part		●	
r_zigzag	sweeps abruptly changing direction to form a zigzag shape		●	
r_ra	parts forming the Devanagari character Ra as they repeat in other characters		●	
r_right	sweeps on the right side of the main vertical in Devanagari		●	
r_bridge	short horizontal or sweep connecting two character sub-shapes		●	

TABLE 6.2: Table describing roles based on parts composites. The last three columns specify the scripts for which the role is being used.

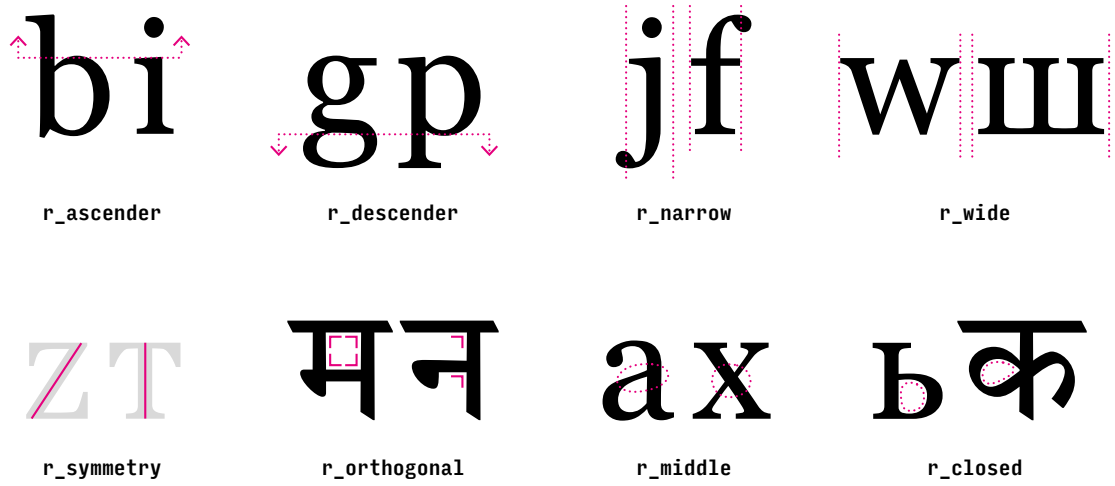


FIGURE 6.18: Diagram showing roles based on character qualities from table 6.3.

Role name	Description	Cyrillic	Devanagari	Latin
r_ascender*	one or more parts of the character ascend above base height	●		●
r_descender*	one or more parts of the character descend below baseline	●		●
r_narrow*	corresponding representative character is among the 20% of narrowest representative characters			●
r_wide*	corresponding representative character is among the 15% (Cyrillic, Latin) or 25% (Devanagari) of widest representative characters	●	●	●
r_symmetry	the configuration of strokes in the character is symmetrical (rotation or reflection symmetry)	●		●
r_orthogonal	stroke parts in the character are mostly joining at right angles		●	
r_middle	stroke part(s) are crossing the middle of the base-height area of the character	●		●
r_closed	character contains a closed counter	●	●	●

TABLE 6.3: Table describing roles based on character qualities. Roles marked with an asterisk are based on an automatic analysis of representative character shapes. The last three columns specify the scripts for which the role is being used.

Note that there is more than one way to divide or juxtapose the roles to categorise the character shapes studied and it seems to be possible to repeatedly come up with new roles. However, introducing a high number of roles would affect the model's parsimony. As features correspond to the model's parameters (see section 6.5), there would be too many parameters to interpret. Thus, the number of roles is kept relatively small.

An alternative way to support a need for a particular role is the analysis of the plots of character clusters from section 5.13. Although they are a simplification, the plots can help spot emerging clusters of character shapes that can then be explained in terms of their shared features (see figure 6.19).

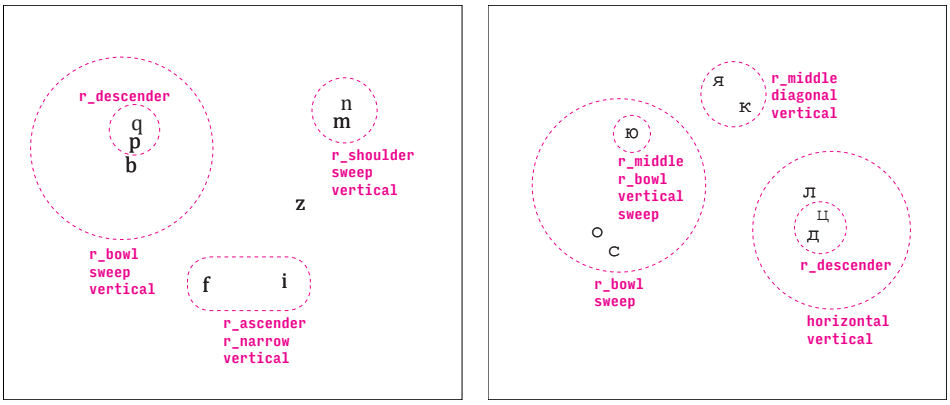


FIGURE 6.19: Left to right: plots of character clusters for typefaces Cambria (Latin) and Courier New (Cyrillic) with some of the clusters (marked with dashed circles) interpreted in terms of their shared features. The distance of the characters in the plots corresponds to their generalized similarity measure derived from the studies. Section 5.13 discusses the way the character cluster plots were made.

The majority of the roles have to be assigned manually using their descriptions as criteria. While admittedly more subjective than the parts discovery, the role discovery process is based on expert knowledge as it builds on parts previously discovered and typographic attributes discussed in section 4.5.1.

Some roles can be assigned to concepts automatically (they are marked with an asterisk in table 6.3) based on an analysis of the corresponding representative character shapes. Four roles were created this way. Roles **r_ascender** and **r_descender** signify that a character shape ascends above or descends below the base height, which is the normal height of most lowercase characters from the Latin and Cyrillic scripts. The salience of these roles corresponds to the proportion of the height of the ascender or descender to the whole ascender or descender area respectively. For example, the salience of **r_ascender** in `h` is 1.0 while the salience in `t` could be around 0.5, as the ascender in `t` is shorter in the representative character for its concept. These two roles are not relevant for any of the concepts for the Devanagari characters.

The roles **r_narrow** and **r_wide** are assigned based on the width of the representative characters. For Latin and Cyrillic, the role **r_wide** is assigned to a concept if its representative character is among the 15% of widest representative characters. For Devanagari, it has to be among the 25%. The role **r_narrow** is assigned to a concept if its representative character is among the 20% of narrowest representative characters. The role **r_narrow** is not used for Cyrillic and Devanagari. In Devanagari there is only one character that is significantly narrower than others (र [Ra]) and in Cyrillic the few narrow characters are nearly as wide as characters with two **verticals**. Those characters can be considered normative.

6.4. Concepts' similarity

Two approaches to calculating similarity have been used to query how participants use the context in their similarity judgements. Both approaches are presented below. The *generalized approach* can be considered context independent, while the *contextual approach* is, as the name suggests, context sensitive.

6.4.1. Generalized approach

In a feature-based approach, the similarity measure of two character concepts is a combination of the total salience of their shared features and the total salience of their distinct features. The question is how to combine these two sets of features and their saliences to produce a useful measure of similarity. The contrast model suggested by Tversky (1977) uses a linear combination.[14] The generalized approach described below is based on a simpler form of Tversky's formula.

The formula for the generalized pair similarity measure is defined below (see the empirical generalized character pair similarity measure in section 5.13). Note that in all the definitions in this section, A, B, C, and D represent character concepts from the model. The components of the formula are defined first:

- **sharedAB:** the sum of the shared salience for features that are in both A and B
- **totalAB:** the sum of the salience for all features in A and B

14. Other common similarity metrics have been tested during the development of this work, e.g. similarities based on cosine, sigmoid, or chi-squared functions. The linear approach showed the best performance when training the model, and it is easier to interpret.

Note that even though A and B might share the same feature, this feature can have different salience in each of these concepts. If the feature is less salient in one concept, the result should be the smaller salience. Hence the term *shared salience*. For example, r typically contains a less salient **vertical** feature than h. The correct result would be twice the salience of a **vertical** in r, once for the feature in r and once for the feature in h.[15]

Thus, the similarity measure of concepts A and B is calculated as:

$$s(A, B) = \frac{\text{sharedAB}}{\text{totalAB}}$$

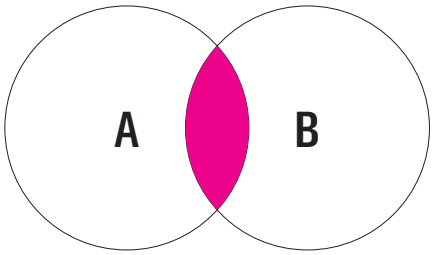


FIGURE 6.20: The Venn diagram shows the set of features in the concepts A (the area of the circle on the left) and B (the area of the circle on the right), and the set of features shared by concepts A and B (the area marked in magenta).

Compare the above definitions and the formula with the Venn diagram (see figure 6.20). It is important to note that the Venn diagram shows sets of features, but the formula operates with feature saliences.

Both the generalized as well as the contextual method described below assume that participants compare all pairs of character shapes in the trial before they respond. Thus, for a triplet of characters conceptualized as A, B, C, they would assess s(A, B), s(A, C), and s(B, C) and keep the pair that is most similar while marking the remaining character as the odd one out. Thus, the probability estimate of C being the odd one out corresponds to s(A, B).

The *generalized predictive method* of the model uses the above formula directly. It calculates the estimate for each concept in a trial, and then makes them all relative and normalized. The three resulting probability estimates are decimal numbers between 0.0 and 1.0 that add up to 1.0. The higher the number the higher the probability of the corresponding character being considered the odd one out.

15. While this may seem intuitive, there are other ways to calculate the sum of shared salience, for example, by multiplying the salience of the shared features in the two characters. These have been tested, but did not outperform the approaches described here.

Note that the generalized similarity measure s(A, B) in trials with character triplets conceptualized as A, B, C and A, B, D is the same, as the third concept does not influence the calculation. However, it does not mean that the probability of the C or D being the odd one out will be the same in the predictions for these two trials, because the individual similarity measures in the triplet are made relative and normalized, i.e. divided by their total sum.

It is also worth noting that if A and B do not have any features in common, their similarity is 0.0. Should this happen for all pair combinations in the triplet, the model cannot provide a prediction, and the triplet of concepts is considered uncomparable. In these cases, the predictive function responds with a prediction where all three responses are equally likely, i.e. f_{iox} = (0.3333, 0.3333, 0.3333).

Note that there are not any uncomparable concepts among the concepts for each script in the model; every pair of concepts has at least one feature in common. However, the predictive function responds with the same prediction when all pairs in the triplet have the same similarity measure. This could happen for triplets like 'i, o, x' in a sans-serif typeface like Calibri, where the corresponding concepts have only one feature in common, **r_symmetry**, and its local salience is identical in all three concepts.

6.4.2. Contextual approach

The *contextual approach* to concept similarity recognizes the influence of the third character concept on the similarity judgement of A and B and uses a context-dependent pair similarity measure to calculate the probability of C being the odd one out (see the empirical specific character pair similarity measure in section 5.12). Again, to make the explanation simpler, the components of the formula are presented first:

- **sharedAB**: the sum of the shared salience for features that are both in A and B
- **uniqueAB**: the sum of the shared salience for features that are both in A and B, but not in C
- **uniqueC**: the sum of the salience for features that are in C, but not in A or B
- **totalABC**: the sum of the salience for all features in A and B and C

The pair similarity measure of A and B in the context of C is calculated as:

$$s_c(A, B) = \frac{\text{sharedAB} + \text{uniqueAB} + \text{uniqueC}}{\text{totalABC}}$$

The formula attempts to model two component judgements the participants have to make when faced with a triplet A, B, C:

i. how similar is A to B: sharedAB / totalABC
ii. how different is C from both A and B: (uniqueC + uniqueAB) / totalABC

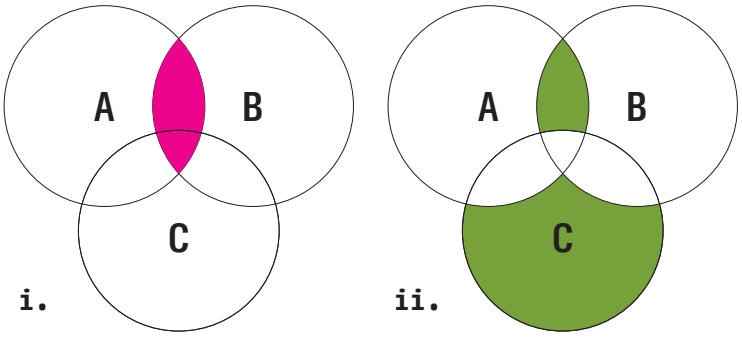


FIGURE 6.21: Each of the Venn diagrams shows: the set of features in the concept A (the area of the circles on the left), B (the area of the circles on the right), and C (the area of the bottom circles). The first diagram corresponds to the first component judgement (marked i. in the text above) and shows the set of features shared between A and B (the area marked in magenta) in comparison to all the features in A, B, and C (the total area of the three circles). The second diagram corresponds to the second component judgement (marked ii. in the text above) and shows the set of features that make C different from the shared features of A and B (the area marked in green) in comparison to all features in A, B, and C (the total area of the three circles).

The Venn diagram in figure 6.21 shows the sets of features used in the contextual approach. Again, note that the formula does not calculate the shared areas in the diagrams, but the shared saliences of the concepts. It may seem odd at first that the third concept, C, could somehow contribute positively towards the similarity of the other two, but it is the unique features in C that are not present in either A or B that can make C seem more different and therefore make A and B look more similar. Analogically, it is the unique features in both A and B that are not in C that make A and B more similar. Also note that the total of the features' salience is from all three character concepts. The *contextual predictive method* uses the contextual pair similarity measures as probability estimates of the contextual concepts being the odd one out. The method makes the values relative and normalized in the same way as was described for the generalized predictive method above.

Potentially fewer uncomparable trials result from this method than from the generalized predictive method (see section 6.4.1), as this method considers the contextual concepts' unique features as well.

It is important to note that from the probabilistic perspective, both of the approaches described above are naive. They assume that all features are independent. However, this may not always be the case. For example, the role **r_ascender** is usually represented with a **vertical** part, thus the two features

are dependent. This is something that ought to be considered during role discovery, but may not always be possible to resolve. Although not perfect, the naive approach gives satisfactory predictions, which justifies its use.

6.5. Model training

The local saliences are set during parts and roles discovery. On the other hand, the global saliences can only be set randomly, as it is not clear in advance which of the features are more or less salient for participants' perception of similarity. The global saliences are the parameters of the model and they are learned from the observations through a process called *supervised training*. The principle is simple: the model training routine iteratively proposes different global saliences and calculates predictions. These predictions are compared to the observations which are provided as an example to replicate. When the new predictions are better than the previous predictions, the currently proposed set of global saliences is adopted. Note that all of the observations, or more precisely the observed response frequencies, used in the training as examples are treated as equal even though they might have been derived from different counts of participant responses.

In order to evaluate the predictions, the program requires a *loss function* that receives predictions and observations on input and returns a single decimal number. This number decreases as the predicted and observed values get closer. The loss function used in the model is a *mean squared error* (L2 norm).[16] The squared differences between the observed and predicted values for the corresponding characters from the trial are calculated first. The mean of these squared differences gives the error for each trial (see figure 6.22). The mean of the trial errors gives an overall error. This overall error is used as the loss function.

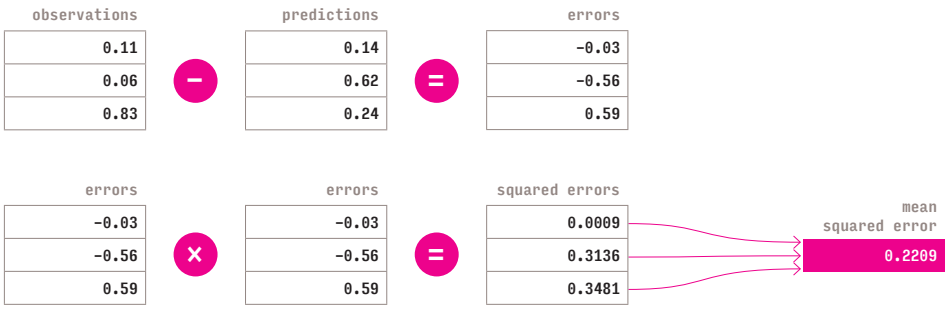


FIGURE 6.22: The process of calculating the mean squared error for a single trial: the differences (errors) between observations and predictions are squared. Their mean gives the mean squared error.

16. Absolute error (L1 norm) produces similar results, but its variance among different trainings was greater than with the mean squared error, which was therefore used.

Another important aspect is how the training routine proposes new global saliences in each iteration. The algorithm used is a version of a *Stochastic gradient descent* (SGD) (Bottou, Curtis & Nocedal, 2016). It modifies the previous global saliences by a small step in a direction that corresponds to the decreasing direction of the loss function, and proposes these modified global saliences as a new alternative. The positivity of feature saliences was enforced during this process as well. When new global saliences were proposed, any negative saliences were converted to zeroes.

To optimize the speed, SGD does not work on the whole set of observations, but divides them into small subsets (batches) and suggests new global saliences for them one by one. This process is repeated for a fixed number of iterations (epochs). The following SGD parameters gave the best results with the model:

- number of total iterations (epochs): 1000
- step by which the saliences are modified: 0.01
- number of observations in each batch: 40 triplet trials

6.6. Technical implementation

As mentioned above, the model consists of three main components:

i. a character concepts database,
ii. typeface mappings, and
iii. a predictive method.

6.6.1. File storage

The character concepts database is stored in Unified Font Object (UFO) files (Leming, van Rossum & van Blokland, 2004). UFO is a non-proprietary, widely supported font format which allows convenient construction of concepts in common font editors such as Glyphs (Seifert & Scheichelbauer, 2016), RoboFont (Berlaen, 2018), or FontLab VI (FontLab, 2017). Individual concepts are stored as glyphs in the UFO file. The glyphs have arbitrary, unique names which identify the concepts as well as the corresponding scripts, e.g. “latn.serif.b”, “latn.sans.b.spurless”, “deva.high-contrast.La”.

An example of a glyph with a character concept is in figure 6.23. Within each concept glyph, stroke parts are represented by manually made bézier paths that approximate strokes’ midlines. Non-stroke parts and roles are represented by anchor points named after the corresponding parts. If there are multiple non-stroke parts or roles, the anchor names get a numeric suffix to disambiguate between them, e.g. “r_shoulder_0, r_shoulder_1”. The representative characters for each concept are stored in a separate UFO file. The diagrams from figure 6.14 and appendix 7 are produced automatically from these UFO files using DrawBot (Rossum, Blokland & Berlaen, 2007).

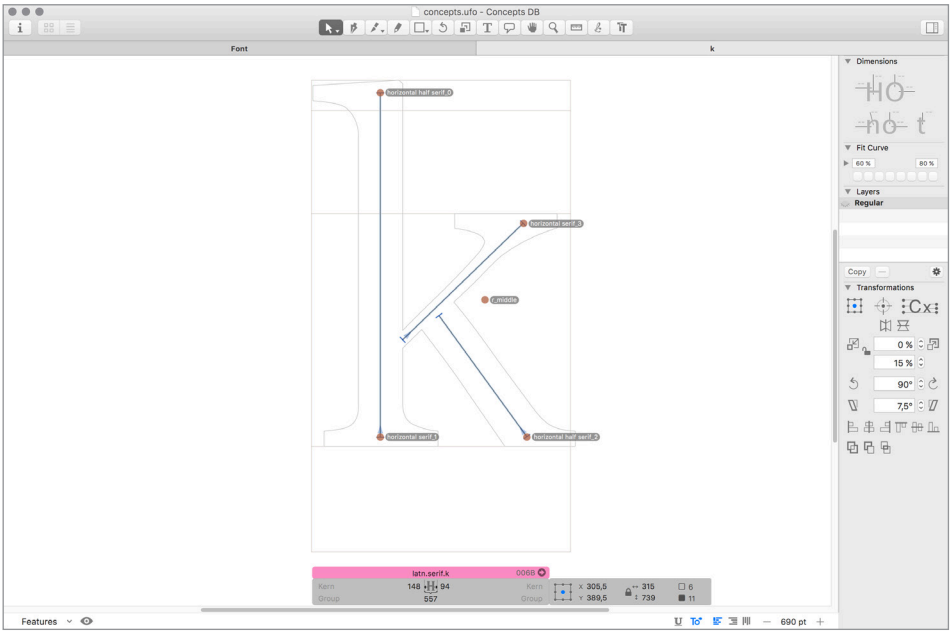


FIGURE 6.23: Character concept in a UFO file viewed here in the application Glyphs, the resulting diagram generated automatically from this concept can be seen in figure 6.14. The contour of the representative character shape from typeface Georgia is in the background for convenience. Otherwise, it is stored in a separate UFO file.

The typeface mappings between characters from a typeface and concepts in the model are stored in YAML files (Ben-Kiki, Evans & Ingerson, 2004).[17] One file is used for each typeface. Each character name is associated with one concept name. Note that the character names used in the model and to save observation data are ASCII-based and have prefixes that identify them as belonging to a specific script, e.g. "latn.a", "cyrl.a", and "deva.a" are three different characters. This is to make them unique and easier to use. Typing the names using three different keyboards, one for each script, would be inconvenient. The YAML files have the following structure. All of the typeface mappings are in appendix 8.

```
typeface:
  name: Arial
  characters:
    studied:
        latn.a: latn.sans.a
        latn.e: latn.sans.e
        latn.l: latn.sans.l
        latn.p: latn.sans.p
        ...
```

17. YAML is a plain text document format for easy specification of structured data.

6.6.2. Data structures

The model is implemented in Python programming language (Python Software Foundation, 2018) as a class called Model. When creating an instance of this class, one can specify:

i. which typeface mappings should be loaded,
ii. script the modelling will focus on,
iii. whether parts should be used as features,
iv. whether roles should be used as features,
v. whether there are some features that should be ignored (this has been only used for testing),
vi. a method used to make predictions, e.g. generalized or contextual.

The model uses the typeface mappings and specified script to find the character concepts required and as it loads them from the concepts database, it also automatically calculates the salience of stroke parts based on the length of each midline's bézier paths. The roles based on the analysis of the representative characters for each concept are automatically created at this point as well (see section 6.3.3 for the criteria used). Once this is finished, the model is ready for training and making predictions.

The calculations of predictions, training, and the model assessment are implemented in Python using vector math with NumPy and SciPy (Jones, Oliphant, Peterson et al., 2001), Pandas (McKinney, 2010), and Sci-kit learn (Pedregosa et al., 2011) libraries.[18] The local saliences are stored in a labelled matrix with concept names in the column index, and a list of all features in the rows index. Parts and roles are listed together. Each column contains a *concept vector* with local saliences of features from the corresponding concept. The salience is 0.0 if the corresponding feature is not present in the concept. The salience can also be higher than 1.0 when a stroke part is longer than a relevant base height or when a feature repeats. When a feature repeats multiple times in the same concept, the total sum of saliences for the individual feature instances is used. The global saliences are saved as an independent vector labelled with feature names. It is called *global-saliences vector* below for the sake of brevity. See figure 6.24 for the overview of the main data structures and their relationship to the concept diagrams. The code for the Model class and the training routine is included in appendix 9.

The observations are represented as a labelled matrix as well. The columns are indexed with a combination of a script name, typeface name, and triplet of character names. Each column contains response frequencies in the order corresponding to the order of characters in the triplet in the column index (see table 6.4).

18. The matrices are represented as Pandas DataFrames, the vectors as Pandas Series.

latin		cyrillic
times-new-roman	verdana	pt-serif
(latn.a, latn.m, latn.u)	(latn.u, latn.w, latn.x)	(cyrl.e, cyrl.er, cyrl.ha)
0.75	0.63	0.20
0.23	0.10	0.17
0.02	0.27	0.63

TABLE 6.4: A few columns from the matrix with observations are shown as an example of the format. The three-level column index is based on a script, typeface, and an alphabetically-ordered triplet of character names. The three response frequencies are ordered vertically in the same way as the triplet. Note that the alphabetic order is always based on the Latin alphabet as the names are in the Latin script. The script level of the index exists only to conveniently filter the observations as the character names are already script-specific.

6.6.3. Calculating predictions

When calculating predictions for a triplet, the model maps the character names from a particular typeface to the corresponding concepts and obtains a triplet of concept vectors (see figure 6.24). Calculating predictions is a two-step process. The concept vectors are first combined as specified by the predictive method used. The result of this step is a *comparison vector* which specifies which features and how locally salient they are in this comparison. In the second step, the predicted probability is calculated as the dot product[19] of the comparison vector and the global-saliences vector.

The generalized predictive method calculates the comparison vector of two concepts as a minimum of the two concept vectors multiplied by two. This makes sense, if a feature is not present in one of the concepts, the resulting salience for this feature should be zero. The multiplication by two is necessary as the shared features appear in both concepts. As discussed in section 6.4, the third contextual character concept from a trial is ignored in the generalized approach.

19. The dot product of two vectors $x = (1, 2, 0)$ and $y = (2, 3, 4)$ is the sum of their item-wise multiplication: $x \cdot y = 1 \times 2 + 2 \times 3 + 0 \times 4 = 8$.

FIGURE 6.24: The diagram shows the process of assigning concept vectors to a particular trial described with a typeface name and triplet of character category names on input. Typeface mappings prescribe which concept should be used with which character category for a particular typeface. Having obtained the concept names from the relevant typeface mapping for a particular typeface (typeface Verdana in this case), the corresponding concept vectors are found in a matrix with all concept vectors (only a small portion is shown here). The corresponding concept diagrams are shown at the bottom for comparison. The concept vectors include a positive number if a particular feature is present and a zero if the feature does not exist in the concept.

The following function receives three concept vectors[20] and returns the comparison vector. Although it may not be immediately apparent, the function uses vector math with NumPy. Note that the example code has been simplified for the sake of clarity.[21] The code used to calculate the presented results is included in appendix 9.

```
def compare_concepts_generalized(a, b, c):
  sharedAB = 2 * np.minimum(a, b)
  totalAB = a + b
  return (sharedAB / totalAB)
```

The comparison vector of two concepts and a third contextual concept using the contextual predictive method is implemented in the following function. It follows the description from section 6.4 and uses boolean math to zero out features that are not present in particular characters.

```
def compare_concepts_contextual(a, b, c):
  # saliences for features in A and B
  sharedAB = 2 * np.minimum(a, b)
  # saliences for features in A and B that are not in C
  uniqueAB = sharedAB * (c == 0)
  # saliences for features in C that are not in A or B
  uniqueC = c * (a == 0) * (b == 0)
  totalABC = a + b + c
  return (sharedAB + uniqueAB + uniqueC) / totalABC
```

A function that calculates a prediction for a particular trial receives a typeface name and triplet of character names. The name of the script is not essential in this case as character names are unique across scripts. The function obtains the concept vectors and calculates the comparison vectors using one of the predictive methods. Afterwards, it takes the dot product of each comparison vector and global-saliences vector to obtain the character pair similarities (see figure 6.25). The similarities are then normalized and taken for predicted response probabilities. When the total sum of the probabilities is zero, it means that the individual concepts are uncomparable using the predictive approach used. In that case, the returned prediction is (0.3333, 0.3333, 0.3333), i.e. a completely undecided response. The following code is for the contextual predictive method. The code for the generalized predictive method would be the same.

20. This is to keep the interface consistent with the contextual predictive method. Only the first two vectors are used.
21. Error handling code and assertions have been removed, calculations have been split into steps that correspond to the components from section 6.4. Optional interfaces of some functions used for testing have been removed.

```
def get_triplet_prediction(typeface_name, triplet_names):
    concept_saliences = []
    for cn in triplet_names:
        # find concept names for the character names
        ccn = self.typefaces[typeface_name].characters[cn]
        # find concept vectors for the concept names
        cc = np.array(self.concepts_matrix[ccn])
        concept_saliences.append(cc)
    a, b, c = concept_saliences
    # get comparison vectors
    A = self.compare_concepts_contextual(b, c, a)
    B = self.compare_concepts_contextual(a, c, b)
    C = self.compare_concepts_contextual(a, b, c)
    # dot of concept vectors with the global-saliences vector
    A = A.dot(self.gsaliences)
    B = B.dot(self.gsaliences)
    C = C.dot(self.gsaliences)
    # normalize frequencies within the whole triplet
    total = A + B + C
    if total == 0:
        # uncomparable triplet
        A = B = C = 0.3333
    else:
        A /= total
        B /= total
        C /= total
    return pd.Series([A, B, C])
```

The model training uses the comparison vectors in a similar fashion. The two-step approach makes the training much faster as the concepts' comparison vectors need only be calculated once at the start. The only values that vary are the global saliences. Therefore, only the second step is repeated for each iteration while tuning the global-saliences vector to make the resulting predictions fit the observations.

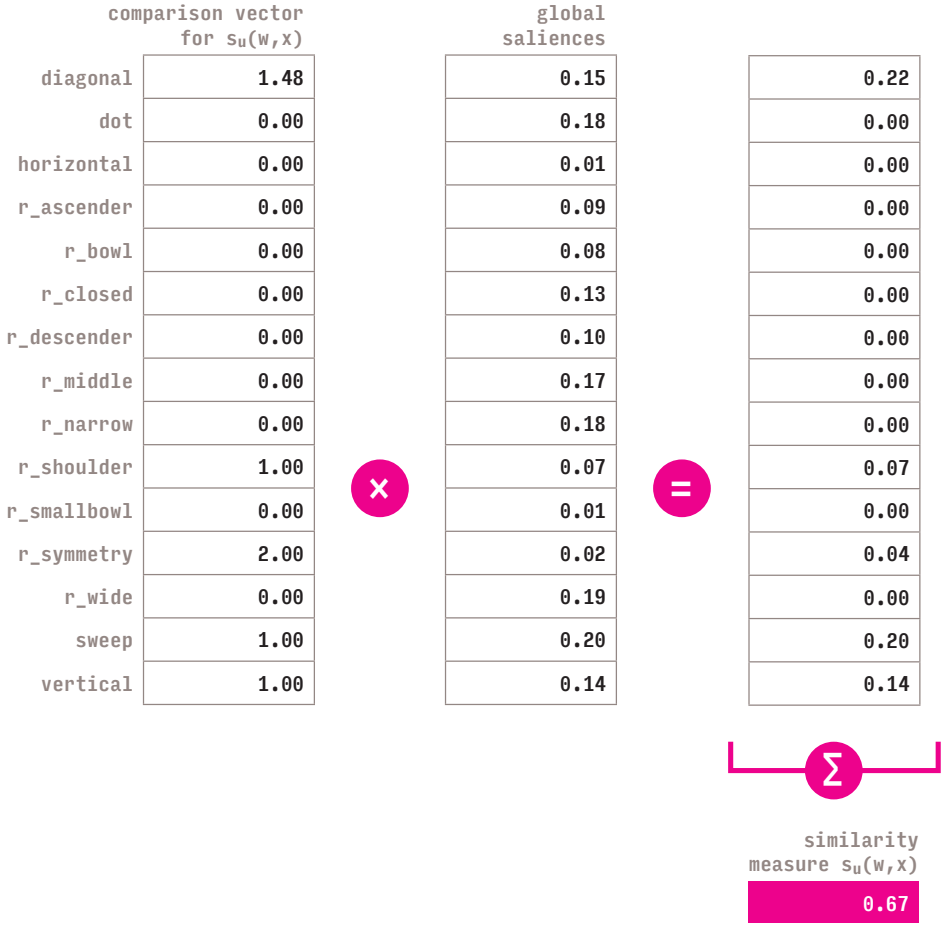


FIGURE 6.25: The similarity measure is calculated as a dot product of the comparison vector and the global-saliences vector. A dot product is a total sum of the item-wise multiplication of two vectors.

6.7. Model assessment

As already mentioned in the introduction to section 6.1, to properly assess a model means to look at its *prediction error*, i.e. how close its predictions are to the observations,[22] but also the number of parameters it uses. Typically, the more parameters the model has, the smaller its prediction error gets during training; however a model with zero prediction error is overfit to the training data and will generalize badly (Friedman, Hastie & Tibshirani, 2001). In an extreme example, a model could have as many parameters as there are responses provided during training and then simply match these parameters to the responses, achieving zero error. Clearly, such a model would end up being very specific and when presented with new, unseen data it would likely fail as it did not generalize any knowledge from the data, but simply replicated it.

22. The data from the studies are treated as the ground truth in the model training and assessment.

To ensure the model is not overfit and that it can be used generally after the training, it is *cross-validated* using some of the observations that were not used during training. Cross-validation tests whether a model can produce good predictions beyond the scope of training. The original, complete data is split randomly into two groups prior to the model training. The training uses 75% of the data while the cross-validation uses 25% of the data. The process of random splitting, training, and cross-validation of each model is repeated multiple times to ensure the results of the cross-validation are general enough. Cross-validating with just one random selection could potentially lead to overly optimistic or pessimistic results if, by chance, the training data contained trials that are significantly more difficult (or easier) to predict than trials from the cross-validation data.

There is also a direct correlation between the number of parameters and the variance of the prediction error. While more parameters decrease the prediction error and increase its variance, fewer parameters typically increase the prediction error and decrease its variance (Friedman, Hastie & Tibshirani, 2001). The goal is to find an optimal number of parameters that help the model achieve the three following objectives at the same time:

i. explain the underlying processes involved in similarity judgements,
ii. have a small variance of the prediction error after different trainings, and
iii. keep the prediction error as small as possible.

This kind of model optimization is usually illustrated by a graph showing a change in the prediction error of the model depending on the change in the number of parameters. Unfortunately, this is not possible with this model, as an arbitrary reduction of the number of features would have random effects on individual trials. The features are assigned manually to the concepts and are sparse, i.e. many concepts have only a few features out of the complete set. Instead, the model has been cross-validated with three different sets of features – parts only, roles only, and roles and parts – and the results have been reported in table 6.5.

The observed and predicted frequencies are compared in several different ways. First of all, the prediction error is represented as a loss (mean squared error) used also during the training (see section 6.5). The smaller the error the closer the model's predictions are to the participants' responses. To illustrate its meaning, it is useful to consider two cases. Possible response frequencies span values between 0.0 and 1.0. If the predictions and observations differed by 0.5 for each character in a trial, the resulting loss would be $0.5 \times 0.5 = 0.25$. If the difference was 0.1 for each character in the trial, the loss would be $0.1 \times 0.1 = 0.01$.

Additionally, a coefficient of determination R^2 (Kmenta, 1986) is provided as a measure of accuracy. It is a proportion of the variance in the predictions that is predictable from the observations. The accuracy is typically a number between 0.0 and 1.0 and the higher it is, the less unexplained variance there is in the predictions.

Alternatively, the model can be assessed as if it was only identifying the odd one out. A comparison between the observed and predicted most popular responses for all trials is provided as a success rate percentage. This indicates how often the model predicts the most popular odd one out correctly, i.e. according to the majority opinion among participants. Nonetheless, it does not account for the model providing the correct response by chance. Therefore, Cohen's kappa statistic is provided as well (see section 5.15.2).

Note that neither the success rate percentage nor Cohen's kappa provide any information about the other, less popular, responses in the trial, and the relative proportion of their probabilities. It only shows whether the most popular response was or was not identical. Consider, for example, that it is less of a problem if the model predicts the overall odd one out wrongly when the two most common responses are very close to each other than if they are very different. To get a better sense of how close all the predictions are to the observations, one has to look at loss and accuracy.

The last important consideration when assessing the model is the variance of its parameters. A model can theoretically predict well, but its features may vary significantly for each training. If that were the case, it would provide very different explanations for the similarity judgements (see section 6.8). In order to assess the variance, standard deviations of the global saliences across different trainings are provided as well.

6.7.1. Convergence of the model

Another important consideration that has to be made before assessing the model's predictions is whether there is a sufficient amount of data for the model to be trained. The whole data set is randomly reduced to 10%, 20%, 30%, … 100% of the total size of 1568 trials. The model, using all parts and roles, is then trained with each of these smaller datasets and its loss is plotted in a graph in relation to the number of epochs of the training. The graphs in figures 6.26 and 6.27 show how the model's loss decreases with each epoch. Each line in the graphs corresponds to one of the smaller data sets.

It is apparent that the model has a sufficient amount of data as the lines converge at around 300 epochs for the generalized predictive method (see figure 6.26) and around 200 epochs for the contextual method (see figure 6.27). After 200–300 epochs, the model's loss is comparable regardless of the size of the data subset used for the training. However, this is not to say that the model will generalize well after training with such a small amount of data.

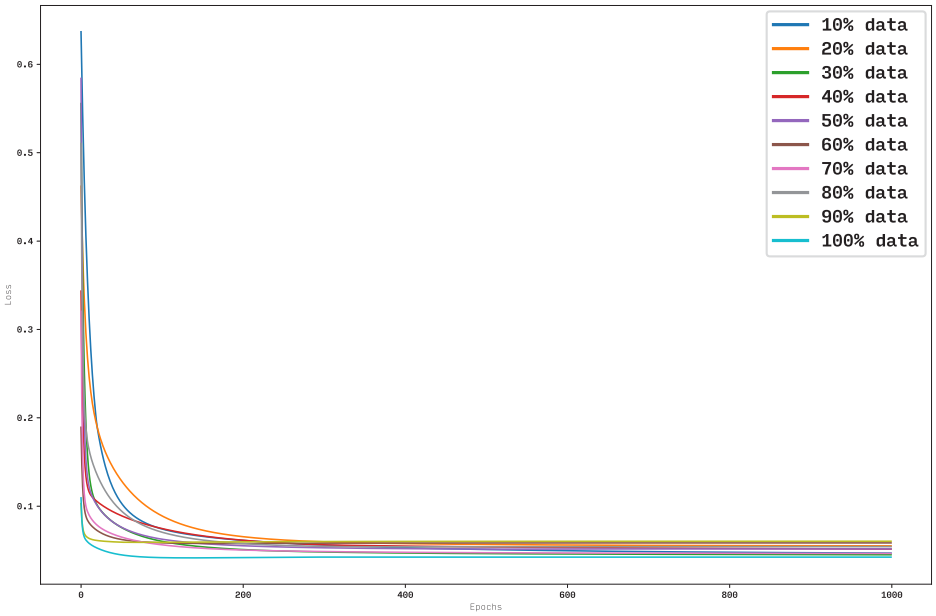


FIGURE 6.26: The progressive training improvement of the model with the generalized predictive method. Loss improvements of the model trained with different sizes of the data sets are marked in different colours. The value of the loss is represented on the y-axis. The number of training epochs is on the x-axis.

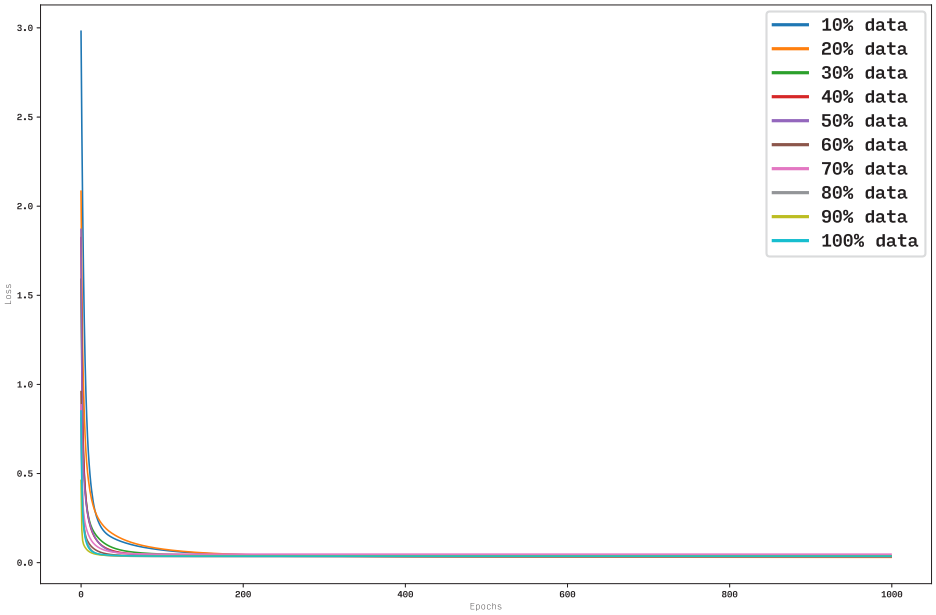


FIGURE 6.27: The progressive training improvement of the model with the contextual predictive method. Loss improvements of the model trained with different sizes of the data sets are marked in different colours. The value of the loss is represented on the y-axis. The number of training epochs is on the x-axis.

6.7.2. Cross-validation

The model was first trained and cross-validated with the data for all the scripts and typefaces using several different settings. Subsequently, the best performing setting was trained and cross-validated with the data for the individual scripts and for the groups of typefaces as they were specified in section 6.2. The model was trained 50 times for each setting. The data set was split randomly into training and evaluation groups every time. Table 6.5 shows the cross-validation results for training on all the data. The means reported are across all training sessions. The columns refer to:

- **Features:** whether parts or roles or both were included in the model; the total number of features (parameters of the model) is provided in parentheses.
- **Method:** predictive method used.
- **Loss:** the mean of the loss function (mean square error) applied on the cross-validation data; the lower the value, the better.
- **Accuracy:** the mean of the coefficient of determination; the higher the value, the better.
- **OOOO:** the mean and standard deviation for the percentage of trials with correctly predicted overall odd one out characters; the higher the value, the better.
- **Kappa:** the mean of Cohen's kappa; the higher the value, the better.
- **Saliences STD:** the mean of standard deviation of global saliences.

Features (no. of features)	Method	Loss	Accuracy	OOOO	Kappa	Saliences STD
Parts only (14)	Generalized	0.06	0.28	65.3 ±2.0%	0.47	0.20
	Contextual	0.05	0.35	64.2 ±2.1%	0.46	0.10
Roles only (24)	Generalized	0.10	−0.26	61.8 ±2.3%	0.42	0.17
	Contextual	0.05	0.39	65.5 ±1.9%	0.48	0.07
Parts & roles (38)	Generalized	0.04	0.44	72.6 ±2.0%	0.58	0.18
	Contextual	0.03	0.62	74.4 ±1.8%	0.61	0.06
Special methods	Majority	0.07	0.15	100.0 ±0.0%	1.00	–
	Oracle	0.00	1.00	100.0 ±0.0%	1.00	–
	Random	0.11	−0.43	32.8 ±1.6%	−0.01	–

TABLE 6.5: The cross-validation results for all the scripts studied together. The model was trained on 1176 trials and cross-validated on 392 trials. None of the special methods (Majority, Oracle, Random) uses the model, hence the column Saliences STD is empty for those. The legend for the columns is in the text of section 6.7.2.

To get a better idea about how good or bad the model could be, three special predictive methods were used as baselines. The *Random* method generates predictions randomly, the *Majority* method always predicts according to the majority opinion among participants, and the *Oracle* method predicts the response frequencies exactly as they were observed. It is not surprising that the Random method evaluates poorly and the Oracle method evaluates best, i.e. with the loss 0.0 and accuracy 1.0. The Majority method shows the diversity of participants' responses well. Even though the OOOO is on 100%, as the overall odd one out success rate is based on the majority of participants' opinion, the loss is high as this method completely disregards the less popular opinions. The relatively high loss and low accuracy show that the participants rarely had one clear answer to the study trials. Note that none of these special predictive methods uses any features from the model. The Random method is algorithmic, while the Majority and Oracle methods are based purely on the observations.

The results from testing with the data for all scripts together show that the contextual predictive method consistently outperforms the generalized predictive method. This suggests that a generalized view of pair similarity where the similarity of two characters is defined absolutely does not provide good insights into similarity judgements in context.

Moreover, comparison of the loss and accuracy for different sets of features shows that the roles-only setting outperforms the parts-only setting and both perform worse than the parts-and-roles setting. This is to be expected, given the increase in the number of model's parameters. Nonetheless, it makes a good case for participants' use of more general notions in their judgements. Also note that the generalized predictive method with the roles-only setting has atypically low accuracy (–0.26) outside the common bounds from 0.0 to 1.0.

Naturally, Cohen's kappa improves with increasing accuracy in all of the results. Its values for the contextual predictive method with all parts and roles included suggest a substantial agreement. This setting has also the lowest loss and variance of the saliences as well as the highest accuracy. Thus, it has been used for cross-validation using the data sets based on individual scripts and typeface groups defined in section 6.2 (see tables 6.6–6.8 for the results).

Data set (train/c-val.)	Features (no. of features)	Loss	Accuracy	OOOO	Kappa	Saliences STD
Sans (126/42)	Parts & roles (15)	0.01	0.82	90.9 ±4.9%	0.86	0.11
Serif (210/70)	Parts & roles (20)	0.03	0.59	74.0 ±4.6%	0.61	0.07
All (336/112)	Parts & roles (20)	0.03	0.64	78.4 ±3.6%	0.67	0.07

TABLE 6.6: The cross-validation results for different data sets for the Cyrillic script. Numbers of trials used for training and cross-validation are provided in parentheses in the Data set column. The legend for the rest of the columns is in the text of section 6.7.2.

Data set (train/c-val.)	Features (no. of features)	Loss	Accuracy	OOOO	Kappa	Saliences STD
Low-contrast (168/56)	Parts & roles (23)	0.03	0.53	71.4 ±5.8%	0.57	0.08
High-contrast (168/56)	Parts & roles (21)	0.03	0.48	66.9 ±4.9%	0.50	0.07
All (336/112)	Parts & roles (23)	0.03	0.50	67.9 ±3.9%	0.52	0.06

TABLE 6.7: The cross-validation results for different data sets for the Devanagari script. Numbers of trials used for training and cross-validation are provided in parentheses in the Data set column. The legend for the rest of the columns is in the text of section 6.7.2.

Data set (train/c-val.)	Features (no. of features)	Loss	Accuracy	OOOO	Kappa	Saliences STD
Sans (252/84)	Parts & roles (15)	0.03	0.71	81.1 ±4.3%	0.68	0.07
Serif (252/84)	Parts & roles (20)	0.02	0.75	82.8 ±4.2%	0.73	0.08
All (462/154)	Parts & roles (20)	0.03	0.70	80.8 ±2.9%	0.69	0.07

TABLE 6.8: The cross-validation results for different data sets for the Latin script. Numbers of trials used for training and cross-validation are provided in parentheses in the Data set column. The legend for the rest of the columns is in the text of section 6.7.2.

The results for Cyrillic and Latin have a relatively high accuracy and Cohen's kappa indicates substantial and almost perfect agreements. However, the accuracy of the model with the Devanagari data sets is substantially lower than with the other scripts despite the loss being comparable. This can probably be attributed to the visual complexity of the script or to insufficient character concepts.

Training the model on individual scripts yields more accurate predictions. This can be explained by different global saliences of the shared features in the individual scripts. For example, the role **r_closed** could have a different salience within Latin and within Devanagari depending on the participants' responses. The same effect applies when training with individual groups of typefaces. And it would also appear if the model was trained with individual typefaces, although this was not done here as the data for individual typefaces are limited and should not be used to draw general conclusions.

Tables 6.5–6.8 show standard deviations of features' global saliences across different trainings of the model. Moreover, figure 6.28 shows the box plots for the distribution of global saliences across all trainings with the complete data set for Devanagari. Although most saliences have a clear median with a small variance, some of the saliences vary substantially more. For example, the saliences for the **nukta** feature vary greatly. This is due to the fact that the **nukta** feature is present in only one character concept and does not influence many predictions. Thus, the SGD algorithm keeps changing it more freely. As the feature is present in only one character, it does not constitute a problem.

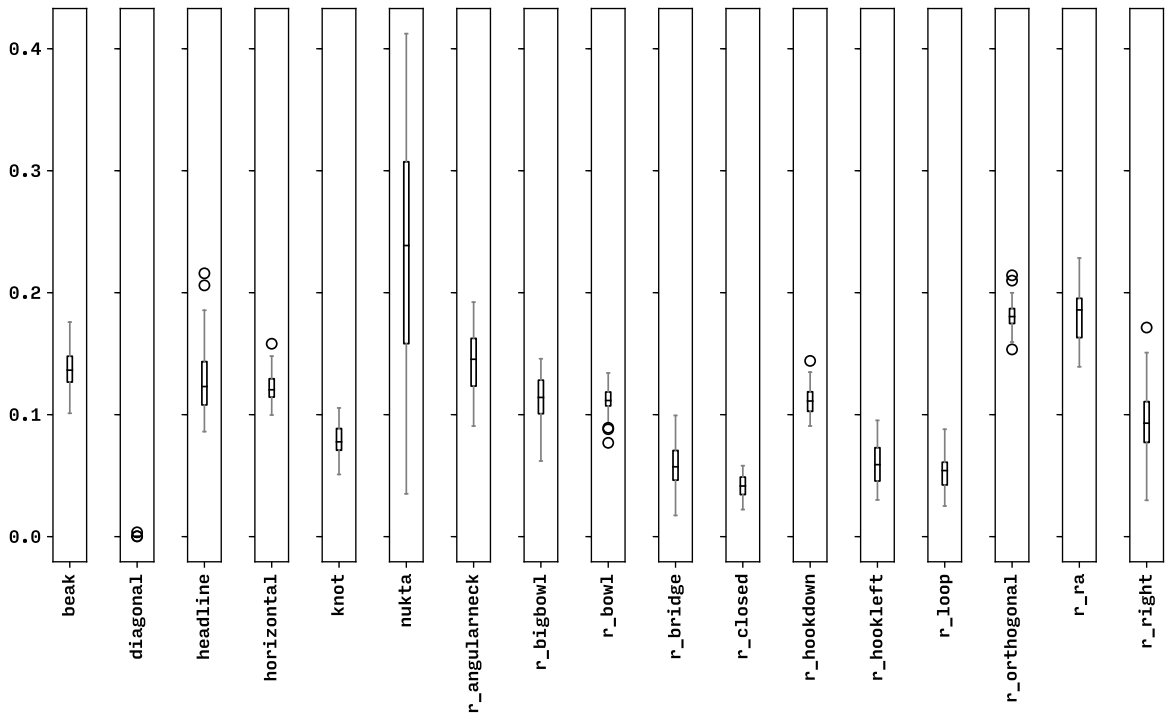


FIGURE 6.28: Box plots for the distribution of global saliences during the 50 training iterations using all data for Devanagari. Only some of the global saliences are shown.

6.8. Predictions and explanations

Having cross-validated the model, it was used to make predictions on the whole data set. The predictions were made for each typeface group separately. A new version of the model has been created for each of the groups. Each time, the model's global saliences were set to the means of the global saliences across all trainings for the specific typeface group. Then, the specific version of the model was used to make predictions for a particular typeface group. The results for the individual groups as well as for the whole data set are reported in table 6.9. The corresponding predicted response frequencies for all trials are included in appendix 2.

The loss is higher than in cross-validation as the data set is bigger and more likely to include substantially different predictions. The accuracy remained approximately the same or improved. Cohen's kappa statistic suggests a substantial and almost perfect agreement for Latin and Cyrillic and moderate agreement for Devanagari. Based on the percentage for the overall assessment, the model selects the odd one out correctly in roughly four out of five trials.

The improved accuracy and Cohen's kappa are to be expected, as the model was trained on the whole data.

Script	Typeface group	Loss	Accuracy	OOOO	Kappa
Cyrillic	Sans	0.04	0.83	92%	0.87
	Serif	0.09	0.61	75%	0.63
Devanagari	Low-contrast	0.08	0.57	73%	0.59
	High-contrast	0.07	0.52	68%	0.52
Latin	Sans	0.08	0.72	83%	0.71
	Serif	0.06	0.77	84%	0.75
All	All	0.07	0.69	79%	0.68

TABLE 6.9: Assessment of the predictions for all trials in the studies made by specific versions of the model fitted to each typeface group. The legend for the columns is in the text of section 6.7.2.

The model can be also used to make predictions for trials that did not appear in the studies. Some of them are presented in table 6.10 to provide an example. As there is no way to evaluate them, it is left to the reader to judge whether they correspond to their personal similarity judgements.

character		freq.
a	a	0.47
b	b	0.25
o	o	0.28

character		freq.
f	f	0.19
i	i	0.62
t	t	0.19

character		freq.
f	f	0.18
j	j	0.38
l	l	0.44

character		freq.
n	n	0.42
r	r	0.07
s	s	0.51

character		freq.
o	o	0.74
v	v	0.02
w	w	0.24

character		freq.
v	v	0.10
w	w	0.35
y	y	0.55

TABLE 6.10: Examples of predictions for trials that did not appear in the studies for the typeface Arial and the Latin script. The overall odd one out is marked with a grey background.

However, the feature-based approach was chosen in order to provide more than mere predictions. Supported by the fact that it has been trained on a large data set of readers' responses, it can be used to hypothesize about the features involved in the similarity judgements. The plots in figures 6.29 and 6.30 show the features and their saliences as they contributed to the modelled similarity judgements. It is important to note that the saliences in the plots are a multiplication of the local saliences of a concept and the global saliences of the model. Naturally, it only makes sense to make these plots for the pairs in triplets where the observations and predictions corresponded well.

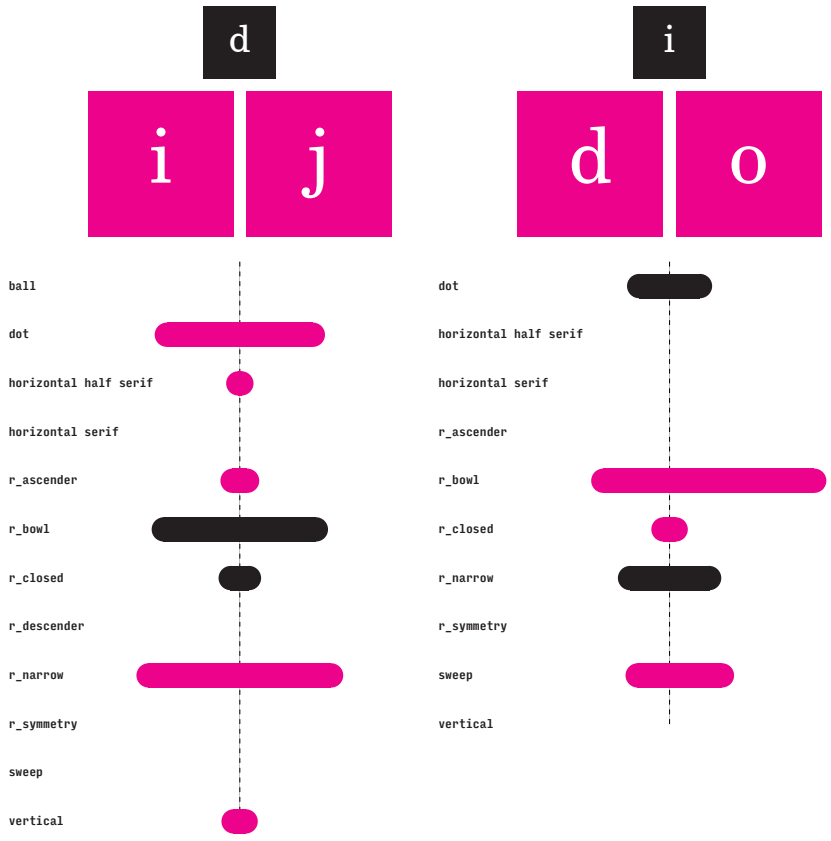


FIGURE 6.29: Plots showing feature saliences involved in the modelled similarity judgements of a pair of two main character concepts (marked in magenta) in the context of a third character concept (marked in black). The bars' lengths signify feature saliences. A bar is proportionally more to the left if the feature is more salient in the left main character. Analogically, for the right main character. The magenta bars are for features from the main character concepts. The black bars are for the features present only in the contextual character concept, i.e. they represent saliences of its distinct features. The features listed are limited to those present in any one of the character concepts.

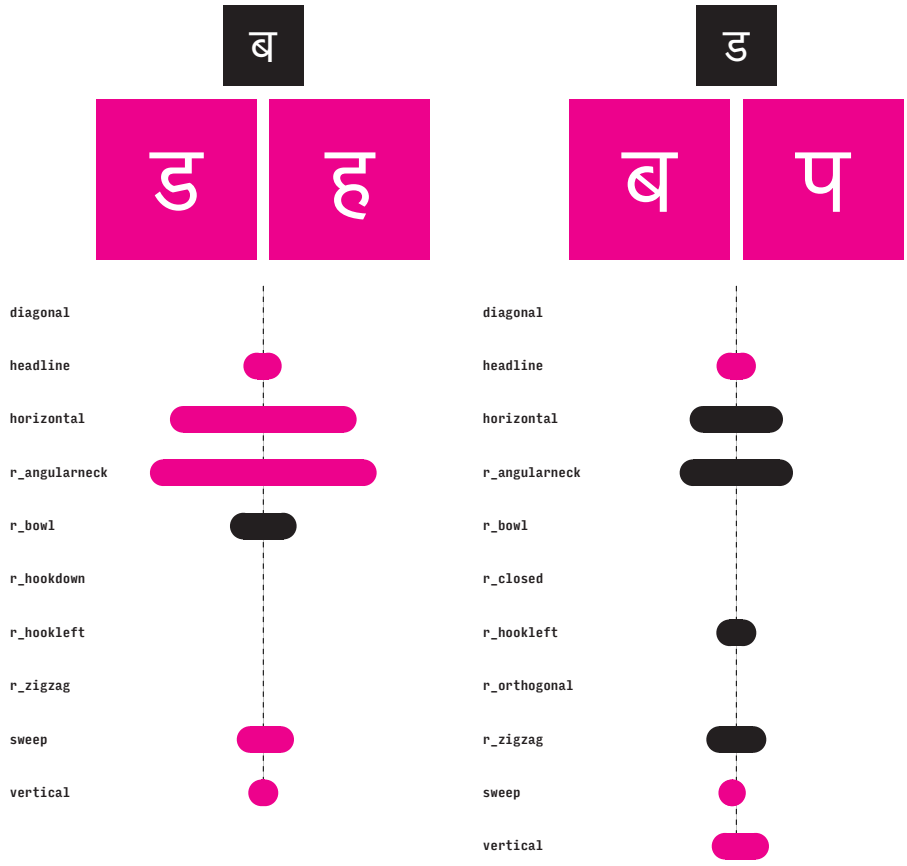


FIGURE 6.30: More plots showing feature saliences involved in the modelled similarity judgements of a pair of two main character concepts. For detailed description see figure 6.29.

And lastly, since predictions for any triplet trial can be generated, the model can compose a complete similarity matrix for a particular character set. The responses for all trials that include a particular character pair were aggregated in the same way as described in section 5.13. The resulting generalized similarity measures have been recorded in a symmetrical similarity matrix. The diagonal was left empty as responses involving pairs of two identical characters could not be reliably predicted. Appendix 10 contains such a matrix made for the typeface Arial and basic Latin alphabet.

This matrix can be compared with a subset of the matrix provided by Simpson et al. (2013), which used Arial for most of the character shapes shown to the participants; the character 'l' was presented in Times New Roman. Their original matrix also contains accented characters and an alternative single-storey version of an 'a', e.g. ɑ. These were removed from the version of their matrix used in this comparison; nonetheless the character 'l' in Times New Roman was kept to keep the character set complete (see also appendix 10). Spearman's rank correlation coefficient (Spearman, 1904) for the matrix

generated by the model and the matrix from Simpson et al. (2013) is 0.753, which suggests a high correlation. The p-value for an uncorrelated system producing data sets that have a Spearman correlation as high as this is close to 0.0, thus highly unlikely.

The above shows that the model can provide reasonably accurate predictions about participants' similarity judgements. It can also provide predictions for new trials or explain the predictions in terms of features and their relative saliences. The model not only agrees with the participants from the studies described in chapter 5, but it also correlates well with the similarity matrix that was constructed recently by researchers.

One can interpret the character concepts included in the model as hypotheses about the underlying principles of character similarity. Each feature corresponds to a visual analogy that can be made when judging similarity of characters. Therefore the assessment of the model is also an assessment of the ability of expert knowledge, which was used to produce these concepts, to explain the underlying principles of character similarity and coherence. As the accuracy rates are relatively high, one can address the first research question by stating that the character concepts included in this research are likely to map to readers' criteria when judging similarity. The likelihood varies depending on the script and the typeface group studied.

The second research question is addressed simply by showing that the model has been used with three different scripts. Although it currently predicts better for Latin and Cyrillic than for Devanagari, there is nothing in principle that would prevent it from being used with other scripts. To answer the second research question directly: the model, and thus the theory used to construct the character concepts, can be used to describe coherence in different scripts.

And lastly, the inaccuracy of the model's predictions can be taken as a measure of difference between the coherence perceived by readers and the coherence intended by designers, which addresses the third research quesion.

7. Conclusion

"The purpose of abstracting is not to be vague, but to create
a new semantic level in which one can be absolutely precise."
— Edsger W. Dijkstra (1972)

This research set out to study the principles underlying the coherence in the design of individual typefaces as perceived by readers. The approach has been cross-disciplinary. On the one hand, the study of perception required an empirical study with psychological considerations. On the other, the findings from the studies had to be compared with a formal theory explaining the underlying principles in terms of typeface design knowledge. This was done using a computational model.

Chapter 2: *Theory and definitions* discussed the different modes of the term character from the specific character shape, and the more general character concept, to the most general character category. These modes were used to define other terms such as a typeface, script, coherence, and style. This allowed for a discussion of character similarity at different levels of generality, and for a view of coherence as a structure of these similarity relationships.

Chapter 3: *Overview of the scripts studied* briefly described Latin, Cyrillic, and Devanagari as visual systems to contextualize the selection of representative characters for the studies with readers in chapter 5. Additionally, it also discussed the origin and nature of the conventional design principles of those scripts.

Chapter 4: *Approaches to character conceptualization* overviewed mental concepts and techniques used by contemporary typeface designers to produce character shapes and achieve coherence. Shapes can be interpreted in terms of contours or strokes. They can also be composed of other shapes or result from a transformation of other shapes. And lastly, more general qualities of shapes can be studied as they appear across multiple different characters.

While chapters 2–4 established the context, chapters 5 and 6 are the main focus of this thesis and contain its original contributions. Chapter 5: *Studies with readers* is the longest in this thesis. Providing sound evidence of readers' perceptions proved perhaps more challenging than formulating a theory to describe the processes involved in this perception.

There have been various empirical methods used to study character similarity, but none of them were deemed appropriate for this research. Methods involving suggestive decomposition of characters or using specific mark-up tools were disregarded right away. The characters could not be interpreted in any way before they were presented to the participants. Neither should the participants be instructed in what constitutes visual similarity.

They had to use their own personal approach. Methods involving challenging readers' perceptions to increase confusability of characters could not be used either as they would be insensitive to typeface design objectives. The appearance of the character shapes had to correspond to normal viewing conditions and could not be obscured to challenge the participants. Methods based on subjective ranking of character pair similarities could not be used either as participants tend to have personal approaches to how they use the ranking scale.

To avoid these issues and to keep the study trials as simple as possible, the participants were presented with a sequence of character triplets produced with a single typeface. They were asked to identify the odd one out in each of these triplets judging by their visual similarity. The task of eliminating the odd one out is also a statement about the complementary pair of characters. This method does not only avoid the disadvantages of the previous methods, it also solicits the similarity judgements in a non-directional way. Most importantly, it permits a study of character similarity with consideration of a specific context. When judging similarity of a character pair within a triplet, the remaining character becomes contextual. Realizing the impact the explicitly shown third character may have on the similarity judgement of the other two characters lays bare the limits of similarity measures that are concerned only with character pairs.

The studies were conducted online to collect a large number of responses from as diverse a group of participants as possible. The diversity of the participants was mapped using a questionnaire. Each of the scripts studied was represented with at least eight different typefaces intended for continuous reading. As it would not be possible to test all triplet combinations of the characters from each script, eight representative character shapes from each of the typefaces were sampled systematically to ensure the resulting triplet trials would contain diverse and challenging similarity judgements. These eight characters were combined in triplets exhaustively, thus there were 56 triplet trials in each study completed by each of the participants. This way, each pair of characters appeared in several different contexts.

The data collected from the studies were analysed to confirm that participants judged the visual similarity and to see what influence their background might have had on their responses. The aggregated responses for individual trials were taken as an indication of an overall opinion and they were combined into observed response frequencies that were used to assess the model in chapter 6. Additionally, similarity matrices for each typeface were produced showing that the context-sensitive method can also provide results in the more conventional format that is currently used by psychologists. In these matrices, the similarity measures are for character pairs with implicit context rather than for pairs within explicitly defined triplets (see appendix 3 for example of these matrices).

The research questions were addressed using computational modelling in chapter 6: *Modelling character similarity and coherence*. The hypothetical underlying principles based on expert knowledge were used to create character concepts to represent each of the characters studied. These concepts formed an essential part of the model that was used to predict participants' responses to trials from the studies in chapter 5. The assessment of the model's ability to predict the responses is also an assessment of the viability of the hypothetical underlying principles.

Various approaches used for theoretical modelling of similarity were considered, including deep learning with artificial neural networks, choice model, or template model. The modelling method used had to achieve good predictions, but it also needed to allow for the model to be trained on the data from the studies and for its parameters to be interpreted directly. It was required in order to directly link to the hypothetical underlying principles of similarity judgements. This is the main reason why neural networks were not used. Their parameters might be interpretable by a human. For example, one can visualize the layers of a convolutional neural network trained on images to reveal specific parts used to identify objects. However, the images of parts are not easily connected directly back to ready-made hypotheses. Having a human interpret every criterion participants might have used would be extremely tedious and potentially subjective. A feature-based approach has been used instead as it can provide insights into the underlying processes and can be trained using an algorithm.

The methods discussed in chapter 4 were used to conceptualize all characters studied using two kinds of features: more specific parts and more general roles. This was done in order to obtain general, but formally precise, character concepts that would allow straightforward comparison across different character categories. These character concepts were stored in the model together with an array of global feature saliences which define the relative importance of the features across the whole set of character concepts. The model included a predictive function that can take concepts and predict similarity judgements in order to provide response frequencies.

The global saliences were initially unknown and thus the model had to be trained in order to set them according to the readers' similarity judgements from the studies. The model was subsequently cross-validated and the results were used to assess the viability of the character concepts used.

The model achieved consistently high scores when predicting response frequencies for each trial as well as when identifying the overall odd one out. This in itself suggests that the hypotheses represented in the character concepts are likely to correspond to the criteria readers used when judging character similarity.

Besides producing predictions for all trials from the studies (see appendix 2), the model was also used to create plots visualizing features and their salience

in individual similarity judgements with specific context (see figures 6.29–6.30). These plots attempt to explain the underlying principles in terms of features and thus may prove useful to designers and psychologists alike.

And lastly, the model was used to produce a similarity matrix for the typeface Arial which was compared with a recent similarity matrix constructed for the same typeface based on the research by Simpson et al. (2013). Although some of the characters from the basic Latin alphabet did not occur in the studies with the typeface Arial in chapter 5, the model was capable of producing a complete similarity matrix that showed high correlation with the matrix by Simpson et al. This further supports the conclusion that the character concepts and the model's predictive function are close approximations of readers' perceptions of character similarity and coherence.

While the model achieved high scores for Latin and Cyrillic, it did not fare as well with Devanagari. This could be explained by the task of judging the visually more complex Devanagari characters being more difficult for the participants, or by the model currently using insufficient concepts for those characters. There is nothing to prevent the present theory from describing perceived coherence with respect to different scripts; however, the concepts for Devanagari characters included in the model might be up for questioning.

To address the last research question, whether coherence perceived by readers differs from the coherence intended by designers, one can look at the discrepancy between the model's predictions which are based on expert knowledge and the data from the studies with readers. While there is a difference and the model's predictions do not fit the readers responses perfectly, it can be said that there seems to be a close correlation between designers' intentions and readers' perceptions.

The original contributions of this research are described below.

7.1. The theory

The theory presented in chapter 6 uses character concepts that are more general than the digital descriptions of characters that typeface designers use in common contemporary fonts. At the same time, these concepts are less general than character categories that are difficult to define visually (see section 2.1). This level of generalization allows one to make precise comments about the typefaces' coherence and scripts' conventions. It also permits portability of the concepts between typefaces and potentially different fields, for, there is nothing that prevents the reuse of the concepts presented in chapter 6 (and appendix 7) in epigraphic or psychological studies, for example.

The method used to create the feature-based character concepts is independent of a particular script. It has been shown to work for the analysis of three different scripts – Cyrillic, Latin, and Devanagari – and it can, and hopefully will be, used to conceptualize other scripts and typefaces. Note that

the features that are not natural to a particular script need not be used. For example, using chirographic features, such as strokes, for a script which does not have a chirographic origin is unnecessary.

Most importantly, the model was shown to be able to predict readers' perceptions of similarity, which makes it a reliable source of hypotheses about the scripts' conventions and the underlying principles of coherence.

Moreover, the model provides a connection that can potentially be used to feed back from the data describing readers' similarity judgements to typeface design practice. The plots showing feature salience from the modelled similarity judgements (see figures 6.29–6.30) are one such example. While this knowledge may seem apparent to a designer familiar with a particular script, the plots may be a useful reference for a designer who is not familiar with the script. Other examples include various applications within the field of computer-aided design or quality assessment as discussed in the introduction (see section 1.3). Psychologists might use the feature-based model to interpret their results in a formal way or even use them to contextualize their findings in the field of character recognition.

7.2. Study format and method

Perhaps the most important contribution of chapter 5 is the method allowing study of the perception of similarity with specific context that is made explicit, with participants making relative, non-directional similarity judgements. Other studies derive or directly solicit character pair similarity judgements where the context is implied by using a particular script. The studies in this research presented triplets of characters where every character pair is accompanied by a third, contextual character. This provides more refined insights into how similarity works and allows the analysis of the influence of an explicit context in similarity judgements.

The format of the studies does not require the characters to be degraded in any way or for the readers to be challenged by other means. Therefore, it is sensitive to typeface design objectives. Non-designers, and psychologists and linguists in particular, are not always sensitive to the fact that the design of typefaces matters. They might be inclined to disregard the stylistic differences between typefaces as non-essential details. However, this is not always the case. The analysis of the data set showed that the results from the studies can be used to assess the effect of a particular typeface design on similarity judgements (see section 5.15.1) and it showed some profound differences.

Moreover, the results from this kind of study can be used to analyze the differences in similarity judgements depending on language fluency, nativity, or design skills (see sections 5.15.2 to 5.15.4). Potentially, it could be also used to study differences between character shapes produced with different typefaces or styles from the same type family.

7.3. Data set

In order to use computational modelling, a rather large amount of data was required. After all, the theory tested does not make basic statements about a few character categories. It makes statements about numerous visual analogies that can occur among these categories. The larger set of hypothetical statements naturally requires more substantial evidence to support it.

The resulting data set is, to my knowledge, one of the largest and most varied data sets among character similarity studies. The total of 1721 participants collectively responded to 1568 unique trials. With the total of 100,072 individual trial responses (see sections 5.9–5.11), the data set seems unprecedented in terms of scale. However, it is also exceptional in terms of the diversity of the participants and the scope of the tested material. Three different scripts were studied under the same conditions.

7.4. Discussion

The predictions of the model seem very promising. However, they are not perfect, and the question remains as to whether a better model could be proposed to achieve better predictions. For example, representing similarity as a transformation difference as mentioned in section 6.1 provides a promising way to improve the model.

Alternatively, one might approach the modelling as a straightforward classification task and identify the overall odd ones out as judged by the majority of participants without attempting to predict all three response frequencies for each triplet. This might permit use of other modelling approaches that could not be used in this research and might be more effective for this particular task.

Deep learning with artificial neural networks, although rejected for this particular research, is definitely a promising direction in studies of similarity. Current neural networks may lack explanatory power, but they provide impressively accurate results and once trained, they can be relatively fast. Thus, they might be effectively applied in computer-aided design systems.

It is worth noting that the model is more effective when making predictions for trials in one script only instead of making predictions for trials from studies with multiple different scripts. While this is supported by the fact that there is a smaller number of parameters to train, it may also suggest that identical features may have different salience in different scripts. This is an important consideration for anyone attempting to design a typeface which supports multiple different scripts while preserving some level of visual coherence across the whole set of characters.

7.5. Harmonious and essential future prospects

Further research could focus on studying featural overlaps between different scripts. For example, a feature that looks like a vertical half serif in Latin could have a different significance (and not just salience) in Cyrillic, because in Cyrillic it may represent different visual analogies among different characters. How should a designer decide which features should be treated consistently across different scripts when visual identity does not suffice? This kind of consideration is crucial in the domain of multi-script typography, where typographic harmonization of different scripts is an important topic. In fact, the very definition of harmonization could be based on coherence as it was formalized in this research. These are interesting topics that remain to be discussed and researched in more detail in the future.

A formalism that is closely based on readers' perceptions of similarity can also provide a means for quality assessment of typefaces. A set of criteria based on character similarity can be used to measure new or existing typeface designs in terms of their coherence. Naturally, this has to be done with a clear understanding of the role of coherence in a particular context. There may be no point in assessing coherence in typefaces that are meant to puzzle and attract the readers by their deliberate incoherence. However, typefaces used for setting continuous texts exhibit a certain set of recurring features, as evidenced in the character-concept diagrams in appendix 7. Such features, or rather the visual analogies they represent, might be expected in more conventional contexts.

The formalism can also guide implementation of parameters in systems such as Metafont (Knuth, 1982), Prototypo (Mathey & Babé, 2014), and others (see chapter 4).

Saying that there is a correlation between designers' intentions and readers' perceptions might seem like stating the obvious. Of course, designers design to affect readers. But the correlation suggests that typeface design should not be viewed narrowly as a stylistic exercise. It is a stylistic exercise built around the perception of connecting different characters and can thus inform our interpretation characters as shapes.

I hope that a research providing a formalism that can be used to study characters as interrelated entities will prove useful in the quest for the visual essence of characters – to truly understand what makes an 'a' an 'a'.

Bibliography

"Ethnologue: Languages of the world." *Ethnologue* (2018): accessed 6. 8. 2018, https://www.ethnologue.com.

ISO/IEC 8859-5: information technology – 8-bit single-byte coded graphic character sets – part 5: Latin/Cyrillic alphabet. International Organization for Standardization, 1999.

ISO 639-3: Codes for the representation of names of languages – part 3: alpha-3 code for comprehensive coverage of languages. International Organization for Standardization, 2007.

"Character, n." *OED online* (2018): accessed 8 August 2018, http://www.oed.com/view/Entry/30639.

The Unicode standard: version 7.0 – core specification. Mountain View, CA: The Unicode Consortium, 2014.

Adams, Debra. "abcdefg (a better constraint driven environment for font generation)." In *Raster Imaging and Digital Typography*, edited by Jacques André and Roger D. Hersch, 54–70. Cambridge University Press, 1989.

Adams, Debra and Richard Southall. "Problems of font quality assessment." In *Raster Imaging and Digital Typography*, edited by Jacques André and Roger D. Hersch, 213–22. Cambridge University Press, 1989.

Adobe. *Adobe Type 1 font format.* Addison-Wesley, 1993.

Adobe. *Designing multiple master typefaces.* Adobe Systems Incorporated, 1995.

Adobe. "Adobe Illustrator CC: vector graphic design software." (2018): accessed 6 August 2018. https://www.adobe.com/products/illustrator.html.

Agüera y Arcas, Blaise. "Temporary matrices and elemental punches in gutenberg's DK type." In *Incunabula and their Readers: Printing, Selling and Using Books in the Fifteenth Century*, edited by Kristian Jensen, 1–12. London: The British Library, 2003.

Ahrens, Tim and Shoko Mugikura. *Size-Specific Adjustments to Type Designs : An investigation of the Principles Guiding the Design of Optical Sizes.* Just Another Foundry, 2014.

Alexeeva, Svetlana and Alena Konina. "Crowded and uncrowded perception of lowercase Cyrillic letters in parafoveal vision: confusion matrices based on error rates." *The European Conference on Visual Perception* (2016).

Amazon Mechanical Turk. "Amazon Mechanical Turk." (2005): accessed 6 August 2018. http://www.mturk.com.

Apple. "Truetype™ reference manual." (n. d.): accessed 6 August 2018. https://developer.apple.com/fonts/TrueType-Reference-Manual/.

Baines, Phil and Andre Haslam. *Type & typography.* London: Laurence King, 2005.

Barnard, G. A. "A new test for 2×2 tables." *Nature* 156 (1945): 177.

Baudin, Fernand. *How typography works: (and why it is important).* London: Lund Humphries Publishers, 1989.

Bauermeister, Benjamin. *A manual of comparative typography: the PANOSE system.* New York: Van Nostrand Reinhold, 1988.

Ben-Kiki, Oren, Clark Evans, and Brian Ingerson. "YAML ain't markup language." (2004): accessed 7. 6. 2018. http://yaml.org.

Berlaen, Frederik. "Robofont." (2018): accessed 7. 6. 2018. http://robofont.com.

Biederman, Irving. "Recognition-by-components: a theory of human image understanding." *Psychological Review* 94, no. 2 (1987): 115–47.

Bigelow, Charles. "Principles of type design for the personal workstation." ATypI Congress in Kiel (1985).

Blesser, B., R. Shillman, C. Cox, T. Kuklinski, J. Ventura, and M. Eden. "Character recognition based on phenomenological attributes." *Visible Language* 7 (1973): 209–23.

Blokland, Erik van and Just van Rossum. *Bad type.* St Bride Library, 2004.

Boles, David B. and John E. Clifford. "An upper- and lowercase alphabetic similarity matrix, with derived generation similarity values." *Behavior Research Methods, Instruments, & Computers* 21 (1989): 579–86.

Bottou, Léon, Frank E. Curtis, and Jorge Nocedal. "Optimization methods for large-scale machine learning." *arXiv* (2016).

Bouma, H. "Visual recognition of isolated lower case letters." *Vision Research* 11 (1971): 459–74.

Carter, Harry Graham. *A view of early typography: up to about 1600.* London: Hyphen Press, 2002.

Carter, Harry Graham. "Fournier on typefounding." In *The Manuel typographique of Pierre-Simon Fournier le jeune*, edited by James Mosley, Darmstadt: 1995.

Catich, Edward. *The origin of the serif: brush writing & Roman letters.* Davenport IA: Catfish Press, 1968.

Clayton, Ewan. *The golden thread: the story of writing.* Counterpoint Berkeley, 2013.

Clopper, Charles J. and Egon S. Pearson. "The use of confidence or fiducial limits illustrated in the case of the binomial." *Biometrika* 26, no. 4 (1934): 404–13.

Cohen, Jacob. "A coefficient of agreement for nominal scales." *Educational and Psychological Measurement* 20, no. 1 (1960): 37–46.

Corel. "CorelDRAW Graphics Suite 2018: Graphic Design Software." (2018): accessed 6 August 2018. https://www.coreldraw.com/en/product/graphic-design-software/.

Coulmas, Florian. *The writing systems of the world* Oxford: Blackwell, 1989.

Coulmas, Florian. *The Blackwell encyclopedia of writing systems.* Oxford: Blackwell Publishers, 1996.

Crump, Matthew J. C., John V. McDonnell, and Todd M. Gureckis. "Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research." *PLOS ONE* 8 (2013):

Dahlström, Erik, Patrick Dengler, Anthony Grasso, Chris Lilley, Cameron McCormack, Doug Schepers, Jonathan Watt, Jon Ferraiolo, Jun Fujisawa, and Dean Jackson. "Scalable vector graphics (SVG)" *W3C* (2011): accessed 6 August 2018. https://www.w3.org/Graphics/SVG/.

Dalvi, Girish. "Anatomy of Devanagari Typefaces." *Design Thoughts* (2009).

Decock, L. and I. Douven. "Similarity after Goodman." *Review of Philosophy and Psychology* 2, no. 1 (2011): 61–75.

Dijkstra, Edsger W. "The humble programmer." *Communications of the ACM* 15, no. 10 (1972): 859–66.

Dunn, Olive Jean. "Multiple comparisons among means." *Journal of the American Statistical Association* 56 (1961): 52–64.

Dwiggins, William Addison. *WAD to RR: a letter about designing type.* Cambridge MA: Harvard College Library, Department of Printing and Graphic Arts, 1940.

Dyson, Mary C., Keith Tam, Clare Leake, and Brian Kwok. "How does expertise contribute to the recognition of Latin and Chinese characters?" In *Digital fonts and reading*, edited by Mary C. Dyson and Ching Y. Suen, World Scientific Publishing, 2015.

D'Agostino, Ralph and Egon S. Pearson. "Tests for departure from normality." *Biometrika* 60, no. 3 (1973): 613–22.

Eco, Umberto. *Otevřené dílo [The open work].* Argo, 2015.

Encyclopædia Britannica. "Leonardo Pisano." *Encyclopædia Britannica online* (2016): accessed 6 August 2018. http://www.britannica.com/biography/Leonardo-Pisano.

Eusebius, [Caesariensis], [Trapezuntius] Georgius, and [Papa V] Nicolaus. *Praeparatio evangelica.* Venice: 1470.

Fiset, Daniel, Caroline Blais, Catherine Éthier-Majcher, Martin Arguin, Daniel Bub, and Frédéric Gosselin. "Features for identification of uppercase and lowercase letters." *Psychological Science* 19 (2008): 1161–68.

Fiset, Daniel, Caroline Blais, Martin Arguin, Karine Tadros, Catherine Éthier-Majcher, Daniel Bub, and Frédéric Gosselin. "The spatio-temporal dynamics of visual letter recognition." *Cognitive Neuropsychology* 26 (2009): 23–35.

Fisher, Ronald A. "On the interpretation of $\chi 2$ from contingency tables, and the calculation of P." *Journal of the Royal Statistical Society* 85, no. 1 (1922): 87–94.

FontLab. “Fontlab VI.” (2017): accessed 6 August 2018. https://www.fontlab.com/font-editor/fontlab-vi/.

Fournier, Pierre-Simon. *Manuel typographique, vol. 1.* Paris: 1706.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning.* Springer, 2001.

Furious Collective. “Formkeep.” (2014): accessed 6 August 2018. http://www.formkeep.com.

Garces, Elena, Aseem Agarwala, Diego Gutierrez, and Aaron Hertzmann. “A similarity measure for illustration style.” *ACM Transactions on Graphics (TOG)* 33, no. 4 (2014): 93.

Gaskell, Philip. “A nomenclature for the letterforms of Roman type.” *Visible Language* 10, no. 1 (1976): 41–51.

Gauthier, I., A. C. Wong, W. G. Hayward, and O. S. Cheung. “Font tuning associated with expertise in letter perception.” *Perception* 35, no. 4 (2006): 541–59.

Laurent Gautier and rpy2 contributors. “RPy2.” (2008): accessed 6 August 2018. https://rpy2.readthedocs.io.

Getform. “Getform.” (2015): accessed 6 August 2018. http://www.getform.org.

Geyer, L. H. “Recognition and confusion of the lowercase alphabet.” *Perception & Psychophysics* 22 (1977): 487–90.

Gombrich, Ernst. “Style.” In *The art of art history: a critical anthology*, New York: Oxford University Press, 2009.

Gordon, Yuri. *Книга про буквы от Аа до Яя [Book of letters from Aa to Яя].* Moscow: Publishing house of the Artemy Lebedev Studio, 2013.

Hahn, Ulrike, Nick Chater, and Lucy B. Richardson. “Similarity as transformation.” *Cognition* 87, no. 1 (2003): 1–32.

Haralambous, Yannis. *Fonts & encodings.* Sebastopol CA: O’Reilly Media, 2007.

Harvey, Michael. *Creative lettering today.* New York: Design Books, 1996.

Hassan, Tamir, Changyuan Hu, and Roger D. Hersch. “Next generation typeface representations: revisiting parametric fonts.” *10th ACM symposium* (2010): 181–84.

Heap. “Heap analytics.” (2013): accessed 6 August 2018. https://heapanalytics.com.

Hofstadter, Douglas Richard. “Meta-font, metamathematics, and metaphysics: comments on Donald Knuth’s ‘The concept of a meta-font’.” *Visible Language* 16, no. 4 (1982): 309–38.

Hofstadter, Douglas Richard and Gary McGraw. *Letter Spirit: an emergent model of the perception and creation of alphabetic style.* CRCC Technical Report 90, Indiana University: 1993.

Hofstadter, Douglas Richard and Gary McGraw. “Letter Spirit: esthetic perception and creative play in the rich microcosm of the Roman alphabet.” In *Fluid concepts & creative analogies: computer models of the fundamental mechanisms of thought*, 407–66. Penguin Books, 1998.

Holbrook, Morris B. “A comparison of methods for measuring the interletter similarity between capital letters.” *Perception & Psychophysics* 17, no. 6 (1975): 532–36.

Hu, Catherine and Thomas Fischer. “Cultural perception and technique as constraints in computational design evolution: a case study of generative Chinese typography.” *Proc. 8th WSEAS International Conference on Computers.* (2004).

Hu, Changyuan. *Synthesis of parametrisable fonts by shape components.* Dissertation, École Polytechnique Fédérale De Lausanne, 1998.

Hu, Changyuan and Roger D. Hersch. “Parametrizable fonts based on shape components.” *IEEE Computer Graphics and Applications* 21 (2001): 70–85.

Huffman, Steve and Alexis Ohanian. “Reddit.” (2005): accessed 6 August 2018. https://reddit.com.

Hunter, J. D. “Matplotlib: a 2d graphics environment.” *Computing In Science & Engineering* 9, no. 3 (2007): 90–95.

Inkscape. “Inkscape: draw freely.” (2018): accessed 6 August 2018. https://inkscape.org.

Johnston, Edward. *Writing & illuminating, & lettering.* London: Sir Isaac Pitman & Sons, 1958.

Jones, Eric, Travis E. Oliphant, Pearu Peterson et al. "Scipy: open source scientific tools for Python." (2001): accessed 6 August 2018, http://www.scipy.org.

Kamermans, Mike. "A primer on bézier curves: a free, online book for when you really need to know how to do bézier things." (2011): accessed 6 August 2018. https://pomax.github.io/bezierinfo/.

Karow, Peter. *Font technology: description [methods] and tools.* Berlin: Springer-Verlag, 1994.

Kelly, George. *The psychology of personal constructs.* New York: Norton, 1955.

Keren, Gideon and Stan Baggen. "Recognition models of alphanumeric characters." *Perception & Psychophysics* 29 (1981): 234–46.

Kinoshita, Sachiko and Leanne Kaplan. "Priming of abstract letter identities in the letter match task." *The Quarterly Journal of Experimental Psychology* 61, no. 12 (2008): 1873–85.

Kinross, Robin. "What is a typeface?" In *Unjustified texts: perspectives on typography*, London: Hyphen Press, 2002.

Kmenta, Jan. *Elements of econometrics.* Macmillan, 1986.

Knuth, Donald Ervin. "The concept of a meta-font." *Visible Language* 16, no. 1 (1982): 3–27.

Knuth, Donald Ervin. "Lessons learned from Metafont." *Visible Language* 19, no. 1 (1985): 35–53.

Knuth, Donald Ervin. *Digital typography.* Stanford CA: CSLI Publications, 1999.

Korger, Hildegard. *Handbook of type and lettering.* Design Press, 1992.

Kruskal, Joseph B. "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis." *Psychometrika* 29, no. 1 (1964): 1–27.

Kruusimagi, Martin, Matthew Terrell, and Daniel Ratzinger. "Call for participants." (2018): accessed 6 August 2018. https://callforparticipants.com.

Lambert, Hester Marjorie. *Introduction to the Devanagari script for students of Sanskrit, Hindi, Marathi, Gujarati and Bengali, etc.* London: Oxford University Press, 1953.

Legros, Lucien Alphonse and John Cameron Grant. *Typographical printing--surfaces: the technology and mechanism of their production.* London: Longmans Green and Co., 1916.

Leming, Tal, Just van Rossum, and Erik van Blokland. "Unified font object (UFO)." (2004): accessed 7. 6. 2018. http://unifiedfontobject.org.

Liu, Lei and Aries Arditi. "How crowding affects letter confusion." *Optometry & Vision Science* 78 (2001): 50–55.

Luce, R Duncan. "Detection and recognition." In *Handbook of mathematical psychology*, 103–89. 1963.

Luce, R. Duncan. "The choice axiom after twenty years." *Journal of Mathematical Psychology* 15, no. 3 (1977): 215–33.

Lun, Zhaoliang, Evangelos Kalogerakis, and Alla Sheffer. "Elements of style: learning perceptual shape style similarity." *ACM Transactions on Graphics* 34 (2015): 84:1–14.

Majoor, Martin. "My type design philosophy." *Typotheque articles on graphic design and typography* (2004): accessed 6 August 2018. https://www.typotheque.com/articles/my_type_design_philosophy.

Marr, David and Herbert Keith Nishihara. "Representation and recognition of the spatial organization of three-dimensional shapes." *Proceedings of the Royal Society* 200, no. 1140 (1978): 269–94.

Mathey, Yannick and Louis-Rémi Babé. "Prototypo: streamlining font creation." (2014): accessed 20 January 2018. http://www.prototypo.io.

Mayzner, Mark S. and Margaret Elizabeth Tresselt. "Tables of single-letter and digram frequency counts for various word-length and letter-position combinations." *Psychonomic Monograph Supplements* 1, no. 2 (1965): 13–32.

McFee, Brian and Gert Lanckriet. "Learning multi-modal similarity." *Journal of Machine Learning Research* 12 (2011): 491–523.

McGraw, Gary and Douglas Richard Hofstadter. "Perception and creation of alphabetic style." *Artificial Intelligence and Creativity: Papers from the 1993 Spring Symposium* (1993).

McGraw, Gary, John Rehling, and Robert L. Goldstone. *Roles in letter perception: human data and computer models.* CRCC Technical Report 68, Indiana University: 1994.

McGraw, Gary. *Letter Spirit (part one): emergent high-level perception of letters using fluid concepts.* Dissertation, Indiana University, 1995.

McKinney, Wes. "Data structures for statistical computing in Python." *Proceedings of the 9th Python in Science Conference 445* (2010): 51–56.

McLean, Ruari. *The Thames and Hudson manual of typography.* London: Thames and Hudson, 1992.

McQueen III, Clyde D., and Raymond G. Beausoleil. "Infinifont: a parametric font generation system." *Electronic Publishing* 6, no. 3 (1993): 117–32.

Medin, Douglas L., Robert L. Goldstone, and Dedre Gentner. "Similarity involving attributes and relations: judgments of similarity and difference are not inverses." *Psychological Science* 1 (1990): 64–69.

Meek, Rob. "Fontstruct." (2010): accessed 6 August 2018. https://fontstruct.com.

Menhart, Oldřich. *Tvorba typografického písma [Creating typefaces].* SPN, 1957.

Microsoft. "Opentype® specification version 1.8.2." *OpenType* (2018): accessed 29. July 2018, https://docs.microsoft.com/en-us/typography/opentype/spec/.

Mohanty, SK. "The Formulation of Parameters for Type Design of Indian Scripts Based on Calligraphic Studies." *Electronic Publishing, Artistic Imaging, and Digital Typography* (1998): 157–66.

Moxon, Joseph. *Mechanick exercises: or, the doctrine of handy-works.* D. Midwinter and T. Leigh, 1703.

Mueller, Shane T. and Christoph T. Weidemann. "Alphabetic letter identification: effects of perceivability, similarity, and bias." *Acta Psychologica* 139 (2012): 19–37.

Naik, Bapurao. *Typography of Devanagari.* Bombay: Directorate of Languages Govt. of Maharashtra, 1971.

Noordzij, Gerrit. "A program for teaching letterforms." In *Dossier A–Z 73*, edited by Fernand Baudin and John Dreyfus, Copenhagen: Association Typographique Internationale, 1973.

Noordzij, Gerrit. "The shape of the stroke." In *Raster Imaging and Digital Typography II*, edited by Jacques André and Robert A. Morris, 34–37. Cambridge University Press, 1991.

Noordzij, Gerrit. *The stroke: theory of writing.* London: Hyphen Press, 2005.

Norvig, Peter. "English letter frequency counts: Mayzner revisited or ETAOIN SRHLDCU." (2013): accessed 20. 1. 2018. http://www.norvig.com/mayzner.html.

Palmer, Stephen E. "Hierarchical structure in perceptual representation." *Cognitive Psychology* 9 (1977): 441–74.

Palmer, Stephen E. "Structural aspects of visual similarity." *Memory & Cognition* 6 (1978): 91–97.

Palmer, Stephen E. *Vision Science: photons to phenomenology* Cambridge MA: MIT Press, 1999.

Pandey, Anshuman. "An Overview of indic fonts for TEX." *TUGboat* 19, no. 2 (1998): 115–20.

Pearson, Karl. "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling." *Philosophical Magazine* 50, no. 302 (1900): 157–75.

Pecina, Martin and David Březina. "Type anatomy 1.0." (2008): accessed 23 July 2018.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. and Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. and Weiss, V. Dubourg, J. Vanderplas, A. and Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. "Scikit-Learn: Machine Learning in Python." *Journal of machine learning research* 12 (2011): 2825–30.

Phillips, John R., Kenneth O. Johnson, and Helen M. Browne. "A comparison of visual and two modes of tactual letter resolution." *Perception & Psychophysics* 34 (1983): 243–49.

Priolkar, Anant Kakba. *The Printing Press in India : Its Beginnings and Early Development : Being a Quartercentenary Commemoration Study of the Advent of Printing in Indian (in 1556).* Bombay: Marathi Samshodhana Mandala, 1958.

Python Software Foundation. "Python 3.7.0 documentation." (2018): accessed 6 August 2018, https://www.python.org.

Rehling, John. *Letter Spirit (part two): modelling creativity in a visual domain.* Dissertation, Indiana University, 2001.

Reich, Lewis N. and Harold E. Bedell. "Relative legibility and confusions of letter acuity targets in the peripheral and central retina." *Optometry and Vision Science* 77 (2000): 270–75.

Riesenhuber, Maximilian and Tomaso Poggio. "Hierarchical models of object recognition in cortex." *Nature Neuroscience* 2, no. 11 (1999): 1019.

Rock, Irvin. *Orientation and Form.* New York: Academic Press, 1973.

Rosa, Eva, Manuel Perea, and Peter Enneson. "The role of letter features in visual-word recognition: evidence from a delayed segment technique." *Acta Psychologica* 169 (2016): 133–42.

Rossum, Just van, Erik van Blokland, and Frederik Berlaen. "Drawbot." (2007): accessed 6 August 2018. http://www.drawbot.com.

Roweis, Sam T. and Lawrence K. Saul. "Nonlinear dimensionality reduction by locally linear embedding." *Science* 290, no. 5500 (2000): 2323–26.

R Core team. "R: A language and environment for statistical computing." (2013): accessed 6 August 2018. http://www.r-project.org.

Sanocki, Thomas. "Visual knowledge underlying letter perception: font-specific, schematic tuning." *Journal of Experimental Psychology: Human Perception and Performance* 13 (1987): 267–78.

Sanocki, Thomas. "Font regularity constraints on the process of letter recognition." *Journal of Experimental Psychology: Human Perception and Performance* 14, no. 3 (1988): 472–80.

Sanocki, Thomas. “Effects of font- and letter-specific experience on the perceptual processing of letters.” *The American Journal of Psychology* 105, no. 3 (1992): 435–58.

Sanocki, Thomas and Mary C. Dyson. “Letter processing and font information during reading: beyond distinctiveness, where vision meets design.” *Attention, Perception, & Psychophysics* 74, no. 1 (2012): 132–45.

Saussure, Ferdinand de. *Course in general linguistics*, edited by Charles Bally and Albert Sechehaye. London: Duckworth, 1972.

Schapiro, Meyer. “Style.” In *Theory and philosophy of art: style, artist, and society: selected papers.* New York: George Braziller, 1994.

Schneider, Uwe. “Datype: a stroke-based typeface design system.” *Computers & Graphics* 22, no. 4 (1998): 515–26.

Schneider, Uwe. “A hybrid approach for stroke-based letterform composition including outline-based methods.” *Computer Graphics Forum* 19 (2000): 243–56.

Schultz, Matthew and Thorsten Joachims. “Learning a distance metric from relative comparisons.” *Advances in Neural Information Processing Systems* (2004): 41–48.

Seifert, Georg and Rainer Erich Scheichelbauer. “Glyphs: create – produce – release, font editing for everyone.” (2016): accessed 7. 6. 2018. https://glyphsapp.com.

Shepard, Roger N. “The analysis of proximities: multidimensional scaling with an unknown distance function. I.” *Psychometrika* 27, no. 2 (1962): 125–40.

Shepard, Roger N. “The analysis of proximities: multidimensional scaling with an unknown distance function. II.” *Psychometrika* 27, no. 3 (1962): 219–46.

Sherif, Ameer M and Hossam A. H. Fahmy. “Meta-designing parameterized Arabic fonts for alqalam.” *TUGBoat* 29 (2008): 435–43.

Simpson, Ian C., Petroula Mousikou, Juan Manuel Montoya, and Sylvia Defior. “A letter visual-similarity matrix for Latin-based alphabets.” *Behavior Research Methods* 45 (2013): 431–39.

Smeijers, Fred. *Counterpunch: making type in the sixteenth century, designing typefaces now.* London: Hyphen Press, 1996.

Sokal, Robert R. and F. James Rohlf. *Biometry: the principles and practices of statistics in biological research.* New York: W. H. Freeman, 1994.

Solpera, Jan. *Klasifikace typografických písem latinkových [Classification of typefaces of latin origin].* Revolver Revue and VŠUP, 2009.

Southall, Richard. "Shape and appearance in typeface design." *Protext* III (1986): 75–86.

Southall, Richard. "Character description techniques in type manufacture." In *Raster Imaging and Digital Typography II*, edited by Jacques André and Robert A. Morris, 16–27. Cambridge University Press, 1991.

Southall, Richard. *Printer's type in the twentieth century: manufacturing and design methods.* London: Oak Knoll Press, 2005.

Spearman, Charles. "The proof and measurement of association between two things." *The American journal of psychology* 15, no. 1 (1904): 72–101.

Thorburn, William M. "The myth of Occam's razor." *Mind* 27, no. 107 (1918): 345–53.

Torgerson, Warren S. "Multidimensional scaling: I. theory and method." *Psychometrika* 17, no. 4 (1952): 401–19.

Townsend, J. T. "Theoretical analysis of an alphabetic confusion matrix." *Perception & Psychophysics* 9, no. 1 (1971): 40–50.

Tracy, Walter. *Letters of credit: a view of type design.* Boston: D. R. Godine, 2003.

Tversky, Amos. "Features of similarity." *Psychological Review* 84 (1977): 327–52.

Twitter. "Twitter." (n. d.): accessed 6 August 2018. https://twitter.com.

Ullman, Stephen. *Style in the French novel.* Cambridge: 1957.

Viera, Anthony J. and Joanne M. Garrett. *Understanding interobserver agreement: the kappa statistic.* Family Medicine 37, no. 5 (2005): 360–63.

Walker, Peter. “Font tuning: a review and new experimental evidence.” *Visual Cognition* 16, no. 8 (2008): 1022–58.

Walker, Sue and Linda Reynolds. “Serifs, sans serifs and infant characters in children’s reading books.” *Information Design Journal* 11, no. 2 (2002): 106–22.

Watson, Ian. *Cognitive design: creating the sets of categories and labels that structure our shared experience.* Dissertation, The State University of New Jersey, 2005.

Wiley, Robert W., Colin Wilson, and Brenda Rapp. “The effects of alphabet and expertise on letter perception.” *Journal of Experimental Psychology: Human Perception and Performance* (2016). Advance online publication.

Wiser, Marianne. “The Role of Intrinsic Axes in Shape Recognition.” *Proceedings of the Third Annual Conference of the Cognitive Science Society* (1981): 184–86.

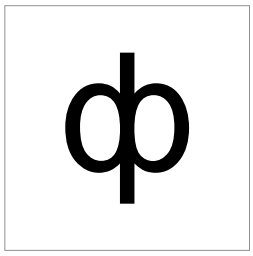
Yefimov, Vladimir. “Civil type and Kis Cyrillic: a modern Russian typeface based on historical forms.” In *Language culture type: international type design in the age of unicode*, edited by John D Berry, ATypI, Graphis, 2002.

Zhang, Zhenyue and Jing Wang. “MLLE: modified locally linear embedding using multiple weights.” *Advances in Neural Information Processing Systems* (2007): 1593–600.

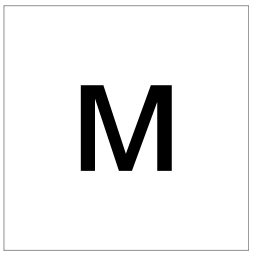
Zhukov, Maxim. “The peculiarities of Cyrillic letterforms: design variation and correlation in Russian typefaces.” *Typography papers* 1 (1996): 5–26.

Appendix 1: character shapes studied

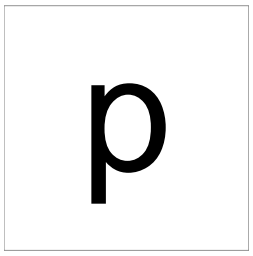
Characters studied as they were presented to the participants of the studies described in chapter 5.


ф

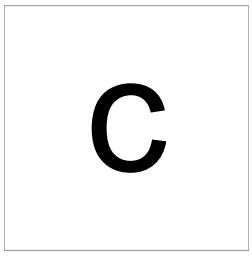
Typeface: Arial
Cyrillic [ef]


м

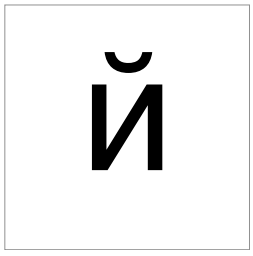
Typeface: Arial
Cyrillic [em]


р

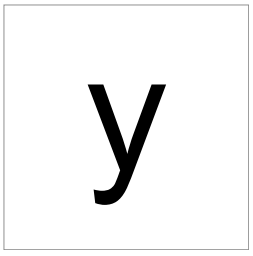
Typeface: Arial
Cyrillic [er]


с

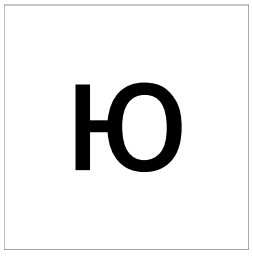
Typeface: Arial
Cyrillic [es]


й

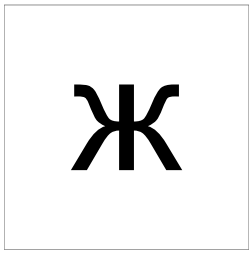
Typeface: Arial
Cyrillic [short i]


у

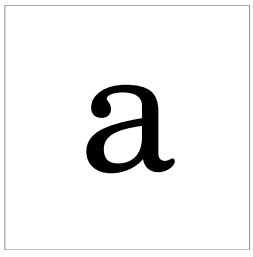
Typeface: Arial
Cyrillic [u]


ю

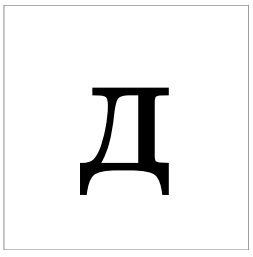
Typeface: Arial
Cyrillic [yu]


ж

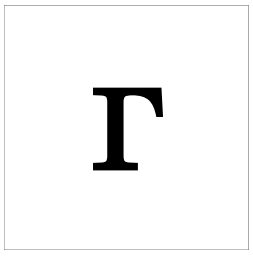
Typeface: Arial
Cyrillic [zhe]


а

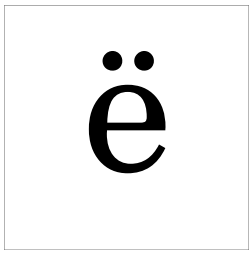
Typeface: Century Schoolbook
Cyrillic [a]


д

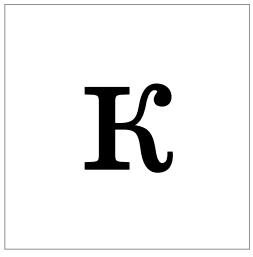
Typeface: Century Schoolbook
Cyrillic [de]


г

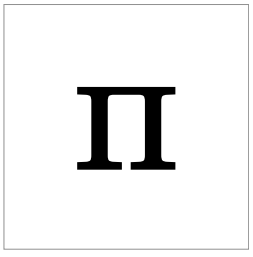
Typeface: Century Schoolbook
Cyrillic [ghe]


ё

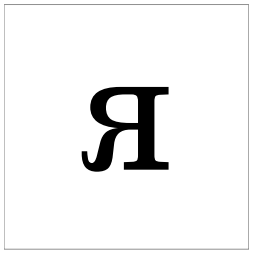
Typeface: Century Schoolbook
Cyrillic [io]


к

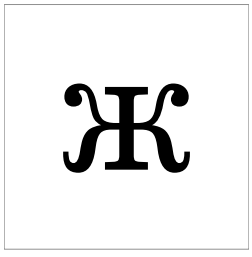
Typeface: Century Schoolbook
Cyrillic [ka]


п

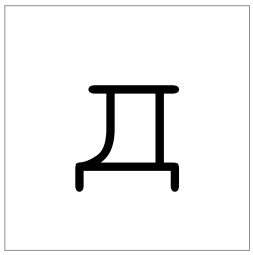
Typeface: Century Schoolbook
Cyrillic [pe]


я

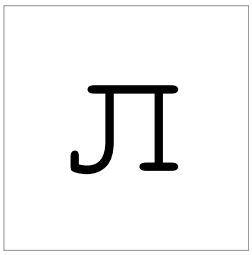
Typeface: Century Schoolbook
Cyrillic [ya]


ж

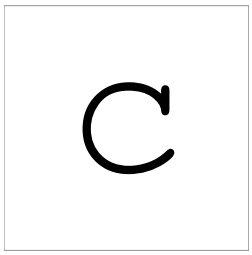
Typeface: Century Schoolbook
Cyrillic [zhe]



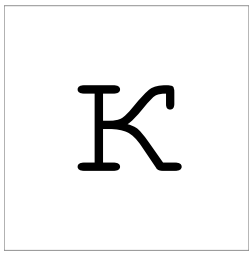
Typeface: Courier New
Cyrillic [de]



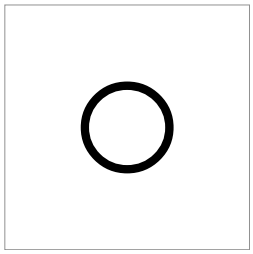
Typeface: Courier New
Cyrillic [el]



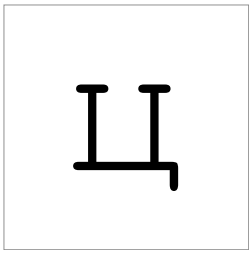
Typeface: Courier New
Cyrillic [es]



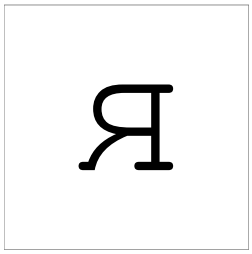
Typeface: Courier New
Cyrillic [ka]



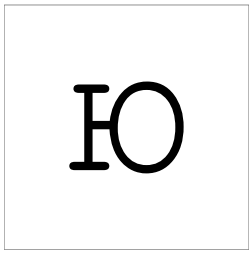
Typeface: Courier New
Cyrillic [o]



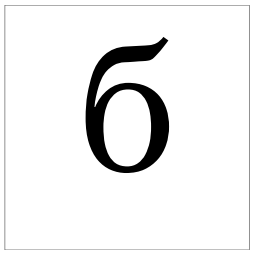
Typeface: Courier New
Cyrillic [tse]



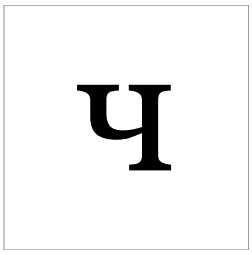
Typeface: Courier New
Cyrillic [ya]



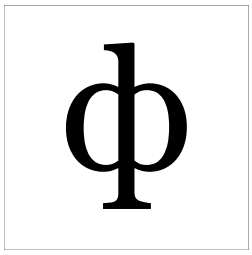
Typeface: Courier New
Cyrillic [yu]



Typeface: Georgia
Cyrillic [be]



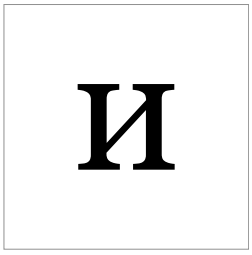
Typeface: Georgia
Cyrillic [che]



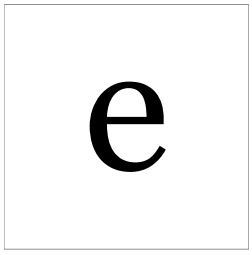
Typeface: Georgia
Cyrillic [ef]

Typeface: Georgia
Cyrillic [em]

Typeface: Georgia
Cyrillic [en]

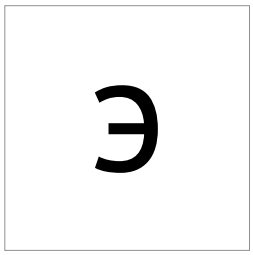


Typeface: Georgia
Cyrillic [i]

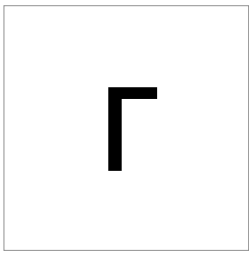


Typeface: Georgia
Cyrillic [ie]

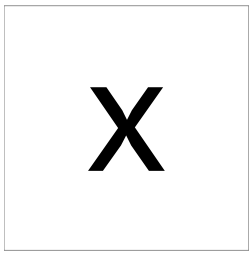
Typeface: Georgia
Cyrillic [ze]



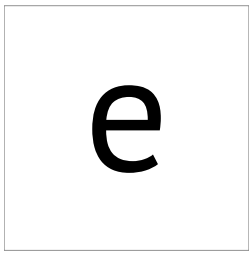
Typeface: PT Sans
Cyrillic [e]



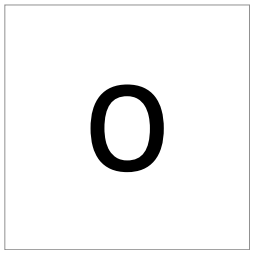
Typeface: PT Sans
Cyrillic [ghe]



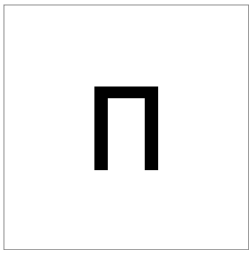
Typeface: PT Sans
Cyrillic [ha]



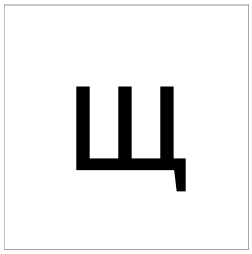
Typeface: PT Sans
Cyrillic [ie]



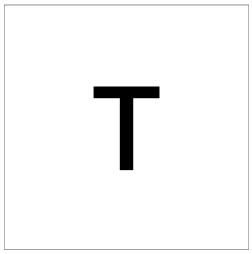
Typeface: PT Sans
Cyrillic [o]



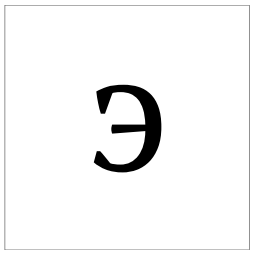
Typeface: PT Sans
Cyrillic [pe]



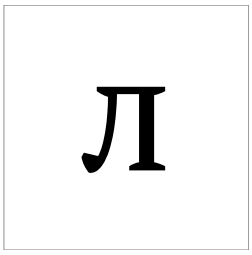
Typeface: PT Sans
Cyrillic [shcha]



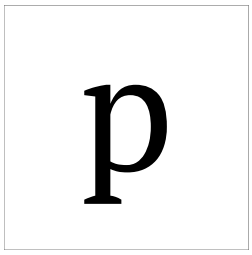
Typeface: PT Sans
Cyrillic [te]



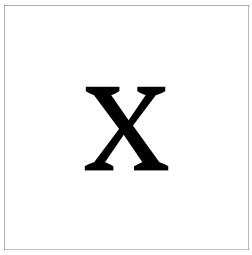
Typeface: PT Serif
Cyrillic [e]



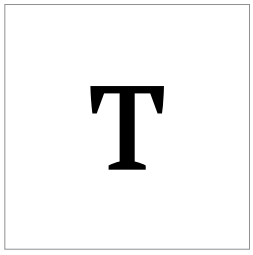
Typeface: PT Serif
Cyrillic [el]



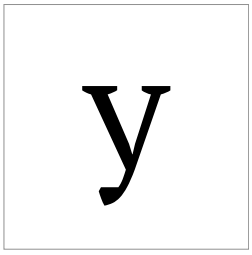
Typeface: PT Serif
Cyrillic [er]



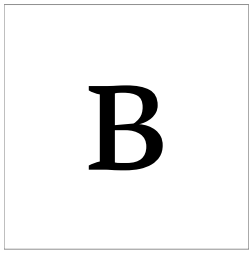
Typeface: PT Serif
Cyrillic [ha]



Typeface: PT Serif
Cyrillic [te]

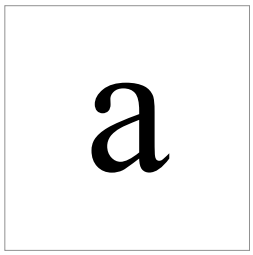


Typeface: PT Serif
Cyrillic [u]

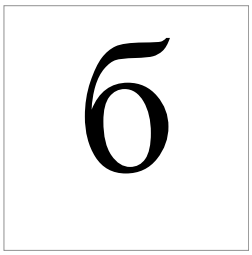


Typeface: PT Serif
Cyrillic [ve]

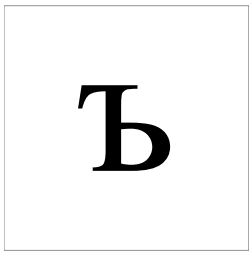
Typeface: PT Serif
Cyrillic [ze]



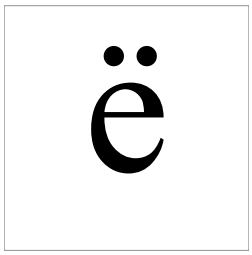
Typeface: Times New Roman
Cyrillic [a]



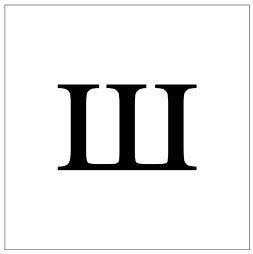
Typeface: Times New Roman
Cyrillic [be]



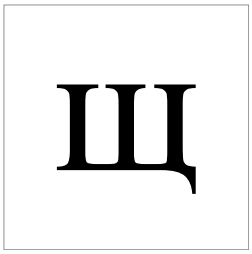
Typeface: Times New Roman
Cyrillic [hard sign]



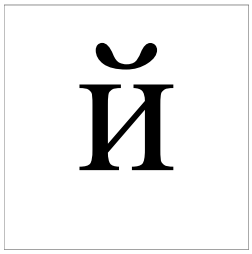
Typeface: Times New Roman
Cyrillic [io]



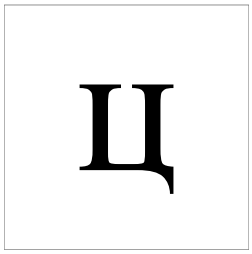
Typeface: Times New Roman
Cyrillic [sha]



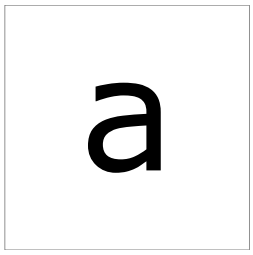
Typeface: Times New Roman
Cyrillic [shcha]



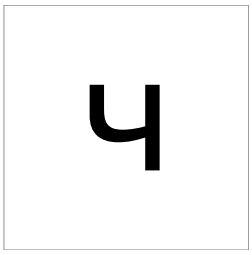
Typeface: Times New Roman
Cyrillic [short i]



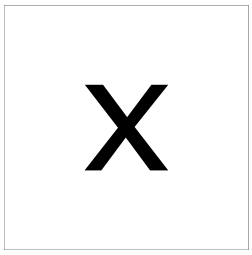
Typeface: Times New Roman
Cyrillic [tse]



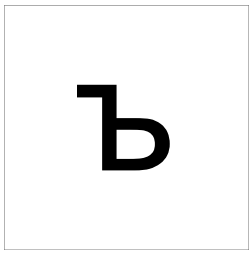
Typeface: Verdana
Cyrillic [a]



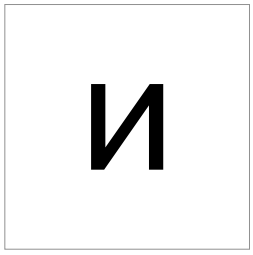
Typeface: Verdana
Cyrillic [che]



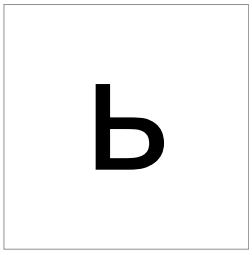
Typeface: Verdana
Cyrillic [ha]



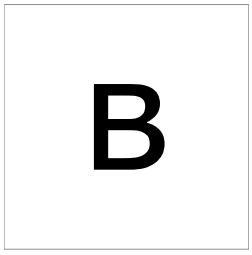
Typeface: Verdana
Cyrillic [hard sign]



Typeface: Verdana
Cyrillic [i]



Typeface: Verdana
Cyrillic [soft sign]



Typeface: Verdana
Cyrillic [ve]

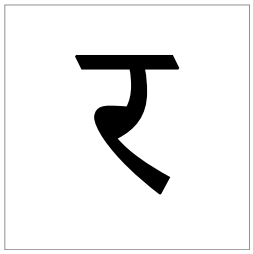
Typeface: Verdana
Cyrillic [yeru]

Typeface: Adobe Devanagari
Devanagari [A]

Typeface: Adobe Devanagari
Devanagari [Ba]

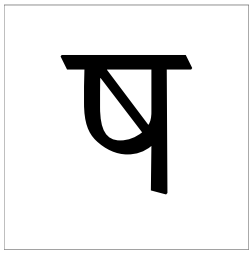
Typeface: Adobe Devanagari
Devanagari [Gha]

Typeface: Adobe Devanagari
Devanagari [Kha]



Typeface: Adobe Devanagari
Devanagari [Ra]

Typeface: Adobe Devanagari
Devanagari [Sa]



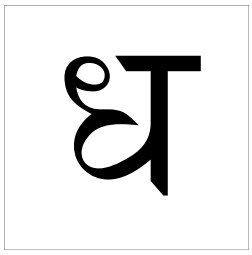
Typeface: Adobe Devanagari
Devanagari [Ssa]

Typeface: Adobe Devanagari
Devanagari [Ta]

Typeface: Devanagari MT
Devanagari [Ca]

Typeface: Devanagari MT
Devanagari [Cha]

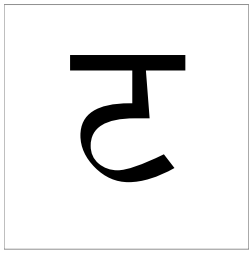
Typeface: Devanagari MT
Devanagari [Dda]



Typeface: Devanagari MT
Devanagari [Dha]

Typeface: Devanagari MT
Devanagari [Gha]

Typeface: Devanagari MT
Devanagari [Ha]

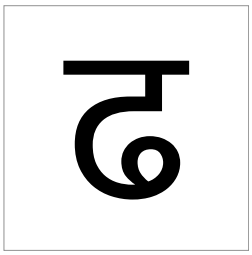


Typeface: Devanagari MT
Devanagari [Tta]

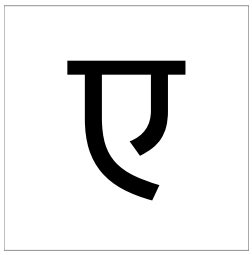
Typeface: Devanagari MT
Devanagari [U]

भ

Typeface: Ek Mukta
Devanagari [Bha]


ढ

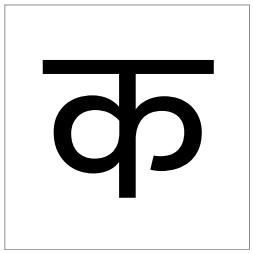
Typeface: Ek Mukta
Devanagari [Ddha]


ए

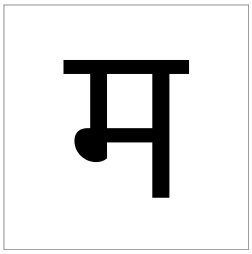
Typeface: Ek Mukta
Devanagari [E]

ज

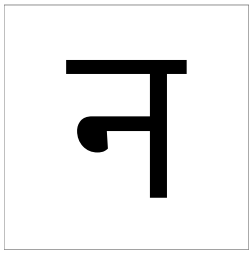
Typeface: Ek Mukta
Devanagari [Ja]


क

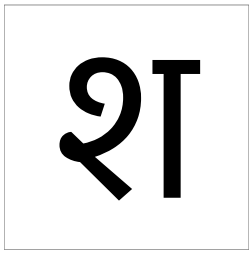
Typeface: Ek Mukta
Devanagari [Ka]


म

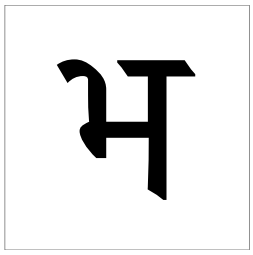
Typeface: Ek Mukta
Devanagari [Ma]


न

Typeface: Ek Mukta
Devanagari [Na]


श

Typeface: Ek Mukta
Devanagari [Sha]


भ

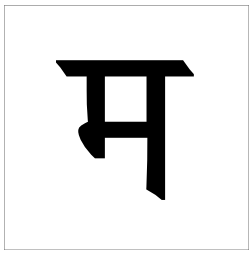
Typeface: ITF Devanagari
Devanagari [Bha]

च

Typeface: ITF Devanagari
Devanagari [Ca]

झ

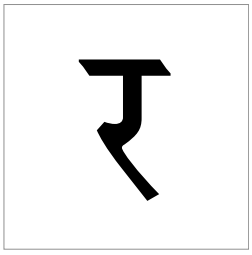
Typeface: ITF Devanagari
Devanagari [Jha]


म

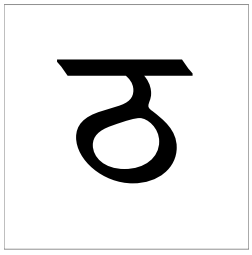
Typeface: ITF Devanagari
Devanagari [Ma]

ञ

Typeface: ITF Devanagari
Devanagari [Nya]


र

Typeface: ITF Devanagari
Devanagari [Ra]

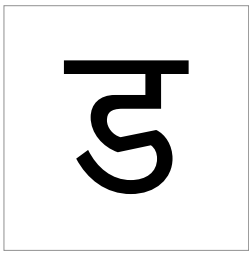

ठ

Typeface: ITF Devanagari
Devanagari [Ttha]

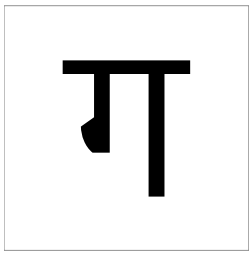
व

Typeface: ITF Devanagari
Devanagari [Va]

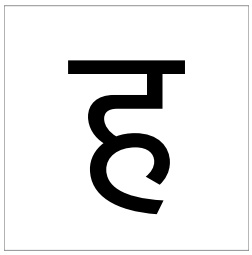
Typeface: Kohinoor Devanagari
Devanagari [Ba]



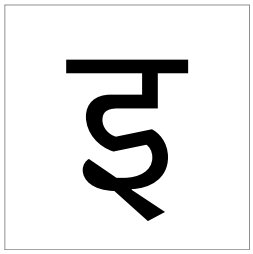
Typeface: Kohinoor Devanagari
Devanagari [Dda]



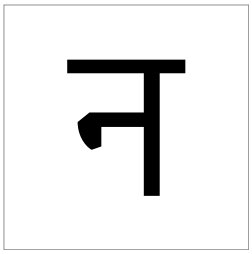
Typeface: Kohinoor Devanagari
Devanagari [Ga]



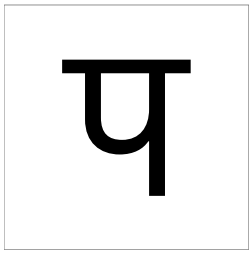
Typeface: Kohinoor Devanagari
Devanagari [Ha]



Typeface: Kohinoor Devanagari
Devanagari [I]

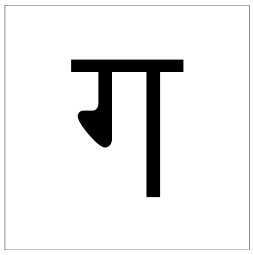


Typeface: Kohinoor Devanagari
Devanagari [Na]

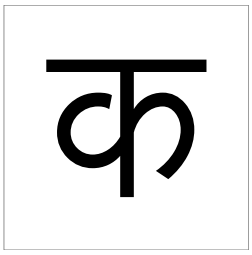


Typeface: Kohinoor Devanagari
Devanagari [Pa]

Typeface: Kohinoor Devanagari
Devanagari [Sa]

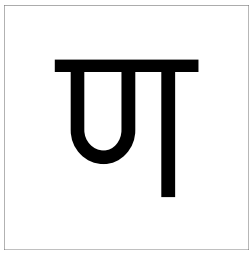


Typeface: Lohit Devanagari
Devanagari [Ga]



Typeface: Lohit Devanagari
Devanagari [Ka]

Typeface: Lohit Devanagari
Devanagari [La]



Typeface: Lohit Devanagari
Devanagari [Nna]

Typeface: Lohit Devanagari
Devanagari [Pa]

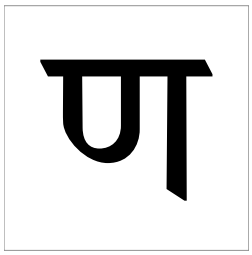
Typeface: Lohit Devanagari
Devanagari [Pha]

Typeface: Lohit Devanagari
Devanagari [Tha]

Typeface: Lohit Devanagari
Devanagari [Ya]

Typeface: Murty Hindi Devanagari [Da]

Typeface: Murty Hindi Devanagari [Ddha]



Typeface: Murty Hindi Devanagari [Nna]

Typeface: Murty Hindi Devanagari [Sha]

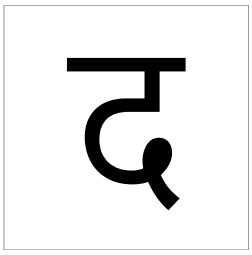
Typeface: Murty Hindi Devanagari [Ssa]

Typeface: Murty Hindi Devanagari [Ta]

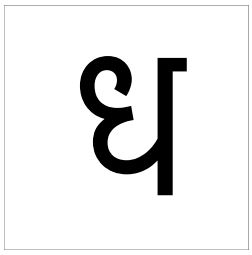
Typeface: Murty Hindi Devanagari [Tha]

Typeface: Murty Hindi Devanagari [Ya]

Typeface: Nirmala UI Devanagari [Cha]



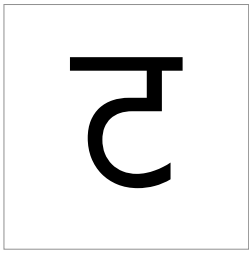
Typeface: Nirmala UI Devanagari [Da]



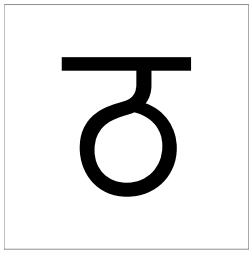
Typeface: Nirmala UI Devanagari [Dha]

Typeface: Nirmala UI Devanagari [La]

Typeface: Nirmala UI Devanagari [Nga]

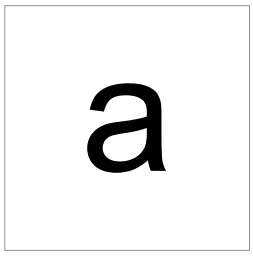


Typeface: Nirmala UI Devanagari [Tta]

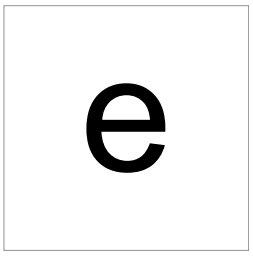


Typeface: Nirmala UI Devanagari [Ttha]

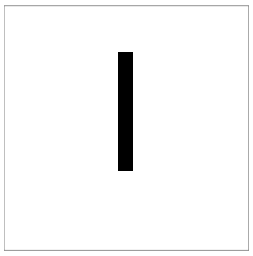
Typeface: Nirmala UI Devanagari [U]



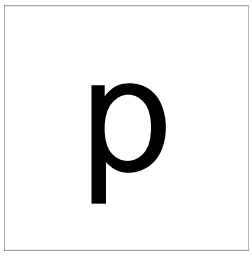
Typeface: Arial
Latin



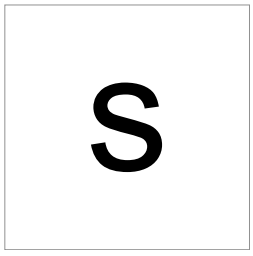
Typeface: Arial
Latin



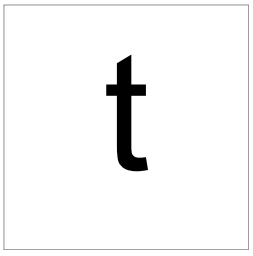
Typeface: Arial
Latin



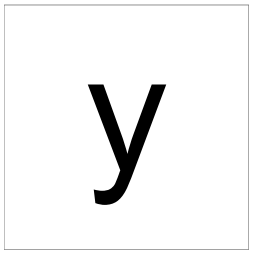
Typeface: Arial
Latin



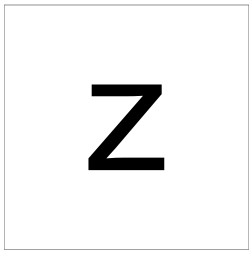
Typeface: Arial
Latin



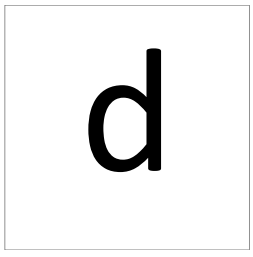
Typeface: Arial
Latin



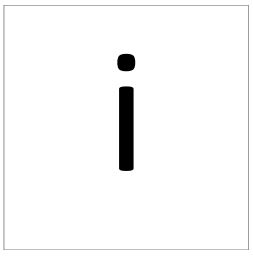
Typeface: Arial
Latin



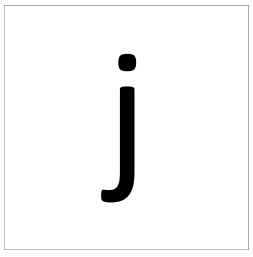
Typeface: Arial
Latin



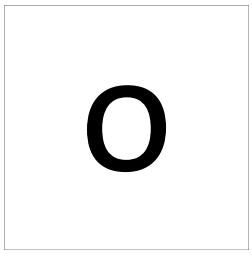
Typeface: Calibri
Latin



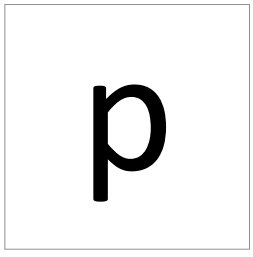
Typeface: Calibri
Latin



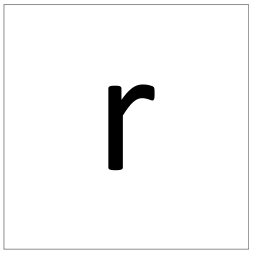
Typeface: Calibri
Latin



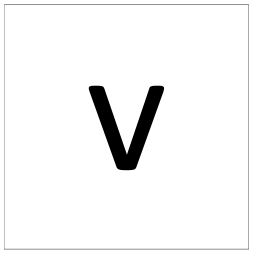
Typeface: Calibri
Latin



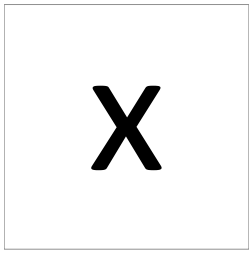
Typeface: Calibri
Latin



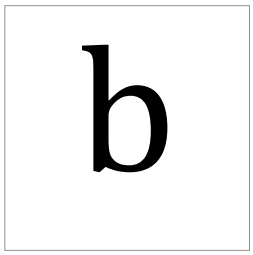
Typeface: Calibri
Latin



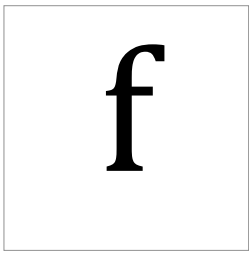
Typeface: Calibri
Latin



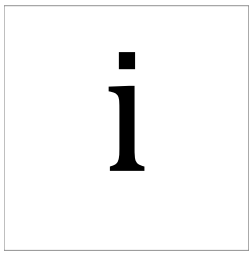
Typeface: Calibri
Latin



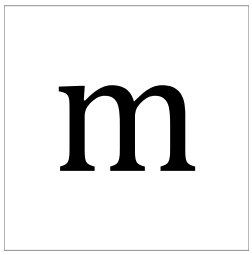
Typeface: Cambria
Latin



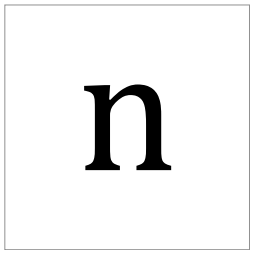
Typeface: Cambria
Latin



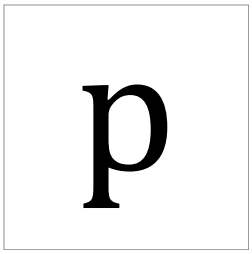
Typeface: Cambria
Latin



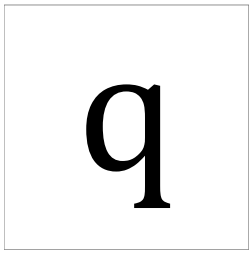
Typeface: Cambria
Latin



Typeface: Cambria
Latin

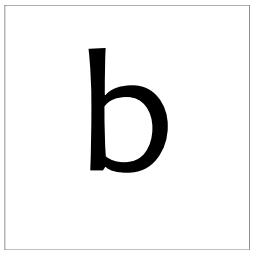


Typeface: Cambria
Latin

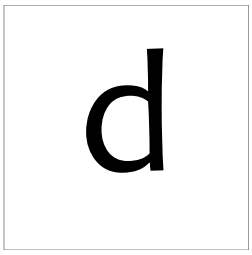


Typeface: Cambria
Latin

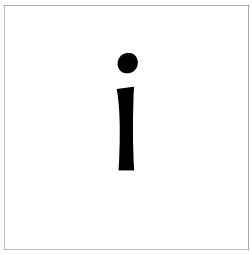
Typeface: Cambria
Latin



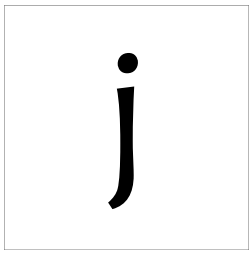
Typeface: Candara
Latin



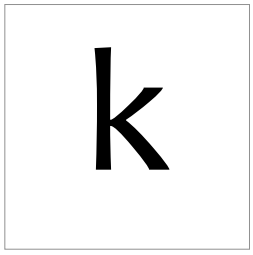
Typeface: Candara
Latin



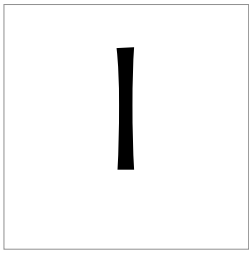
Typeface: Candara
Latin



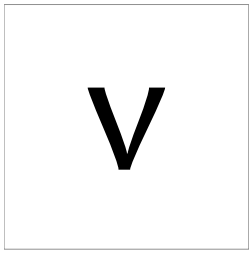
Typeface: Candara
Latin



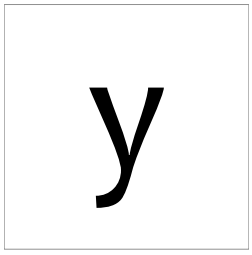
Typeface: Candara
Latin



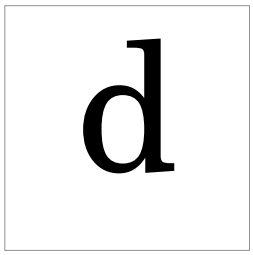
Typeface: Candara
Latin



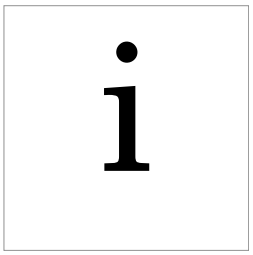
Typeface: Candara
Latin



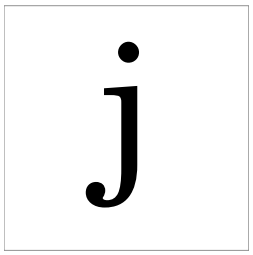
Typeface: Candara
Latin



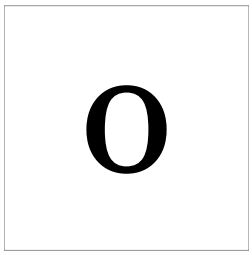
Typeface: Century Schoolbook
Latin



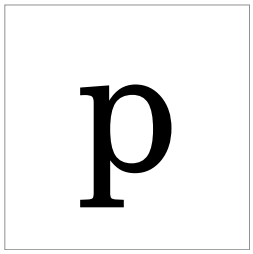
Typeface: Century Schoolbook
Latin



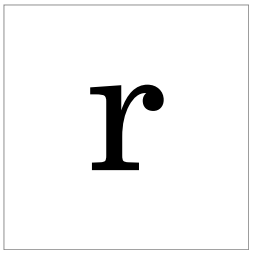
Typeface: Century Schoolbook
Latin



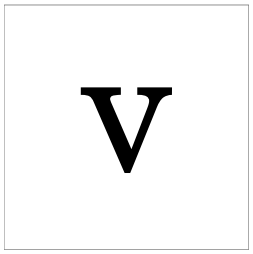
Typeface: Century Schoolbook
Latin



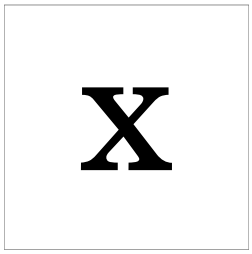
Typeface: Century Schoolbook
Latin



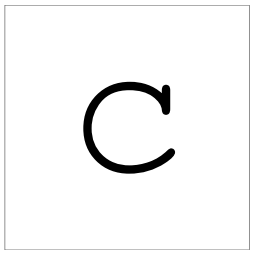
Typeface: Century Schoolbook
Latin



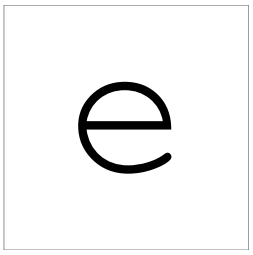
Typeface: Century Schoolbook
Latin



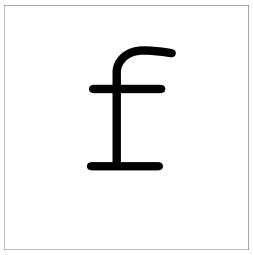
Typeface: Century Schoolbook
Latin



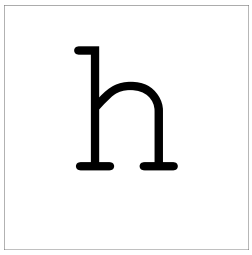
Typeface: Courier New
Latin



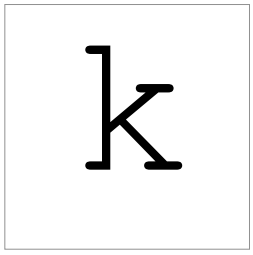
Typeface: Courier New
Latin



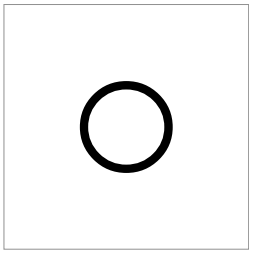
Typeface: Courier New
Latin



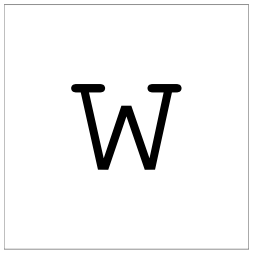
Typeface: Courier New
Latin



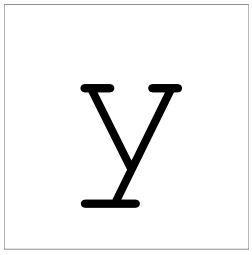
Typeface: Courier New
Latin



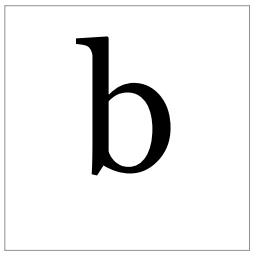
Typeface: Courier New
Latin



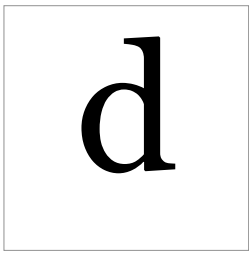
Typeface: Courier New
Latin



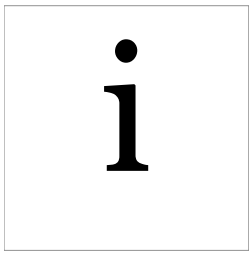
Typeface: Courier New
Latin



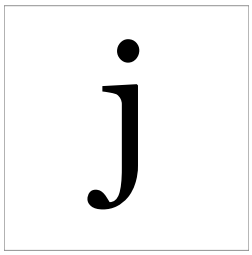
Typeface: Georgia
Latin



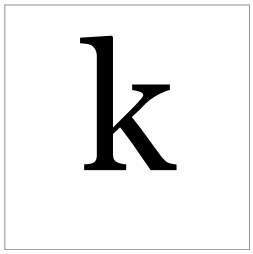
Typeface: Georgia
Latin



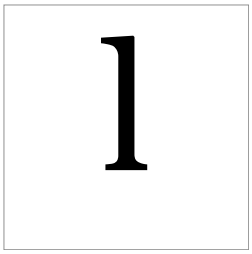
Typeface: Georgia
Latin



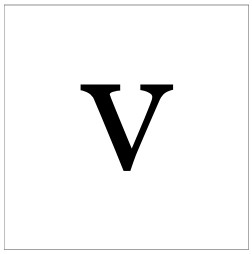
Typeface: Georgia
Latin



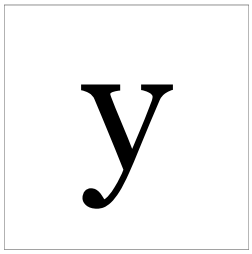
Typeface: Georgia
Latin



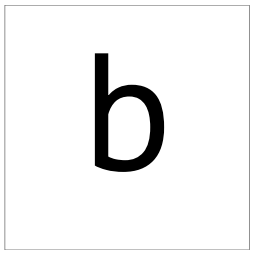
Typeface: Georgia
Latin



Typeface: Georgia
Latin

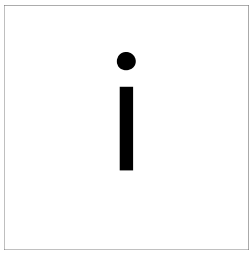


Typeface: Georgia
Latin

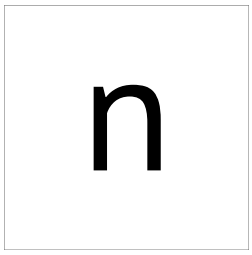


Typeface: PT Sans
Latin

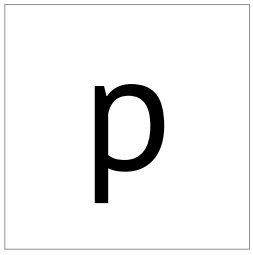
Typeface: PT Sans
Latin



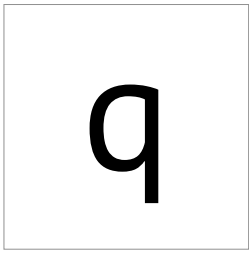
Typeface: PT Sans
Latin



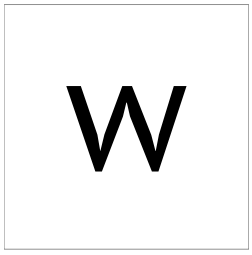
Typeface: PT Sans
Latin



Typeface: PT Sans
Latin

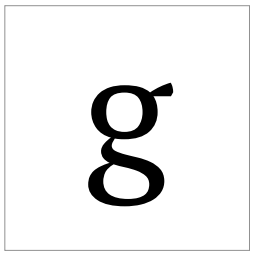


Typeface: PT Sans
Latin

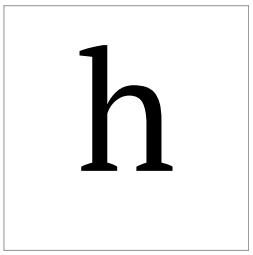


Typeface: PT Sans
Latin

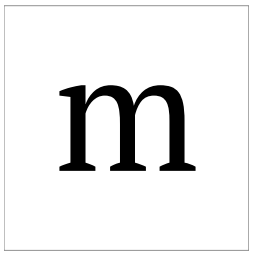
Typeface: PT Sans
Latin



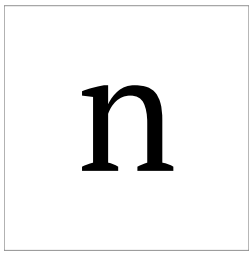
Typeface: PT Serif
Latin



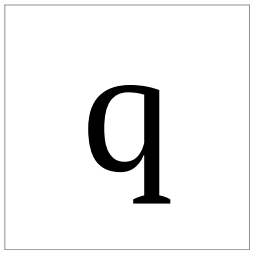
Typeface: PT Serif
Latin



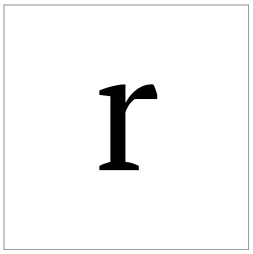
Typeface: PT Serif
Latin



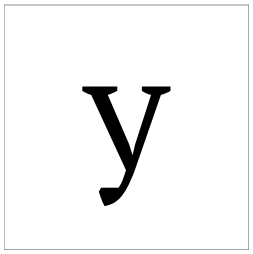
Typeface: PT Serif
Latin



Typeface: PT Serif
Latin

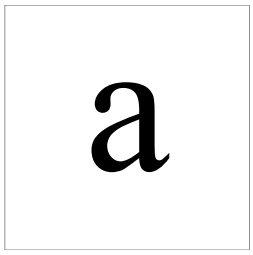


Typeface: PT Serif
Latin

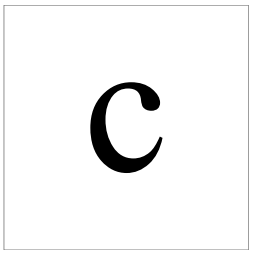


Typeface: PT Serif
Latin

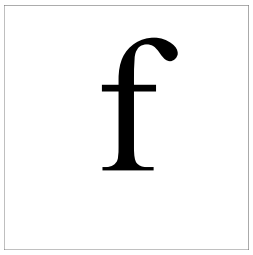
Typeface: PT Serif
Latin



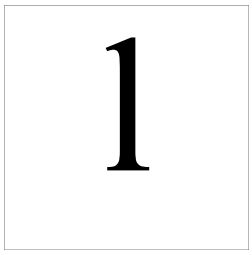
Typeface: Times New Roman
Latin



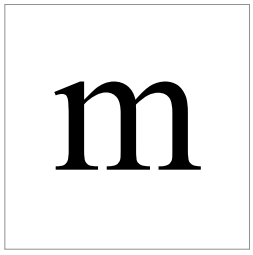
Typeface: Times New Roman
Latin



Typeface: Times New Roman
Latin

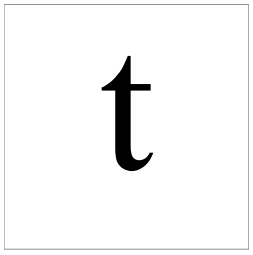


Typeface: Times New Roman
Latin

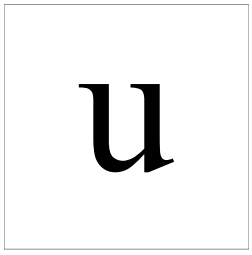


Typeface: Times New Roman
Latin

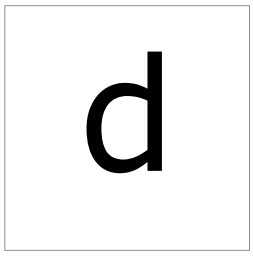
Typeface: Times New Roman
Latin



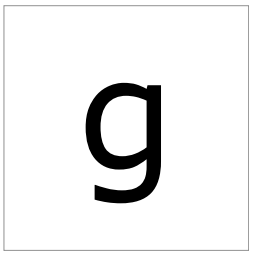
Typeface: Times New Roman
Latin



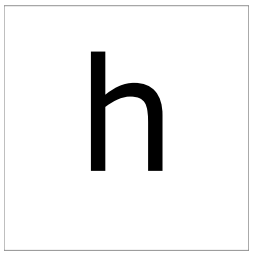
Typeface: Times New Roman
Latin



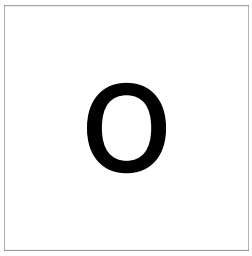
Typeface: Verdana
Latin



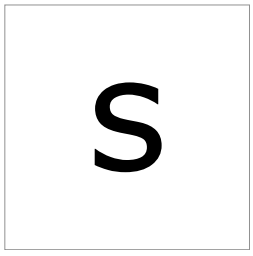
Typeface: Verdana
Latin



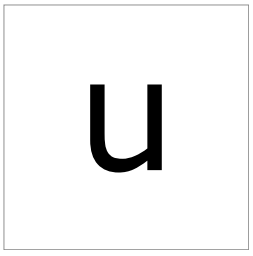
Typeface: Verdana
Latin



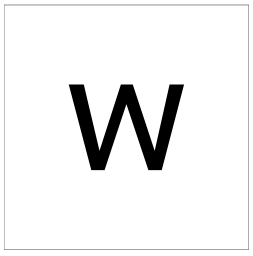
Typeface: Verdana
Latin



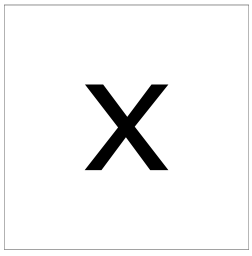
Typeface: Verdana
Latin



Typeface: Verdana
Latin



Typeface: Verdana
Latin



Typeface: Verdana
Latin

Appendix 2: observations and predictions

Responses (observations) and predictions for all the triplet trials from the studies described in chapter 5. The columns are:

- **character:** characters' name and appearance
- **studies – count:** observed response counts
- **studies – freq.:** observed response frequencies
- **studies – confidence interval:** binomial confidence intervals
- **model – freq.:** predicted response frequencies

The overall odd one out is marked with a grey background. This is done separately for the observations and predictions. Thus, misalignment of the grey backgrounds indicates disagreement between the observed and predicted responses.

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	7	0.10	(0.04, 0.2)	0.09
em	м	61	0.88	(0.78, 0.95)	0.78
er	р	1	0.01	(0.0, 0.08)	0.13

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	19	0.28	(0.17, 0.4)	0.19
em	м	45	0.65	(0.53, 0.76)	0.56
es	с	5	0.07	(0.02, 0.16)	0.25

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	52	0.75	(0.64, 0.85)	0.59
em	м	3	0.04	(0.01, 0.12)	0.14
short i	й	14	0.20	(0.12, 0.32)	0.27

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	43	0.62	(0.5, 0.74)	0.47
em	м	20	0.29	(0.19, 0.41)	0.28
u	у	6	0.09	(0.03, 0.18)	0.25

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	9	0.13	(0.06, 0.23)	0.16
em	м	57	0.83	(0.72, 0.91)	0.65
yu	ю	3	0.04	(0.01, 0.12)	0.19

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	55	0.80	(0.68, 0.88)	0.54
em	м	5	0.07	(0.02, 0.16)	0.30
zhe	ж	9	0.13	(0.06, 0.23)	0.16

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	20	0.29	(0.19, 0.41)	0.25
er	р	1	0.01	(0.0, 0.08)	0.17
es	с	48	0.70	(0.57, 0.8)	0.57

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	5	0.07	(0.02, 0.16)	0.13
er	р	2	0.03	(0.0, 0.1)	0.09
short i	й	62	0.90	(0.8, 0.96)	0.78

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	11	0.16	(0.08, 0.27)	0.18
er	р	2	0.03	(0.0, 0.1)	0.09
u	у	56	0.81	(0.7, 0.9)	0.72

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	17	0.25	(0.15, 0.36)	0.24
er	р	7	0.10	(0.04, 0.2)	0.28
yu	ю	45	0.65	(0.53, 0.76)	0.47

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	7	0.10	(0.04, 0.2)	0.09
er	р	4	0.06	(0.02, 0.14)	0.21
zhe	ж	58	0.84	(0.73, 0.92)	0.70

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	18	0.26	(0.16, 0.38)	0.22
es	с	2	0.03	(0.0, 0.1)	0.20
short i	й	49	0.71	(0.59, 0.81)	0.58

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	15	0.22	(0.13, 0.33)	0.23
es	с	19	0.28	(0.17, 0.4)	0.26
u	у	35	0.51	(0.38, 0.63)	0.50

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	26	0.38	(0.26, 0.5)	0.35
es	с	37	0.54	(0.41, 0.66)	0.42
yu	ю	6	0.09	(0.03, 0.18)	0.23

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	17	0.25	(0.15, 0.36)	0.19
es	с	9	0.13	(0.06, 0.23)	0.32
zhe	ж	43	0.62	(0.5, 0.74)	0.49

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	45	0.65	(0.53, 0.76)	0.43
short i	й	23	0.33	(0.22, 0.46)	0.36
u	у	1	0.01	(0.0, 0.08)	0.21

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	9	0.13	(0.06, 0.23)	0.19
short i	й	56	0.81	(0.7, 0.9)	0.67
yu	ю	4	0.06	(0.02, 0.14)	0.15

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	35	0.51	(0.38, 0.63)	0.49
short i	й	26	0.38	(0.26, 0.5)	0.36
zhe	ж	8	0.12	(0.05, 0.22)	0.15

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	8	0.12	(0.05, 0.22)	0.06
u	у	54	0.78	(0.67, 0.87)	0.66
yu	ю	7	0.10	(0.04, 0.2)	0.28

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	31	0.45	(0.33, 0.57)	0.34
u	у	17	0.25	(0.15, 0.36)	0.35
zhe	ж	21	0.30	(0.2, 0.43)	0.30

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	7	0.10	(0.04, 0.2)	0.33
yu	ю	8	0.12	(0.05, 0.22)	0.18
zhe	ж	54	0.78	(0.67, 0.87)	0.49

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	53	0.77	(0.65, 0.86)	0.73
er	р	13	0.19	(0.1, 0.3)	0.13
es	с	3	0.04	(0.01, 0.12)	0.14

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	4	0.06	(0.02, 0.14)	0.20
er	р	49	0.71	(0.59, 0.81)	0.59
short i	й	16	0.23	(0.14, 0.35)	0.21

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	35	0.51	(0.38, 0.63)	0.38
er	р	27	0.39	(0.28, 0.52)	0.45
u	у	7	0.10	(0.04, 0.2)	0.16

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	49	0.71	(0.59, 0.81)	0.73
er	р	13	0.19	(0.1, 0.3)	0.14
yu	ю	7	0.10	(0.04, 0.2)	0.13

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	2	0.03	(0.0, 0.1)	0.24
er	р	53	0.77	(0.65, 0.86)	0.59
zhe	ж	14	0.20	(0.12, 0.32)	0.17

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	7	0.10	(0.04, 0.2)	0.16
es	с	46	0.67	(0.54, 0.78)	0.65
short i	й	16	0.23	(0.14, 0.35)	0.19

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	5	0.07	(0.02, 0.16)	0.28
es	с	41	0.59	(0.47, 0.71)	0.55
u	у	23	0.33	(0.22, 0.46)	0.16

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	52	0.75	(0.64, 0.85)	0.77
es	с	7	0.10	(0.04, 0.2)	0.12
yu	ю	10	0.14	(0.07, 0.25)	0.11

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	1	0.01	(0.0, 0.08)	0.24
es	с	51	0.74	(0.62, 0.84)	0.63
zhe	ж	17	0.25	(0.15, 0.36)	0.12

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	11	0.16	(0.08, 0.27)	0.31
short i	й	14	0.20	(0.12, 0.32)	0.27
u	у	44	0.64	(0.51, 0.75)	0.42

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	2	0.03	(0.0, 0.1)	0.18
short i	й	21	0.30	(0.2, 0.43)	0.21
yu	ю	46	0.67	(0.54, 0.78)	0.61

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	1	0.01	(0.0, 0.08)	0.23
short i	й	35	0.51	(0.38, 0.63)	0.44
zhe	ж	33	0.48	(0.36, 0.6)	0.33

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	6	0.09	(0.03, 0.18)	0.12
u	у	29	0.42	(0.3, 0.55)	0.28
yu	ю	34	0.49	(0.37, 0.62)	0.60

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	–	–	–	0.15
u	у	45	0.65	–	0.56
zhe	ж	24	0.35	–	0.28

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	4	0.06	(0.02, 0.14)	0.46
yu	ю	52	0.75	(0.64, 0.85)	0.49
zhe	ж	13	0.19	(0.1, 0.3)	0.05

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	9	0.13	(0.06, 0.23)	0.14
es	с	1	0.01	(0.0, 0.08)	0.16
short i	й	59	0.86	(0.75, 0.93)	0.70

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	3	0.04	(0.01, 0.12)	0.13
es	с	23	0.33	(0.22, 0.46)	0.24
u	у	43	0.62	(0.5, 0.74)	0.63

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	10	0.14	(0.07, 0.25)	0.32
es	с	37	0.54	(0.41, 0.66)	0.38
yu	ю	22	0.32	(0.21, 0.44)	0.29

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	9	0.13	(0.06, 0.23)	0.17
es	с	2	0.03	(0.0, 0.1)	0.18
zhe	ж	58	0.84	(0.73, 0.92)	0.65

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	27	0.39	(0.28, 0.52)	0.39
short i	й	39	0.57	(0.44, 0.68)	0.42
u	у	3	0.04	(0.01, 0.12)	0.19

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	8	0.12	(0.05, 0.22)	0.15
short i	й	57	0.83	(0.72, 0.91)	0.70
yu	ю	4	0.06	(0.02, 0.14)	0.15

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	34	0.49	(0.37, 0.62)	0.53
short i	й	21	0.30	(0.2, 0.43)	0.26
zhe	ж	14	0.20	(0.12, 0.32)	0.21

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	5	0.07	(0.02, 0.16)	0.02
u	у	48	0.70	(0.57, 0.8)	0.69
yu	ю	16	0.23	(0.14, 0.35)	0.29

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	22	0.32	(0.21, 0.44)	0.35
u	у	5	0.07	(0.02, 0.16)	0.24
zhe	ж	42	0.61	(0.48, 0.72)	0.40

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	11	0.16	(0.08, 0.27)	0.33
yu	ю	1	0.01	(0.0, 0.08)	0.09
zhe	ж	57	0.83	(0.72, 0.91)	0.58

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
es	с	32	0.46	(0.34, 0.59)	0.46
short i	й	22	0.32	(0.21, 0.44)	0.34
u	у	15	0.22	(0.13, 0.33)	0.20

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
es	с	3	0.04	(0.01, 0.12)	0.13
short i	й	59	0.86	(0.75, 0.93)	0.73
yu	ю	7	0.10	(0.04, 0.2)	0.13

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
es	с	33	0.48	(0.36, 0.6)	0.56
short i	й	21	0.30	(0.2, 0.43)	0.26
zhe	ж	15	0.22	(0.13, 0.33)	0.18

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
es	с	2	0.03	(0.0, 0.1)	0.02
u	у	57	0.83	(0.72, 0.91)	0.76
yu	ю	10	0.14	(0.07, 0.25)	0.23

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
es	с	32	0.46	(0.34, 0.59)	0.41
u	у	12	0.17	(0.09, 0.28)	0.26
zhe	ж	25	0.36	(0.25, 0.49)	0.32

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
es	с	6	0.09	(0.03, 0.18)	0.37
yu	ю	5	0.07	(0.02, 0.16)	0.07
zhe	ж	58	0.84	(0.73, 0.92)	0.56

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
short i	й	21	0.30	(0.2, 0.43)	0.19
u	у	13	0.19	(0.1, 0.3)	0.29
yu	ю	35	0.51	(0.38, 0.63)	0.52

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
short i	й	14	0.20	(0.12, 0.32)	0.28
u	у	16	0.23	(0.14, 0.35)	0.42
zhe	ж	39	0.57	(0.44, 0.68)	0.29

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
short i	й	23	0.33	(0.22, 0.46)	0.46
yu	ю	28	0.41	(0.29, 0.53)	0.43
zhe	ж	18	0.26	(0.16, 0.38)	0.11

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
u	у	25	0.36	(0.25, 0.49)	0.59
yu	ю	28	0.41	(0.29, 0.53)	0.35
zhe	ж	16	0.23	(0.14, 0.35)	0.06

Arial (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	48	0.77	(0.65, 0.87)	0.47
de	д	4	0.06	(0.02, 0.16)	0.07
ghe	г	10	0.16	(0.08, 0.28)	0.46

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	2	0.03	(0.0, 0.11)	0.19
de	д	38	0.61	(0.48, 0.73)	0.40
io	ё	22	0.35	(0.24, 0.49)	0.41

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	27	0.44	(0.31, 0.57)	0.24
de	д	25	0.40	(0.28, 0.54)	0.46
ka	к	10	0.16	(0.08, 0.28)	0.29

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	55	0.89	(0.78, 0.95)	0.38
de	д	3	0.05	(0.01, 0.13)	0.14
pe	п	4	0.06	(0.02, 0.16)	0.48

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	9	0.15	(0.07, 0.26)	0.27
de	д	47	0.76	(0.63, 0.86)	0.50
ya	я	6	0.10	(0.04, 0.2)	0.23

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	22	0.35	(0.24, 0.49)	0.24
de	д	19	0.31	(0.2, 0.44)	0.44
zhe	ж	21	0.34	(0.22, 0.47)	0.32

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	4	0.06	(0.02, 0.16)	0.13
ghe	г	45	0.73	(0.6, 0.83)	0.58
io	ё	13	0.21	(0.12, 0.33)	0.29

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	35	0.56	(0.43, 0.69)	0.37
ghe	г	20	0.32	(0.21, 0.45)	0.58
ka	к	7	0.11	(0.05, 0.22)	0.04

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	54	0.87	(0.76, 0.94)	0.83
ghe	г	3	0.05	(0.01, 0.13)	0.13
pe	п	5	0.08	(0.03, 0.18)	0.04

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	13	0.21	(0.12, 0.33)	0.37
ghe	г	43	0.69	(0.56, 0.8)	0.59
ya	я	6	0.10	(0.04, 0.2)	0.04

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	17	0.27	(0.17, 0.4)	0.38
ghe	г	22	0.35	(0.24, 0.49)	0.53
zhe	ж	23	0.37	(0.25, 0.5)	0.09

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	1	0.02	(0.0, 0.09)	0.23
io	ё	28	0.45	(0.32, 0.58)	0.46
ka	к	33	0.53	(0.4, 0.66)	0.31

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	6	0.10	(0.04, 0.2)	0.15
io	ё	13	0.21	(0.12, 0.33)	0.27
pe	п	43	0.69	(0.56, 0.8)	0.58

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	1	0.02	(0.0, 0.09)	0.26
io	ё	44	0.71	(0.58, 0.82)	0.49
ya	я	17	0.27	(0.17, 0.4)	0.26

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	5	0.08	(0.03, 0.18)	0.25
io	ё	24	0.39	(0.27, 0.52)	0.44
zhe	ж	33	0.53	(0.4, 0.66)	0.32

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	35	0.56	(0.43, 0.69)	0.44
ka	к	5	0.08	(0.03, 0.18)	0.03
pe	п	22	0.35	(0.24, 0.49)	0.53

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	31	0.50	(0.37, 0.63)	0.48
ka	к	29	0.47	(0.34, 0.6)	0.32
ya	я	2	0.03	(0.0, 0.11)	0.20

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	56	0.90	(0.8, 0.96)	0.59
ka	к	1	0.02	(0.0, 0.09)	0.19
zhe	ж	5	0.08	(0.03, 0.18)	0.22

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	12	0.19	(0.1, 0.31)	0.34
pe	п	43	0.69	(0.56, 0.8)	0.62
ya	я	7	0.11	(0.05, 0.22)	0.03

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	25	0.40	(0.28, 0.54)	0.44
pe	п	23	0.37	(0.25, 0.5)	0.48
zhe	ж	14	0.23	(0.13, 0.35)	0.08

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	22	0.35	(0.24, 0.49)	0.47
ya	я	3	0.05	(0.01, 0.13)	0.20
zhe	ж	37	0.60	(0.46, 0.72)	0.33

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	д	4	0.06	(0.02, 0.16)	0.03
ghe	г	5	0.08	(0.03, 0.18)	0.35
io	ё	53	0.85	(0.74, 0.93)	0.62

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	д	11	0.18	(0.09, 0.3)	0.31
ghe	г	6	0.10	(0.04, 0.2)	0.34
ka	к	45	0.73	(0.6, 0.83)	0.35

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	д	33	0.53	(0.4, 0.66)	0.63
ghe	г	24	0.39	(0.27, 0.52)	0.12
pe	п	5	0.08	(0.03, 0.18)	0.25

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	д	8	0.13	(0.06, 0.24)	0.32
ghe	г	12	0.19	(0.1, 0.31)	0.37
ya	я	42	0.68	(0.55, 0.79)	0.31

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	д	6	0.10	(0.04, 0.2)	0.33
ghe	г	4	0.06	(0.02, 0.16)	0.25
zhe	ж	52	0.84	(0.72, 0.92)	0.42

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	д	9	0.15	(0.07, 0.26)	0.35
io	ё	50	0.81	(0.69, 0.9)	0.42
ka	к	3	0.05	(0.01, 0.13)	0.23

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	д	–	–	–	0.11
io	ё	58	0.94	–	0.51
pe	п	4	0.06	–	0.38

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	д	18	0.29	(0.18, 0.42)	0.36
io	ё	40	0.65	(0.51, 0.76)	0.45
ya	я	4	0.06	(0.02, 0.16)	0.19

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	д	10	0.16	(0.08, 0.28)	0.35
io	ё	40	0.65	(0.51, 0.76)	0.40
zhe	ж	12	0.19	(0.1, 0.31)	0.25

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	д	8	0.13	(0.06, 0.24)	0.46
ka	к	47	0.76	(0.63, 0.86)	0.22
pe	п	7	0.11	(0.05, 0.22)	0.32

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	д	40	0.65	(0.51, 0.76)	0.62
ka	к	14	0.23	(0.13, 0.35)	0.23
ya	я	8	0.13	(0.06, 0.24)	0.16

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	д	48	0.77	(0.65, 0.87)	0.71
ka	к	5	0.08	(0.03, 0.18)	0.12
zhe	ж	9	0.15	(0.07, 0.26)	0.17

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	д	4	0.06	(0.02, 0.16)	0.35
pe	п	3	0.05	(0.01, 0.13)	0.41
ya	я	55	0.89	(0.78, 0.95)	0.24

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	4	0.06	(0.02, 0.16)	0.48
pe	П	2	0.03	(0.0, 0.11)	0.23
zhe	Ж	56	0.90	(0.8, 0.96)	0.29

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	37	0.60	(0.46, 0.72)	0.60
ya	Я	5	0.08	(0.03, 0.18)	0.14
zhe	Ж	20	0.32	(0.21, 0.45)	0.26

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	7	0.11	(0.05, 0.22)	0.45
io	ё	51	0.82	(0.7, 0.91)	0.52
ka	К	4	0.06	(0.02, 0.16)	0.03

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	2	0.03	(0.0, 0.11)	0.11
io	ё	55	0.89	(0.78, 0.95)	0.87
pe	П	5	0.08	(0.03, 0.18)	0.03

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	25	0.40	(0.28, 0.54)	0.46
io	ё	34	0.55	(0.42, 0.68)	0.50
ya	Я	3	0.05	(0.01, 0.13)	0.04

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	18	0.29	(0.18, 0.42)	0.44
io	ё	33	0.53	(0.4, 0.66)	0.50
zhe	Ж	11	0.18	(0.09, 0.3)	0.07

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	6	0.10	(0.04, 0.2)	0.31
ka	К	53	0.85	(0.74, 0.93)	0.55
pe	П	3	0.05	(0.01, 0.13)	0.14

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	40	0.65	(0.51, 0.76)	0.66
ka	К	5	0.08	(0.03, 0.18)	0.16
ya	Я	17	0.27	(0.17, 0.4)	0.18

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	50	0.81	(0.69, 0.9)	0.68
ka	К	2	0.03	(0.0, 0.11)	0.14
zhe	Ж	10	0.16	(0.08, 0.28)	0.18

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	6	0.10	(0.04, 0.2)	0.26
pe	П	1	0.02	(0.0, 0.09)	0.17
ya	Я	55	0.89	(0.78, 0.95)	0.57

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	3	0.05	(0.01, 0.13)	0.29
pe	П	1	0.02	(0.0, 0.09)	0.14
zhe	Ж	58	0.94	(0.84, 0.98)	0.57

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	38	0.61	(0.48, 0.73)	0.60
ya	Я	8	0.13	(0.06, 0.24)	0.19
zhe	Ж	16	0.26	(0.16, 0.38)	0.21

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
io	ё	47	0.76	(0.63, 0.86)	0.55
ka	К	7	0.11	(0.05, 0.22)	0.03
pe	П	8	0.13	(0.06, 0.24)	0.41

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
io	ё	49	0.79	(0.67, 0.88)	0.62
ka	К	8	0.13	(0.06, 0.24)	0.22
ya	Я	5	0.08	(0.03, 0.18)	0.16

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
io	ё	51	0.82	(0.7, 0.91)	0.68
ka	К	4	0.06	(0.02, 0.16)	0.15
zhe	Ж	7	0.11	(0.05, 0.22)	0.16

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
io	ё	32	0.52	(0.39, 0.65)	0.47
pe	П	27	0.44	(0.31, 0.57)	0.48
ya	Я	3	0.05	(0.01, 0.13)	0.05

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
io	ё	35	0.56	(0.43, 0.69)	0.53
pe	П	16	0.26	(0.16, 0.38)	0.40
zhe	Ж	11	0.18	(0.09, 0.3)	0.07

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
io	ё	44	0.71	(0.58, 0.82)	0.59
ya	Я	4	0.06	(0.02, 0.16)	0.17
zhe	Ж	14	0.23	(0.13, 0.35)	0.23

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ka	К	6	0.10	(0.04, 0.2)	0.14
pe	П	46	0.74	(0.62, 0.84)	0.64
ya	Я	10	0.16	(0.08, 0.28)	0.22

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ka	К	2	0.03	(0.0, 0.11)	0.17
pe	П	48	0.77	(0.65, 0.87)	0.62
zhe	Ж	12	0.19	(0.1, 0.31)	0.21

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ka	К	3	0.05	(0.01, 0.13)	0.29
ya	Я	43	0.69	(0.56, 0.8)	0.38
zhe	Ж	16	0.26	(0.16, 0.38)	0.32

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
pe	П	40	0.65	(0.51, 0.76)	0.57
ya	Я	7	0.11	(0.05, 0.22)	0.23
zhe	Ж	15	0.24	(0.14, 0.37)	0.20

Century Schoolbook (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	3	0.04	(0.01, 0.13)	0.19
el	Л	4	0.06	(0.02, 0.15)	0.11
es	С	60	0.90	(0.8, 0.96)	0.70

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	10	0.15	(0.07, 0.26)	0.34
el	Л	3	0.04	(0.01, 0.13)	0.24
ka	К	54	0.81	(0.69, 0.89)	0.42

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	–	–	–	0.09
el	Л	7	0.10	–	0.16
o	О	60	0.90	–	0.75

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	7	0.10	(0.04, 0.2)	0.23
el	Л	25	0.37	(0.26, 0.5)	0.22
tse	Ц	35	0.52	(0.4, 0.65)	0.55

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	9	0.13	(0.06, 0.24)	0.22
el	Л	4	0.06	(0.02, 0.15)	0.28
ya	Я	54	0.81	(0.69, 0.89)	0.50

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	3	0.04	(0.01, 0.13)	0.25
el	Л	5	0.07	(0.02, 0.17)	0.33
yu	Ю	59	0.88	(0.78, 0.95)	0.42

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	7	0.10	(0.04, 0.2)	0.22
es	С	52	0.78	(0.66, 0.87)	0.58
ka	К	8	0.12	(0.05, 0.22)	0.21

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	54	0.81	(0.69, 0.89)	0.68
es	С	5	0.07	(0.02, 0.17)	0.17
o	О	8	0.12	(0.05, 0.22)	0.15

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	4	0.06	(0.02, 0.15)	0.10
es	С	61	0.91	(0.82, 0.97)	0.70
tse	Ц	2	0.03	(0.0, 0.1)	0.20

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	14	0.21	(0.12, 0.33)	0.13
es	С	43	0.64	(0.52, 0.76)	0.59
ya	Я	10	0.15	(0.07, 0.26)	0.28

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	38	0.57	(0.44, 0.69)	0.19
es	С	21	0.31	(0.21, 0.44)	0.65
yu	Ю	8	0.12	(0.05, 0.22)	0.16

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	5	0.07	(0.02, 0.17)	0.13
ka	К	12	0.18	(0.1, 0.29)	0.25
o	О	50	0.75	(0.63, 0.84)	0.61

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	6	0.09	(0.03, 0.18)	0.21
ka	К	57	0.85	(0.74, 0.93)	0.39
tse	Ц	4	0.06	(0.02, 0.15)	0.40

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	29	0.43	(0.31, 0.56)	0.49
ka	К	17	0.25	(0.16, 0.37)	0.23
ya	Я	21	0.31	(0.21, 0.44)	0.28

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	27	0.40	(0.28, 0.53)	0.27
ka	К	12	0.18	(0.1, 0.29)	0.42
yu	Ю	28	0.42	(0.3, 0.54)	0.31

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	2	0.03	(0.0, 0.1)	0.00
o	О	59	0.88	(0.78, 0.95)	0.75
tse	Ц	6	0.09	(0.03, 0.18)	0.25

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	18	0.27	(0.17, 0.39)	0.25
o	О	43	0.64	(0.52, 0.76)	0.53
ya	Я	6	0.09	(0.03, 0.18)	0.22

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	45	0.67	(0.55, 0.78)	0.35
o	О	17	0.25	(0.16, 0.37)	0.58
yu	Ю	5	0.07	(0.02, 0.17)	0.07

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	1	0.01	(0.0, 0.08)	0.09
tse	Ц	10	0.15	(0.07, 0.26)	0.43
ya	Я	56	0.84	(0.73, 0.92)	0.48

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	1	0.01	(0.0, 0.08)	0.12
tse	Ц	1	0.01	(0.0, 0.08)	0.45
yu	Ю	65	0.97	(0.9, 1.0)	0.43

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
de	Д	28	0.42	(0.3, 0.54)	0.41
ya	Я	10	0.15	(0.07, 0.26)	0.38
yu	Ю	29	0.43	(0.31, 0.56)	0.21

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	Л	5	0.07	(0.02, 0.17)	0.17
es	С	58	0.87	(0.76, 0.94)	0.60
ka	К	4	0.06	(0.02, 0.15)	0.23

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	Л	55	0.82	(0.71, 0.9)	0.66
es	С	1	0.01	(0.0, 0.08)	0.09
o	О	11	0.16	(0.08, 0.27)	0.25

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	Л	3	0.04	(0.01, 0.13)	0.07
es	С	62	0.93	(0.83, 0.98)	0.67
tse	Ц	2	0.03	(0.0, 0.1)	0.27

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	Л	8	0.12	(0.05, 0.22)	0.10
es	С	50	0.75	(0.63, 0.84)	0.52
ya	Я	9	0.13	(0.06, 0.24)	0.38

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	Л	35	0.52	(0.4, 0.65)	0.16
es	С	22	0.33	(0.22, 0.45)	0.57
yu	Ю	10	0.15	(0.07, 0.26)	0.27

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	Л	1	0.01	(0.0, 0.08)	0.11
ka	К	7	0.10	(0.04, 0.2)	0.17
o	О	59	0.88	(0.78, 0.95)	0.72

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	Л	8	0.12	(0.05, 0.22)	0.17
ka	К	43	0.64	(0.52, 0.76)	0.36
tse	Ц	16	0.24	(0.14, 0.36)	0.47

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	Л	27	0.40	(0.28, 0.53)	0.46
ka	К	25	0.37	(0.26, 0.5)	0.16
ya	Я	15	0.22	(0.13, 0.34)	0.38

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	Л	19	0.28	(0.18, 0.41)	0.24
ka	К	7	0.10	(0.04, 0.2)	0.32
yu	Ю	41	0.61	(0.49, 0.73)	0.44

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	Л	6	0.09	(0.03, 0.18)	0.00
o	О	59	0.88	(0.78, 0.95)	0.81
tse	Ц	2	0.03	(0.0, 0.1)	0.19

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	Л	15	0.22	(0.13, 0.34)	0.28
o	О	47	0.70	(0.58, 0.81)	0.52
ya	Я	5	0.07	(0.02, 0.17)	0.20

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	Л	47	0.70	(0.58, 0.81)	0.39
o	О	16	0.24	(0.14, 0.36)	0.57
yu	Ю	4	0.06	(0.02, 0.15)	0.05

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	Л	1	0.01	(0.0, 0.08)	0.09
tse	Ц	24	0.36	(0.24, 0.48)	0.37
ya	Я	42	0.63	(0.5, 0.74)	0.53

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	Л	5	0.07	(0.02, 0.17)	0.12
tse	Ц	6	0.09	(0.03, 0.18)	0.37
yu	Ю	56	0.84	(0.73, 0.92)	0.51

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	Л	24	0.36	(0.24, 0.48)	0.46
ya	Я	8	0.12	(0.05, 0.22)	0.35
yu	Ю	35	0.52	(0.4, 0.65)	0.19

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
es	С	-	-	-	0.11
ka	К	56	0.84	-	0.64
o	О	11	0.16	-	0.25

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
es	С	49	0.73	(0.61, 0.83)	0.50
ka	К	10	0.15	(0.07, 0.26)	0.17
tse	Ц	8	0.12	(0.05, 0.22)	0.33

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
es	С	50	0.75	(0.63, 0.84)	0.67
ka	К	5	0.07	(0.02, 0.17)	0.07
ya	Я	12	0.18	(0.1, 0.29)	0.25

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
es	С	31	0.46	(0.34, 0.59)	0.54
ka	К	24	0.36	(0.24, 0.48)	0.21
yu	Ю	12	0.18	(0.1, 0.29)	0.25

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
es	С	2	0.03	(0.0, 0.1)	0.05
o	О	12	0.18	(0.1, 0.29)	0.19
tse	Ц	53	0.79	(0.67, 0.88)	0.77

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
es	С	4	0.06	(0.02, 0.15)	0.26
o	О	7	0.10	(0.04, 0.2)	0.07
ya	Я	56	0.84	(0.73, 0.92)	0.68

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
es	С	12	0.18	(0.1, 0.29)	0.32
o	О	7	0.10	(0.04, 0.2)	0.10
yu	Ю	48	0.72	(0.59, 0.82)	0.58

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
es	С	36	0.54	(0.41, 0.66)	0.42
tse	Ц	17	0.25	(0.16, 0.37)	0.26
ya	Я	14	0.21	(0.12, 0.33)	0.32

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
es	С	14	0.21	(0.12, 0.33)	0.44
tse	Ц	40	0.60	(0.47, 0.72)	0.35
yu	Ю	13	0.19	(0.11, 0.31)	0.21

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
es	С	28	0.42	(0.3, 0.54)	0.69
ya	Я	34	0.51	(0.38, 0.63)	0.24
yu	Ю	5	0.07	(0.02, 0.17)	0.07

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ka	К	8	0.12	(0.05, 0.22)	0.12
o	О	57	0.85	(0.74, 0.93)	0.62
tse	Ц	2	0.03	(0.0, 0.1)	0.26

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ka	К	14	0.21	(0.12, 0.33)	0.22
o	О	51	0.76	(0.64, 0.86)	0.69
ya	Я	2	0.03	(0.0, 0.1)	0.10

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ka	К	41	0.61	(0.49, 0.73)	0.40
o	О	23	0.34	(0.23, 0.47)	0.53
yu	Ю	3	0.04	(0.01, 0.13)	0.07

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ka	К	9	0.13	(0.06, 0.24)	0.08
tse	Ц	36	0.54	(0.41, 0.66)	0.63
ya	Я	22	0.33	(0.22, 0.45)	0.29

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ka	К	8	0.12	(0.05, 0.22)	0.24
tse	Ц	26	0.39	(0.27, 0.51)	0.41
yu	Ю	33	0.49	(0.37, 0.62)	0.35

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ka	К	19	0.28	(0.18, 0.41)	0.36
ya	Я	22	0.33	(0.22, 0.45)	0.22
yu	Ю	26	0.39	(0.27, 0.51)	0.43

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
o	О	44	0.66	(0.53, 0.77)	0.42
tse	Ц	18	0.27	(0.17, 0.39)	0.44
ya	Я	5	0.07	(0.02, 0.17)	0.15

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
o	О	17	0.25	(0.16, 0.37)	0.42
tse	Ц	45	0.67	(0.55, 0.78)	0.58
yu	Ю	5	0.07	(0.02, 0.17)	0.00

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
o	О	23	0.34	(0.23, 0.47)	0.59
ya	Я	41	0.61	(0.49, 0.73)	0.31
yu	Ю	3	0.04	(0.01, 0.13)	0.10

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
tse	Ц	37	0.55	(0.43, 0.67)	0.62
ya	Я	15	0.22	(0.13, 0.34)	0.27
yu	Ю	15	0.22	(0.13, 0.34)	0.11

Courier New (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	12	0.16	(0.09, 0.26)	0.33
che	ч	53	0.71	(0.59, 0.81)	0.57
ef	ф	10	0.13	(0.07, 0.23)	0.10

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	50	0.67	(0.55, 0.77)	0.68
che	ч	4	0.05	(0.01, 0.13)	0.20
em	м	21	0.28	(0.18, 0.4)	0.12

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	61	0.81	(0.71, 0.89)	0.82
che	ч	4	0.05	(0.01, 0.13)	0.05
en	н	10	0.13	(0.07, 0.23)	0.14

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	58	0.77	(0.66, 0.86)	0.68
che	ч	1	0.01	(0.0, 0.07)	0.21
i	и	16	0.21	(0.13, 0.32)	0.11

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	30	0.40	(0.29, 0.52)	0.25
che	ч	40	0.53	(0.41, 0.65)	0.68
ie	е	5	0.07	(0.02, 0.15)	0.07

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	38	0.51	(0.39, 0.62)	0.44
che	ч	30	0.40	(0.29, 0.52)	0.35
ze	з	7	0.09	(0.04, 0.18)	0.21

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	10	0.13	(0.07, 0.23)	0.36
ef	ф	11	0.15	(0.08, 0.25)	0.08
em	м	54	0.72	(0.6, 0.82)	0.56

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	10	0.13	(0.07, 0.23)	0.29
ef	ф	13	0.17	(0.1, 0.28)	0.07
en	н	52	0.69	(0.58, 0.79)	0.64

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	4	0.05	(0.01, 0.13)	0.31
ef	ф	11	0.15	(0.08, 0.25)	0.11
i	и	60	0.80	(0.69, 0.88)	0.58

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	6	0.08	(0.03, 0.17)	0.19
ef	ф	33	0.44	(0.33, 0.56)	0.45
ie	е	36	0.48	(0.36, 0.6)	0.37

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	11	0.15	(0.08, 0.25)	0.09
ef	ф	20	0.27	(0.17, 0.38)	0.32
ze	з	44	0.59	(0.47, 0.7)	0.59

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	74	0.99	-	0.81
em	м	1	0.01	-	0.01
en	н	-	-	-	0.18

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	66	0.88	(0.78, 0.94)	0.98
em	м	3	0.04	(0.01, 0.11)	0.02
i	и	6	0.08	(0.03, 0.17)	0.01

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	23	0.31	(0.21, 0.42)	0.01
em	м	47	0.63	(0.51, 0.74)	0.82
ie	е	5	0.07	(0.02, 0.15)	0.16

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	26	0.35	(0.24, 0.47)	0.21
em	м	46	0.61	(0.49, 0.72)	0.47
ze	з	3	0.04	(0.01, 0.11)	0.32

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	71	0.95	-	0.82
en	н	-	-	-	0.18
i	и	4	0.05	-	0.00

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	25	0.33	(0.23, 0.45)	0.28
en	н	44	0.59	(0.47, 0.7)	0.72
ie	е	6	0.08	(0.03, 0.17)	0.00

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	30	0.40	(0.29, 0.52)	0.45
en	н	41	0.55	(0.43, 0.66)	0.42
ze	з	4	0.05	(0.01, 0.13)	0.13

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	34	0.45	(0.34, 0.57)	0.01
i	и	39	0.52	(0.4, 0.64)	0.81
ie	е	2	0.03	(0.0, 0.09)	0.17

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	25	0.33	(0.23, 0.45)	0.21
i	и	48	0.64	(0.52, 0.75)	0.46
ze	з	2	0.03	(0.0, 0.09)	0.33

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	34	0.45	(0.34, 0.57)	0.34
ie	е	14	0.19	(0.11, 0.29)	0.08
ze	з	27	0.36	(0.25, 0.48)	0.58

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	ч	5	0.07	(0.02, 0.15)	0.31
ef	ф	56	0.75	(0.63, 0.84)	0.45
em	м	14	0.19	(0.11, 0.29)	0.24

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	ч	-	-	-	0.15
ef	ф	64	0.85	-	0.62
en	н	11	0.15	-	0.23

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	ч	1	0.01	(0.0, 0.07)	0.29
ef	ф	52	0.69	(0.58, 0.79)	0.47
i	и	22	0.29	(0.19, 0.41)	0.24

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	ч	27	0.36	(0.25, 0.48)	0.38
ef	ф	36	0.48	(0.36, 0.6)	0.30
ie	е	12	0.16	(0.09, 0.26)	0.32

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	ч	26	0.35	(0.24, 0.47)	0.14
ef	ф	37	0.49	(0.38, 0.61)	0.44
ze	з	12	0.16	(0.09, 0.26)	0.42

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	ч	48	0.64	-	0.39
em	м	27	0.36	-	0.43
en	н	-	-	-	0.17

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	58	0.77	(0.66, 0.86)	0.66
em	М	14	0.19	(0.11, 0.29)	0.19
i	И	3	0.04	(0.01, 0.11)	0.15

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	9	0.12	(0.06, 0.22)	0.04
em	М	24	0.32	(0.22, 0.44)	0.38
ie	е	42	0.56	(0.44, 0.67)	0.58

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	4	0.05	(0.01, 0.13)	0.05
em	М	25	0.33	(0.23, 0.45)	0.40
ze	З	46	0.61	(0.49, 0.72)	0.56

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	66	0.88	(0.78, 0.94)	0.42
en	Н	2	0.03	(0.0, 0.09)	0.18
i	И	7	0.09	(0.04, 0.18)	0.40

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	7	0.09	(0.04, 0.18)	0.19
en	Н	12	0.16	(0.09, 0.26)	0.20
ie	е	56	0.75	(0.63, 0.84)	0.61

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	4	0.05	(0.01, 0.13)	0.16
en	Н	16	0.21	(0.13, 0.32)	0.23
ze	З	55	0.73	(0.62, 0.83)	0.61

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	12	0.16	(0.09, 0.26)	0.04
i	И	24	0.32	(0.22, 0.44)	0.37
ie	е	39	0.52	(0.4, 0.64)	0.59

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	6	0.08	(0.03, 0.17)	0.05
i	И	21	0.28	(0.18, 0.4)	0.39
ze	З	48	0.64	(0.52, 0.75)	0.57

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	48	0.64	(0.52, 0.75)	0.48
ie	е	11	0.15	(0.08, 0.25)	0.28
ze	З	16	0.21	(0.13, 0.32)	0.24

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	74	0.99	–	0.59
em	М	1	0.01	–	0.13
en	Н	–	–	–	0.28

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	71	0.95	–	0.73
em	М	–	–	–	0.13
i	И	4	0.05	–	0.14

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	22	0.29	(0.19, 0.41)	0.07
em	М	40	0.53	(0.41, 0.65)	0.47
ie	е	13	0.17	(0.1, 0.28)	0.45

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	23	0.31	(0.21, 0.42)	0.22
em	М	40	0.53	(0.41, 0.65)	0.23
ze	З	12	0.16	(0.09, 0.26)	0.55

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	69	0.92	(0.83, 0.97)	0.63
en	Н	1	0.01	(0.0, 0.07)	0.25
i	И	5	0.07	(0.02, 0.15)	0.12

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	33	0.44	(0.33, 0.56)	0.32
en	Н	33	0.44	(0.33, 0.56)	0.43
ie	е	9	0.12	(0.06, 0.22)	0.25

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	32	0.43	(0.31, 0.55)	0.43
en	н	36	0.48	(0.36, 0.6)	0.21
ze	з	7	0.09	(0.04, 0.18)	0.36

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	37	0.49	(0.38, 0.61)	0.08
i	и	34	0.45	(0.34, 0.57)	0.50
ie	е	4	0.05	(0.01, 0.13)	0.42

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	31	0.41	(0.3, 0.53)	0.24
i	и	36	0.48	(0.36, 0.6)	0.24
ze	з	8	0.11	(0.05, 0.2)	0.51

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ef	ф	43	0.57	(0.45, 0.69)	0.53
ie	е	12	0.16	(0.09, 0.26)	0.07
ze	з	20	0.27	(0.17, 0.38)	0.40

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	52	0.69	(0.58, 0.79)	0.29
en	н	14	0.19	(0.11, 0.29)	0.51
i	и	9	0.12	(0.06, 0.22)	0.20

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	4	0.05	–	0.31
en	н	–	–	–	0.00
ie	е	71	0.95	–	0.69

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	7	0.09	(0.04, 0.18)	0.30
en	н	2	0.03	(0.0, 0.09)	0.00
ze	з	66	0.88	(0.78, 0.94)	0.70

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	9	0.12	(0.06, 0.22)	0.01
i	и	1	0.01	(0.0, 0.07)	0.00
ie	е	65	0.87	(0.77, 0.93)	0.99

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	6	0.08	(0.03, 0.17)	0.01
i	и	4	0.05	(0.01, 0.13)	0.00
ze	з	65	0.87	(0.77, 0.93)	0.99

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
em	м	59	0.79	(0.68, 0.87)	0.71
ie	е	9	0.12	(0.06, 0.22)	0.17
ze	з	7	0.09	(0.04, 0.18)	0.13

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
en	н	8	0.11	(0.05, 0.2)	0.00
i	и	2	0.03	(0.0, 0.09)	0.29
ie	е	65	0.87	(0.77, 0.93)	0.71

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
en	н	2	0.03	(0.0, 0.09)	0.00
i	и	3	0.04	(0.01, 0.11)	0.27
ze	з	70	0.93	(0.85, 0.98)	0.73

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
en	н	58	0.77	(0.66, 0.86)	0.54
ie	е	10	0.13	(0.07, 0.23)	0.24
ze	з	7	0.09	(0.04, 0.18)	0.22

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
i	и	53	0.71	(0.59, 0.81)	0.70
ie	е	7	0.09	(0.04, 0.18)	0.17
ze	з	15	0.20	(0.12, 0.31)	0.13

Georgia (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	41	0.66	(0.53, 0.78)	0.54
ghe	г	12	0.19	(0.1, 0.31)	0.21
ha	х	9	0.15	(0.07, 0.26)	0.25

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	3	0.05	(0.01, 0.13)	0.05
ghe	Г	57	0.92	(0.82, 0.97)	0.91
ie	е	2	0.03	(0.0, 0.11)	0.04

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	6	0.10	(0.04, 0.2)	0.06
ghe	Г	48	0.77	(0.65, 0.87)	0.82
o	О	8	0.13	(0.06, 0.24)	0.12

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	58	0.94	(0.84, 0.98)	0.88
ghe	Г	3	0.05	(0.01, 0.13)	0.03
pe	П	1	0.02	(0.0, 0.09)	0.09

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	35	0.56	(0.43, 0.69)	0.73
ghe	Г	9	0.15	(0.07, 0.26)	0.02
shcha	Щ	18	0.29	(0.18, 0.42)	0.25

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	58	0.94	(0.84, 0.98)	0.90
ghe	Г	1	0.02	(0.0, 0.09)	0.02
te	Т	3	0.05	(0.01, 0.13)	0.08

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	2	0.03	(0.0, 0.11)	0.05
ha	Х	56	0.90	(0.8, 0.96)	0.91
ie	е	4	0.06	(0.02, 0.16)	0.04

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	5	0.08	(0.03, 0.18)	0.14
ha	Х	52	0.84	(0.72, 0.92)	0.78
o	О	5	0.08	(0.03, 0.18)	0.08

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	43	0.69	(0.56, 0.8)	0.62
ha	Х	7	0.11	(0.05, 0.22)	0.18
pe	П	12	0.19	(0.1, 0.31)	0.20

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	33	0.53	(0.4, 0.66)	0.45
ha	Х	6	0.10	(0.04, 0.2)	0.18
shcha	Щ	23	0.37	(0.25, 0.5)	0.37

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	45	0.73	(0.6, 0.83)	0.63
ha	Х	8	0.13	(0.06, 0.24)	0.17
te	Т	9	0.15	(0.07, 0.26)	0.20

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	14	0.23	(0.13, 0.35)	0.33
ie	е	3	0.05	(0.01, 0.13)	0.25
o	О	45	0.73	(0.6, 0.83)	0.41

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	3	0.05	–	0.05
ie	е	–	–	–	0.04
pe	П	59	0.95	–	0.91

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	3	0.05	(0.01, 0.13)	0.04
ie	е	4	0.06	(0.02, 0.16)	0.03
shcha	Щ	55	0.89	(0.78, 0.95)	0.93

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	4	0.06	(0.02, 0.16)	0.05
ie	е	2	0.03	(0.0, 0.11)	0.03
te	Т	56	0.90	(0.8, 0.96)	0.91

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	2	0.03	(0.0, 0.11)	0.14
o	О	8	0.13	(0.06, 0.24)	0.07
pe	П	52	0.84	(0.72, 0.92)	0.79

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	2	0.03	(0.0, 0.11)	0.05
o	О	9	0.15	(0.07, 0.26)	0.08
shcha	Щ	51	0.82	(0.7, 0.91)	0.87

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	6	0.10	(0.04, 0.2)	0.14
o	О	5	0.08	(0.03, 0.18)	0.06
te	Т	51	0.82	(0.7, 0.91)	0.80

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	41	0.66	(0.53, 0.78)	0.72
pe	П	6	0.10	(0.04, 0.2)	0.06
shcha	Щ	15	0.24	(0.14, 0.37)	0.22

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	58	0.94	–	0.96
pe	П	–	–	–	0.02
te	Т	4	0.06	–	0.02

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	Э	33	0.53	(0.4, 0.66)	0.70
shcha	Щ	18	0.29	(0.18, 0.42)	0.23
te	Т	11	0.18	(0.09, 0.3)	0.07

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	7	0.11	(0.05, 0.22)	0.21
ha	Х	14	0.23	(0.13, 0.35)	0.26
ie	е	41	0.66	(0.53, 0.78)	0.53

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	6	0.10	(0.04, 0.2)	0.30
ha	Х	12	0.19	(0.1, 0.31)	0.16
o	О	44	0.71	(0.58, 0.82)	0.54

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	8	0.13	(0.06, 0.24)	0.22
ha	Х	49	0.79	(0.67, 0.88)	0.78
pe	П	5	0.08	(0.03, 0.18)	0.00

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	10	0.16	(0.08, 0.28)	0.00
ha	Х	22	0.35	(0.24, 0.49)	0.63
shcha	Щ	30	0.48	(0.35, 0.61)	0.37

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	3	0.05	(0.01, 0.13)	0.19
ha	Х	58	0.94	(0.84, 0.98)	0.81
te	Т	1	0.02	(0.0, 0.09)	0.00

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	56	0.90	(0.8, 0.96)	0.87
ie	е	2	0.03	(0.0, 0.11)	0.03
o	О	4	0.06	(0.02, 0.16)	0.10

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	2	0.03	(0.0, 0.11)	0.04
ie	е	59	0.95	(0.87, 0.99)	0.87
pe	П	1	0.02	(0.0, 0.09)	0.09

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	1	0.02	(0.0, 0.09)	0.03
ie	е	37	0.60	(0.46, 0.72)	0.72
shcha	Щ	24	0.39	(0.27, 0.52)	0.25

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	1	0.02	(0.0, 0.09)	0.03
ie	е	58	0.94	(0.84, 0.98)	0.88
te	Т	3	0.05	(0.01, 0.13)	0.08

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	4	0.06	(0.02, 0.16)	0.13
o	О	56	0.90	(0.8, 0.96)	0.87
pe	П	2	0.03	(0.0, 0.11)	0.00

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	2	0.03	(0.0, 0.11)	0.00
o	О	37	0.60	(0.46, 0.72)	0.77
shcha	Щ	23	0.37	(0.25, 0.5)	0.23

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	2	0.03	(0.0, 0.11)	0.12
o	О	57	0.92	(0.82, 0.97)	0.88
te	Т	3	0.05	(0.01, 0.13)	0.00

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	8	0.13	(0.06, 0.24)	0.21
pe	П	2	0.03	(0.0, 0.11)	0.23
shcha	Щ	52	0.84	(0.72, 0.92)	0.56

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	10	0.16	(0.08, 0.28)	0.54
pe	П	41	0.66	(0.53, 0.78)	0.23
te	Т	11	0.18	(0.09, 0.3)	0.23

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ghe	Г	2	0.03	(0.0, 0.11)	0.16
shcha	Щ	58	0.94	(0.84, 0.98)	0.59
te	Т	2	0.03	(0.0, 0.11)	0.25

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	57	0.92	(0.82, 0.97)	0.83
ie	е	2	0.03	(0.0, 0.11)	0.12
o	О	3	0.05	(0.01, 0.13)	0.05

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	10	0.16	(0.08, 0.28)	0.20
ie	е	43	0.69	(0.56, 0.8)	0.60
pe	П	9	0.15	(0.07, 0.26)	0.20

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	3	0.05	(0.01, 0.13)	0.20
ie	е	33	0.53	(0.4, 0.66)	0.44
shcha	Щ	26	0.42	(0.3, 0.55)	0.37

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	9	0.15	(0.07, 0.26)	0.19
ie	е	47	0.76	(0.63, 0.86)	0.61
te	Т	6	0.10	(0.04, 0.2)	0.20

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	17	0.27	(0.17, 0.4)	0.20
o	О	40	0.65	(0.51, 0.76)	0.58
pe	П	5	0.08	(0.03, 0.18)	0.23

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	2	0.03	(0.0, 0.11)	0.13
o	О	38	0.61	(0.48, 0.73)	0.44
shcha	Щ	22	0.35	(0.24, 0.49)	0.44

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	14	0.23	(0.13, 0.35)	0.20
o	О	43	0.69	(0.56, 0.8)	0.58
te	Т	5	0.08	(0.03, 0.18)	0.23

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	27	0.44	(0.31, 0.57)	0.55
pe	П	7	0.11	(0.05, 0.22)	0.00
shcha	Щ	28	0.45	(0.32, 0.58)	0.45

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	56	0.90	(0.8, 0.96)	0.85
pe	П	3	0.05	(0.01, 0.13)	0.07
te	Т	3	0.05	(0.01, 0.13)	0.07

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	22	0.35	(0.24, 0.49)	0.51
shcha	Щ	29	0.47	(0.34, 0.6)	0.49
te	Т	11	0.18	(0.09, 0.3)	0.00

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ie	е	3	0.05	(0.01, 0.13)	0.11
o	о	3	0.05	(0.01, 0.13)	0.07
pe	п	56	0.90	(0.8, 0.96)	0.83

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ie	е	2	0.03	(0.0, 0.11)	0.02
o	о	2	0.03	(0.0, 0.11)	0.08
shcha	щ	58	0.94	(0.84, 0.98)	0.90

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ie	е	2	0.03	(0.0, 0.11)	0.11
o	о	2	0.03	(0.0, 0.11)	0.06
te	т	58	0.94	(0.84, 0.98)	0.83

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ie	е	39	0.63	(0.5, 0.75)	0.70
pe	п	4	0.06	(0.02, 0.16)	0.07
shcha	щ	19	0.31	(0.2, 0.44)	0.22

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ie	е	57	0.92	(0.82, 0.97)	0.94
pe	п	2	0.03	(0.0, 0.11)	0.03
te	т	3	0.05	(0.01, 0.13)	0.03

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ie	е	36	0.58	(0.45, 0.7)	0.68
shcha	щ	20	0.32	(0.21, 0.45)	0.24
te	т	6	0.10	(0.04, 0.2)	0.08

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
o	о	39	0.63	(0.5, 0.75)	0.70
pe	п	1	0.02	(0.0, 0.09)	0.00
shcha	щ	22	0.35	(0.24, 0.49)	0.30

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
o	о	57	0.92	(0.82, 0.97)	0.91
pe	п	2	0.03	(0.0, 0.11)	0.04
te	т	3	0.05	(0.01, 0.13)	0.04

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
o	о	39	0.63	(0.5, 0.75)	0.68
shcha	щ	20	0.32	(0.21, 0.45)	0.32
te	т	3	0.05	(0.01, 0.13)	0.00

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
pe	п	3	0.05	(0.01, 0.13)	0.12
shcha	щ	47	0.76	(0.63, 0.86)	0.69
te	т	12	0.19	(0.1, 0.31)	0.18

PT Sans (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	16	0.27	(0.16, 0.4)	0.42
el	л	33	0.55	(0.42, 0.68)	0.31
er	р	11	0.18	(0.1, 0.3)	0.26

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	40	0.67	(0.53, 0.78)	0.23
el	л	4	0.07	(0.02, 0.16)	0.31
ha	х	16	0.27	(0.16, 0.4)	0.46

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	45	0.75	(0.62, 0.85)	0.69
el	л	7	0.12	(0.05, 0.23)	0.06
te	т	8	0.13	(0.06, 0.25)	0.25

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	45	0.75	(0.62, 0.85)	0.46
el	л	6	0.10	(0.04, 0.21)	0.17
u	у	9	0.15	(0.07, 0.27)	0.37

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	17	0.28	(0.17, 0.41)	0.38
el	л	36	0.60	(0.47, 0.72)	0.38
ve	в	7	0.12	(0.05, 0.23)	0.24

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	4	0.07	(0.02, 0.16)	0.26
el	л	52	0.87	(0.75, 0.94)	0.58
ze	з	4	0.07	(0.02, 0.16)	0.16

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	12	0.20	(0.11, 0.32)	0.17
er	р	10	0.17	(0.08, 0.29)	0.31
ha	х	38	0.63	(0.5, 0.75)	0.52

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	17	0.28	(0.17, 0.41)	0.37
er	р	11	0.18	(0.1, 0.3)	0.17
te	т	32	0.53	(0.4, 0.66)	0.46

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	21	0.35	(0.23, 0.48)	0.41
er	р	6	0.10	(0.04, 0.21)	0.18
u	у	33	0.55	(0.42, 0.68)	0.42

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	20	0.33	(0.22, 0.47)	0.30
er	р	29	0.48	(0.35, 0.62)	0.39
ve	в	11	0.18	(0.1, 0.3)	0.32

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	3	0.05	(0.01, 0.14)	0.08
er	р	51	0.85	(0.73, 0.93)	0.60
ze	з	6	0.10	(0.04, 0.21)	0.32

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	44	0.73	(0.6, 0.84)	0.44
ha	х	9	0.15	(0.07, 0.27)	0.21
te	т	7	0.12	(0.05, 0.23)	0.35

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	45	0.75	(0.62, 0.85)	0.55
ha	х	6	0.10	(0.04, 0.21)	0.18
u	у	9	0.15	(0.07, 0.27)	0.27

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	12	0.20	(0.11, 0.32)	0.26
ha	х	46	0.77	(0.64, 0.87)	0.53
ve	в	2	0.03	(0.0, 0.12)	0.21

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	6	0.10	–	0.22
ha	х	54	0.90	–	0.69
ze	з	–	–	–	0.09

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	37	0.62	(0.48, 0.74)	0.47
te	т	8	0.13	(0.06, 0.25)	0.30
u	у	15	0.25	(0.15, 0.38)	0.23

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	14	0.23	(0.13, 0.36)	0.31
te	т	44	0.73	(0.6, 0.84)	0.53
ve	в	2	0.03	(0.0, 0.12)	0.16

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	10	0.17	(0.08, 0.29)	0.19
te	т	49	0.82	(0.7, 0.9)	0.73
ze	з	1	0.02	(0.0, 0.09)	0.08

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	3	0.05	(0.01, 0.14)	0.25
u	у	47	0.78	(0.66, 0.88)	0.54
ve	в	10	0.17	(0.08, 0.29)	0.20

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	2	0.03	(0.0, 0.12)	0.21
u	у	54	0.90	(0.79, 0.96)	0.73
ze	з	4	0.07	(0.02, 0.16)	0.07

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
e	э	27	0.45	(0.32, 0.58)	0.35
ve	в	31	0.52	(0.38, 0.65)	0.46
ze	з	2	0.03	(0.0, 0.12)	0.19

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	л	4	0.07	(0.02, 0.16)	0.17
er	р	34	0.57	(0.43, 0.69)	0.25
ha	х	22	0.37	(0.25, 0.5)	0.58

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	л	6	0.10	(0.04, 0.21)	0.13
er	р	43	0.72	(0.59, 0.83)	0.50
te	т	11	0.18	(0.1, 0.3)	0.36

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	л	20	0.33	(0.22, 0.47)	0.29
er	р	26	0.43	(0.31, 0.57)	0.33
u	у	14	0.23	(0.13, 0.36)	0.38

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	л	45	0.75	(0.62, 0.85)	0.26
er	р	8	0.13	(0.06, 0.25)	0.33
ve	в	7	0.12	(0.05, 0.23)	0.41

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	л	29	0.48	(0.35, 0.62)	0.10
er	р	17	0.28	(0.17, 0.41)	0.39
ze	з	14	0.23	(0.13, 0.36)	0.51

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	л	21	0.35	(0.23, 0.48)	0.26
ha	х	29	0.48	(0.35, 0.62)	0.69
te	т	10	0.17	(0.08, 0.29)	0.05

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	л	28	0.47	(0.34, 0.6)	0.44
ha	х	13	0.22	(0.12, 0.34)	0.50
u	у	19	0.32	(0.2, 0.45)	0.06

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	л	10	0.17	(0.08, 0.29)	0.29
ha	х	22	0.37	(0.25, 0.5)	0.56
ve	в	28	0.47	(0.34, 0.6)	0.15

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	л	10	0.17	(0.08, 0.29)	0.38
ha	х	20	0.33	(0.22, 0.47)	0.49
ze	з	30	0.50	(0.37, 0.63)	0.13

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	л	4	0.07	(0.02, 0.16)	0.08
te	т	21	0.35	(0.23, 0.48)	0.37
u	у	35	0.58	(0.45, 0.71)	0.56

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	л	6	0.10	(0.04, 0.21)	0.09
te	т	12	0.20	(0.11, 0.32)	0.33
ve	в	42	0.70	(0.57, 0.81)	0.58

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	л	5	0.08	(0.03, 0.18)	0.06
te	т	12	0.20	(0.11, 0.32)	0.26
ze	з	43	0.72	(0.59, 0.83)	0.68

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	л	1	0.02	(0.0, 0.09)	0.14
u	у	23	0.38	(0.26, 0.52)	0.46
ve	в	36	0.60	(0.47, 0.72)	0.39

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	л	5	0.08	(0.03, 0.18)	0.21
u	у	17	0.28	(0.17, 0.41)	0.38
ze	з	38	0.63	(0.5, 0.75)	0.41

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
el	Л	54	0.90	(0.79, 0.96)	0.46
ve	В	3	0.05	(0.01, 0.14)	0.24
ze	З	3	0.05	(0.01, 0.14)	0.30

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	44	0.73	(0.6, 0.84)	0.45
ha	х	13	0.22	(0.12, 0.34)	0.37
te	т	3	0.05	(0.01, 0.14)	0.18

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	36	0.60	(0.47, 0.72)	0.53
ha	х	17	0.28	(0.17, 0.41)	0.42
u	у	7	0.12	(0.05, 0.23)	0.04

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	12	0.20	(0.11, 0.32)	0.34
ha	х	46	0.77	(0.64, 0.87)	0.51
ve	В	2	0.03	(0.0, 0.12)	0.14

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	13	0.22	(0.12, 0.34)	0.49
ha	х	41	0.68	(0.55, 0.8)	0.30
ze	З	6	0.10	(0.04, 0.21)	0.22

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	20	0.33	(0.22, 0.47)	0.29
te	т	24	0.40	(0.28, 0.53)	0.42
u	у	16	0.27	(0.16, 0.4)	0.28

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	12	0.20	(0.11, 0.32)	0.24
te	т	45	0.75	(0.62, 0.85)	0.42
ve	В	3	0.05	(0.01, 0.14)	0.34

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	16	0.27	(0.16, 0.4)	0.29
te	т	29	0.48	(0.35, 0.62)	0.26
ze	З	15	0.25	(0.15, 0.38)	0.45

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	3	0.05	(0.01, 0.14)	0.20
u	у	44	0.73	(0.6, 0.84)	0.39
ve	В	13	0.22	(0.12, 0.34)	0.41

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	5	0.08	(0.03, 0.18)	0.36
u	у	36	0.60	(0.47, 0.72)	0.16
ze	З	19	0.32	(0.2, 0.45)	0.49

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
er	р	47	0.78	(0.66, 0.88)	0.57
ve	В	2	0.03	(0.0, 0.12)	0.09
ze	З	11	0.18	(0.1, 0.3)	0.34

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	х	5	0.08	(0.03, 0.18)	0.20
te	т	37	0.62	(0.48, 0.74)	0.49
u	у	18	0.30	(0.19, 0.43)	0.30

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	х	16	0.27	(0.16, 0.4)	0.33
te	т	14	0.23	(0.13, 0.36)	0.32
ve	В	30	0.50	(0.37, 0.63)	0.35

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	х	10	0.17	(0.08, 0.29)	0.20
te	т	9	0.15	(0.07, 0.27)	0.43
ze	З	41	0.68	(0.55, 0.8)	0.37

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	х	3	0.05	(0.01, 0.14)	0.20
u	у	19	0.32	(0.2, 0.45)	0.28
ve	В	38	0.63	(0.5, 0.75)	0.51

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	6	0.10	(0.04, 0.21)	0.20
u	У	15	0.25	(0.15, 0.38)	0.32
ze	З	39	0.65	(0.52, 0.77)	0.48

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	53	0.88	(0.77, 0.95)	0.61
ve	В	3	0.05	(0.01, 0.14)	0.23
ze	З	4	0.07	(0.02, 0.16)	0.15

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
te	Т	2	0.03	(0.0, 0.12)	0.27
u	У	29	0.48	(0.35, 0.62)	0.35
ve	В	29	0.48	(0.35, 0.62)	0.37

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
te	Т	4	0.07	(0.02, 0.16)	0.36
u	У	18	0.30	(0.19, 0.43)	0.23
ze	З	38	0.63	(0.5, 0.75)	0.40

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
te	Т	45	0.75	(0.62, 0.85)	0.63
ve	В	4	0.07	(0.02, 0.16)	0.15
ze	З	11	0.18	(0.1, 0.3)	0.22

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
u	У	53	0.88	(0.77, 0.95)	0.65
ve	В	4	0.07	(0.02, 0.16)	0.22
ze	З	3	0.05	(0.01, 0.14)	0.13

PT Serif (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	7	0.10	(0.04, 0.19)	0.11
be	б	26	0.36	(0.25, 0.48)	0.61
hard sign	ъ	40	0.55	(0.43, 0.66)	0.27

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	17	0.23	(0.14, 0.35)	0.51
be	б	18	0.25	(0.15, 0.36)	0.27
io	ё	38	0.52	(0.4, 0.64)	0.22

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	7	0.10	(0.04, 0.19)	0.18
be	б	10	0.14	(0.07, 0.24)	0.24
sha	ш	56	0.77	(0.65, 0.86)	0.58

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	4	0.05	(0.02, 0.13)	0.16
be	б	7	0.10	(0.04, 0.19)	0.22
shcha	щ	62	0.85	(0.75, 0.92)	0.61

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	10	0.14	(0.07, 0.24)	0.28
be	б	8	0.11	(0.05, 0.2)	0.33
short i	й	55	0.75	(0.64, 0.85)	0.40

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	8	0.11	(0.05, 0.2)	0.17
be	б	13	0.18	(0.1, 0.29)	0.25
tse	ц	52	0.71	(0.59, 0.81)	0.58

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	7	0.10	(0.04, 0.19)	0.26
hard sign	ъ	42	0.58	(0.45, 0.69)	0.25
io	ё	24	0.33	(0.22, 0.45)	0.49

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	25	0.34	(0.24, 0.46)	0.19
hard sign	ъ	7	0.10	(0.04, 0.19)	0.10
sha	ш	41	0.56	(0.44, 0.68)	0.72

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	31	0.42	(0.31, 0.55)	0.30
hard sign	ъ	7	0.10	(0.04, 0.19)	0.02
shcha	щ	35	0.48	(0.36, 0.6)	0.68

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	32	0.44	(0.32, 0.56)	0.16
hard sign	ъ	8	0.11	(0.05, 0.2)	0.23
short i	й	33	0.45	(0.34, 0.57)	0.61

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	37	0.51	(0.39, 0.63)	0.33
hard sign	ъ	11	0.15	(0.08, 0.25)	0.03
tse	ц	25	0.34	(0.24, 0.46)	0.64

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	5	0.07	(0.02, 0.15)	0.12
io	ё	12	0.16	(0.09, 0.27)	0.25
sha	ш	56	0.77	(0.65, 0.86)	0.62

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	7	0.10	(0.04, 0.19)	0.11
io	ё	13	0.18	(0.1, 0.29)	0.24
shcha	щ	53	0.73	(0.61, 0.82)	0.65

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	25	0.34	(0.24, 0.46)	0.21
io	ё	6	0.08	(0.03, 0.17)	0.33
short i	й	42	0.58	(0.45, 0.69)	0.46

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	7	0.10	(0.04, 0.19)	0.13
io	ё	11	0.15	(0.08, 0.25)	0.26
tse	ц	55	0.75	(0.64, 0.85)	0.61

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	70	0.96	–	0.88
sha	ш	–	–	–	0.02
shcha	щ	3	0.04	–	0.10

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	50	0.68	(0.57, 0.79)	0.54
sha	ш	7	0.10	(0.04, 0.19)	0.36
short i	й	16	0.22	(0.13, 0.33)	0.10

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	66	0.90	(0.81, 0.96)	0.80
sha	ш	1	0.01	(0.0, 0.07)	0.08
tse	ц	6	0.08	(0.03, 0.17)	0.12

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	53	0.73	(0.61, 0.82)	0.49
shcha	щ	11	0.15	(0.08, 0.25)	0.42
short i	й	9	0.12	(0.06, 0.22)	0.09

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	70	0.96	(0.88, 0.99)	0.91
shcha	щ	2	0.03	(0.0, 0.1)	0.07
tse	ц	1	0.01	(0.0, 0.07)	0.02

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	52	0.71	(0.59, 0.81)	0.50
short i	й	17	0.23	(0.14, 0.35)	0.10
tse	ц	4	0.05	(0.02, 0.13)	0.40

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	11	0.15	(0.08, 0.25)	0.33
hard sign	ъ	48	0.66	(0.54, 0.76)	0.55
io	ё	14	0.19	(0.11, 0.3)	0.12

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	38	0.52	(0.4, 0.64)	0.39
hard sign	ъ	10	0.14	(0.07, 0.24)	0.21
sha	ш	25	0.34	(0.24, 0.46)	0.40

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	41	0.56	(0.44, 0.68)	0.50
hard sign	ъ	7	0.10	(0.04, 0.19)	0.12
shcha	щ	25	0.34	(0.24, 0.46)	0.38

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	35	0.48	(0.36, 0.6)	0.34
hard sign	ъ	10	0.14	(0.07, 0.24)	0.35
short i	й	28	0.38	(0.27, 0.5)	0.31

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	47	0.64	(0.52, 0.75)	0.55
hard sign	ъ	10	0.14	(0.07, 0.24)	0.12
tse	ц	16	0.22	(0.13, 0.33)	0.33

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	7	0.10	(0.04, 0.19)	0.07
io	ё	15	0.21	(0.12, 0.32)	0.15
sha	ш	51	0.70	(0.58, 0.8)	0.78

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	14	0.19	(0.11, 0.3)	0.06
io	ё	9	0.12	(0.06, 0.22)	0.14
shcha	щ	50	0.68	(0.57, 0.79)	0.80

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	23	0.32	(0.21, 0.43)	0.13
io	ё	6	0.08	(0.03, 0.17)	0.22
short i	й	44	0.60	(0.48, 0.72)	0.65

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	5	0.07	(0.02, 0.15)	0.08
io	ё	14	0.19	(0.11, 0.3)	0.15
tse	ц	54	0.74	(0.62, 0.84)	0.77

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	70	0.96	(0.88, 0.99)	0.92
sha	ш	2	0.03	(0.0, 0.1)	0.00
shcha	щ	1	0.01	(0.0, 0.07)	0.08

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	51	0.70	(0.58, 0.8)	0.64
sha	ш	10	0.14	(0.07, 0.24)	0.30
short i	й	12	0.16	(0.09, 0.27)	0.06

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	71	0.97	–	0.85
sha	ш	–	–	–	0.05
tse	ц	2	0.03	–	0.09

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	59	0.81	(0.7, 0.89)	0.58
shcha	щ	9	0.12	(0.06, 0.22)	0.36
short i	й	5	0.07	(0.02, 0.15)	0.06

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	72	0.99	–	0.95
shcha	щ	–	–	–	0.05
tse	ц	1	0.01	–	0.00

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
be	б	56	0.77	(0.65, 0.86)	0.61
short i	й	10	0.14	(0.07, 0.24)	0.06
tse	ц	7	0.10	(0.04, 0.19)	0.33

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
hard sign	ъ	6	0.08	(0.03, 0.17)	0.10
io	ё	35	0.48	(0.36, 0.6)	0.33
sha	ш	32	0.44	(0.32, 0.56)	0.57

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
hard sign	ъ	6	0.08	(0.03, 0.17)	0.03
io	ё	50	0.68	(0.57, 0.79)	0.42
shcha	щ	17	0.23	(0.14, 0.35)	0.54

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
hard sign	ъ	32	0.44	(0.32, 0.56)	0.23
io	ё	30	0.41	(0.3, 0.53)	0.30
short i	й	11	0.15	(0.08, 0.25)	0.47

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
hard sign	Ъ	8	0.11	(0.05, 0.2)	0.04
io	ё	45	0.62	(0.5, 0.73)	0.45
tse	Ц	20	0.27	(0.18, 0.39)	0.51

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
hard sign	Ъ	68	0.93	(0.85, 0.98)	0.74
sha	Ш	3	0.04	(0.01, 0.12)	0.19
shcha	Щ	2	0.03	(0.0, 0.1)	0.07

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
hard sign	Ъ	45	0.62	(0.5, 0.73)	0.48
sha	Ш	8	0.11	(0.05, 0.2)	0.33
short i	Й	20	0.27	(0.18, 0.39)	0.19

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
hard sign	Ъ	62	0.85	(0.75, 0.92)	0.64
sha	Ш	10	0.14	(0.07, 0.24)	0.28
tse	Ц	1	0.01	(0.0, 0.07)	0.08

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
hard sign	Ъ	40	0.55	(0.43, 0.66)	0.38
shcha	Щ	8	0.11	(0.05, 0.2)	0.32
short i	Й	25	0.34	(0.24, 0.46)	0.30

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
hard sign	Ъ	66	0.90	(0.81, 0.96)	0.72
shcha	Щ	6	0.08	(0.03, 0.17)	0.17
tse	Ц	1	0.01	(0.0, 0.07)	0.11

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
hard sign	Ъ	47	0.64	(0.52, 0.75)	0.37
short i	Й	24	0.33	(0.22, 0.45)	0.34
tse	Ц	2	0.03	(0.0, 0.1)	0.29

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
io	ё	69	0.95	(0.87, 0.98)	0.90
sha	Ш	2	0.03	(0.0, 0.1)	0.01
shcha	Щ	2	0.03	(0.0, 0.1)	0.09

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
io	ё	50	0.68	(0.57, 0.79)	0.61
sha	Ш	20	0.27	(0.18, 0.39)	0.31
short i	Й	3	0.04	(0.01, 0.12)	0.07

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
io	ё	67	0.92	(0.83, 0.97)	0.83
sha	Ш	4	0.05	(0.02, 0.13)	0.07
tse	Ц	2	0.03	(0.0, 0.1)	0.10

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
io	ё	49	0.67	(0.55, 0.78)	0.57
shcha	Щ	17	0.23	(0.14, 0.35)	0.36
short i	Й	7	0.10	(0.04, 0.19)	0.06

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
io	ё	69	0.95	(0.87, 0.98)	0.92
shcha	Щ	2	0.03	(0.0, 0.1)	0.06
tse	Ц	2	0.03	(0.0, 0.1)	0.02

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
io	ё	48	0.66	(0.54, 0.76)	0.58
short i	Й	4	0.05	(0.02, 0.13)	0.08
tse	Ц	21	0.29	(0.19, 0.41)	0.33

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
sha	Ш	3	0.04	–	0.13
shcha	Щ	–	–	–	0.23
short i	Й	70	0.96	–	0.64

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
sha	Ш	36	0.49	(0.37, 0.61)	0.42
shcha	Щ	1	0.01	(0.0, 0.07)	0.18
tse	Ц	36	0.49	(0.37, 0.61)	0.40

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
sha	Ш	14	0.19	(0.11, 0.3)	0.20
short i	Й	53	0.73	(0.61, 0.82)	0.53
tse	Ц	6	0.08	(0.03, 0.17)	0.27

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
shcha	Щ	5	0.07	(0.02, 0.15)	0.18
short i	Й	64	0.88	(0.78, 0.94)	0.68
tse	Ц	4	0.05	(0.02, 0.13)	0.15

Times New Roman (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	13	0.22	(0.12, 0.34)	0.31
che	ч	3	0.05	(0.01, 0.14)	0.02
ha	х	44	0.73	(0.6, 0.84)	0.67

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	7	0.12	(0.05, 0.23)	0.14
che	ч	31	0.52	(0.38, 0.65)	0.51
hard sign	ъ	22	0.37	(0.25, 0.5)	0.35

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	19	0.32	(0.2, 0.45)	0.38
che	ч	15	0.25	(0.15, 0.38)	0.03
i	и	26	0.43	(0.31, 0.57)	0.59

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	7	0.12	(0.05, 0.23)	0.15
che	ч	33	0.55	(0.42, 0.68)	0.51
soft sign	ь	20	0.33	(0.22, 0.47)	0.34

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	8	0.13	(0.06, 0.25)	0.16
che	ч	41	0.68	(0.55, 0.8)	0.45
ve	в	11	0.18	(0.1, 0.3)	0.39

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	5	0.08	(0.03, 0.18)	0.14
che	ч	24	0.40	(0.28, 0.53)	0.49
yeru	ы	31	0.52	(0.38, 0.65)	0.37

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	5	0.08	(0.03, 0.18)	0.15
ha	х	42	0.70	(0.57, 0.81)	0.79
hard sign	ъ	13	0.22	(0.12, 0.34)	0.05

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	39	0.65	(0.52, 0.77)	0.76
ha	х	20	0.33	(0.22, 0.47)	0.17
i	и	1	0.02	(0.0, 0.09)	0.08

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	2	0.03	(0.0, 0.12)	0.16
ha	х	53	0.88	(0.77, 0.95)	0.79
soft sign	ь	5	0.08	(0.03, 0.18)	0.05

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	8	0.13	(0.06, 0.25)	0.17
ha	х	49	0.82	(0.7, 0.9)	0.77
ve	в	3	0.05	(0.01, 0.14)	0.06

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	2	0.03	(0.0, 0.12)	0.16
ha	х	47	0.78	(0.66, 0.88)	0.75
yeru	ы	11	0.18	(0.1, 0.3)	0.10

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	7	0.12	(0.05, 0.23)	0.20
hard sign	ъ	15	0.25	(0.15, 0.38)	0.07
i	и	38	0.63	(0.5, 0.75)	0.73

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	53	0.88	(0.77, 0.95)	0.49
hard sign	ъ	3	0.05	(0.01, 0.14)	0.26
soft sign	ь	4	0.07	(0.02, 0.16)	0.26

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	28	0.47	(0.34, 0.6)	0.51
hard sign	ъ	19	0.32	(0.2, 0.45)	0.26
ve	в	13	0.22	(0.12, 0.34)	0.23

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	32	0.53	(0.4, 0.66)	0.47
hard sign	ъ	13	0.22	(0.12, 0.34)	0.24
yeru	ы	15	0.25	(0.15, 0.38)	0.29

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	5	0.08	(0.03, 0.18)	0.20
i	и	39	0.65	(0.52, 0.77)	0.73
soft sign	ь	16	0.27	(0.16, 0.4)	0.07

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	8	0.13	(0.06, 0.25)	0.22
i	и	45	0.75	(0.62, 0.85)	0.70
ve	в	7	0.12	(0.05, 0.23)	0.08

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	11	0.18	(0.1, 0.3)	0.22
i	и	34	0.57	(0.43, 0.69)	0.68
yeru	ы	15	0.25	(0.15, 0.38)	0.10

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	31	0.52	(0.38, 0.65)	0.50
soft sign	ь	16	0.27	(0.16, 0.4)	0.27
ve	в	13	0.22	(0.12, 0.34)	0.24

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	42	0.70	(0.57, 0.81)	0.48
soft sign	ь	6	0.10	(0.04, 0.21)	0.24
yeru	ы	12	0.20	(0.11, 0.32)	0.28

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
a	а	26	0.43	(0.31, 0.57)	0.48
ve	в	9	0.15	(0.07, 0.27)	0.22
yeru	ы	25	0.42	(0.29, 0.55)	0.30

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	ч	4	0.07	(0.02, 0.16)	0.15
ha	х	32	0.53	(0.4, 0.66)	0.55
hard sign	ъ	24	0.40	(0.28, 0.53)	0.30

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	ч	31	0.52	(0.38, 0.65)	0.55
ha	х	20	0.33	(0.22, 0.47)	0.35
i	и	9	0.15	(0.07, 0.27)	0.10

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	ч	4	0.07	(0.02, 0.16)	0.15
ha	х	37	0.62	(0.48, 0.74)	0.56
soft sign	ь	19	0.32	(0.2, 0.45)	0.30

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	ч	11	0.18	(0.1, 0.3)	0.16
ha	х	38	0.63	(0.5, 0.75)	0.53
ve	в	11	0.18	(0.1, 0.3)	0.32

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	ч	5	0.08	(0.03, 0.18)	0.14
ha	х	34	0.57	(0.43, 0.69)	0.53
yeru	ы	21	0.35	(0.23, 0.48)	0.33

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	ч	5	0.08	(0.03, 0.18)	0.18
hard sign	ъ	27	0.45	(0.32, 0.58)	0.37
i	и	28	0.47	(0.34, 0.6)	0.44

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	ч	55	0.92	–	0.73
hard sign	ъ	5	0.08	–	0.14
soft sign	ь	–	–	–	0.13

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	39	0.65	(0.52, 0.77)	0.71
hard sign	Ъ	14	0.23	(0.13, 0.36)	0.15
ve	В	7	0.12	(0.05, 0.23)	0.15

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	43	0.72	(0.59, 0.83)	0.70
hard sign	Ъ	6	0.10	(0.04, 0.21)	0.14
yeru	Ы	11	0.18	(0.1, 0.3)	0.16

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	8	0.13	(0.06, 0.25)	0.19
i	И	31	0.52	(0.38, 0.65)	0.45
soft sign	Ь	21	0.35	(0.23, 0.48)	0.37

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	12	0.20	(0.11, 0.32)	0.20
i	И	32	0.53	(0.4, 0.66)	0.41
ve	В	16	0.27	(0.16, 0.4)	0.40

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	19	0.32	(0.2, 0.45)	0.20
i	И	19	0.32	(0.2, 0.45)	0.42
yeru	Ы	22	0.37	(0.25, 0.5)	0.38

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	40	0.67	(0.53, 0.78)	0.70
soft sign	Ь	12	0.20	(0.11, 0.32)	0.15
ve	В	8	0.13	(0.06, 0.25)	0.15

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	47	0.78	(0.66, 0.88)	0.71
soft sign	Ь	6	0.10	(0.04, 0.21)	0.14
yeru	Ы	7	0.12	(0.05, 0.23)	0.16

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
che	Ч	38	0.63	(0.5, 0.75)	0.67
ve	В	6	0.10	(0.04, 0.21)	0.15
yeru	Ы	16	0.27	(0.16, 0.4)	0.18

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	13	0.22	(0.12, 0.34)	0.24
hard sign	Ъ	43	0.72	(0.59, 0.83)	0.67
i	И	4	0.07	(0.02, 0.16)	0.08

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	54	0.90	(0.79, 0.96)	0.96
hard sign	Ъ	4	0.07	(0.02, 0.16)	0.02
soft sign	Ь	2	0.03	(0.0, 0.12)	0.02

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	42	0.70	(0.57, 0.81)	0.96
hard sign	Ъ	13	0.22	(0.12, 0.34)	0.02
ve	В	5	0.08	(0.03, 0.18)	0.02

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	48	0.80	(0.68, 0.89)	0.92
hard sign	Ъ	4	0.07	(0.02, 0.16)	0.02
yeru	Ы	8	0.13	(0.06, 0.25)	0.06

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	19	0.32	(0.2, 0.45)	0.25
i	И	4	0.07	(0.02, 0.16)	0.08
soft sign	Ь	37	0.62	(0.48, 0.74)	0.67

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	17	0.28	(0.17, 0.41)	0.24
i	И	7	0.12	(0.05, 0.23)	0.08
ve	В	36	0.60	(0.47, 0.72)	0.67

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	19	0.32	-	0.30
i	И	-	-	-	0.07
yeru	Ы	41	0.68	-	0.63

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	50	0.83	(0.71, 0.92)	0.96
soft sign	Ь	8	0.13	(0.06, 0.25)	0.02
ve	В	2	0.03	(0.0, 0.12)	0.02

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	48	0.80	(0.68, 0.89)	0.92
soft sign	Ь	4	0.07	(0.02, 0.16)	0.02
yeru	Ы	8	0.13	(0.06, 0.25)	0.06

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
ha	Х	47	0.78	(0.66, 0.88)	0.91
ve	В	1	0.02	(0.0, 0.09)	0.02
yeru	Ы	12	0.20	(0.11, 0.32)	0.07

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
hard sign	Ъ	6	0.10	(0.04, 0.21)	0.05
i	И	53	0.88	(0.77, 0.95)	0.90
soft sign	Ь	1	0.02	(0.0, 0.09)	0.05

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
hard sign	Ъ	12	0.20	(0.11, 0.32)	0.06
i	И	43	0.72	(0.59, 0.83)	0.88
ve	В	5	0.08	(0.03, 0.18)	0.06

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
hard sign	Ъ	8	0.13	(0.06, 0.25)	0.08
i	И	45	0.75	(0.62, 0.85)	0.85
yeru	Ы	7	0.12	(0.05, 0.23)	0.08

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
hard sign	Ъ	16	0.27	(0.16, 0.4)	0.33
soft sign	Ь	1	0.02	(0.0, 0.09)	0.34
ve	В	43	0.72	(0.59, 0.83)	0.33

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
hard sign	Ъ	17	0.28	(0.17, 0.41)	0.32
soft sign	Ь	1	0.02	(0.0, 0.09)	0.32
yeru	Ы	42	0.70	(0.57, 0.81)	0.36

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
hard sign	Ъ	7	0.12	(0.05, 0.23)	0.31
ve	В	37	0.62	(0.48, 0.74)	0.31
yeru	Ы	16	0.27	(0.16, 0.4)	0.38

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
i	И	45	0.75	(0.62, 0.85)	0.88
soft sign	Ь	11	0.18	(0.1, 0.3)	0.06
ve	В	4	0.07	(0.02, 0.16)	0.06

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
i	И	41	0.68	(0.55, 0.8)	0.85
soft sign	Ь	8	0.13	(0.06, 0.25)	0.08
yeru	Ы	11	0.18	(0.1, 0.3)	0.08

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
i	И	36	0.60	(0.47, 0.72)	0.81
ve	В	13	0.22	(0.12, 0.34)	0.10
yeru	Ы	11	0.18	(0.1, 0.3)	0.09

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
soft sign	Ь	2	0.03	(0.0, 0.12)	0.31
ve	В	40	0.67	(0.53, 0.78)	0.32
yeru	Ы	18	0.30	(0.19, 0.43)	0.37

Verdana (Cyrillic)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	27	0.54	(0.39, 0.68)	0.60
Ba	ब	20	0.40	(0.26, 0.55)	0.30
Gha	घ	3	0.06	(0.01, 0.17)	0.10

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	27	0.54	(0.39, 0.68)	0.50
Ba	ब	13	0.26	(0.15, 0.4)	0.23
Ka	ख	10	0.20	(0.1, 0.34)	0.27

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	19	0.38	(0.25, 0.53)	0.35
Ba	ब	14	0.28	(0.16, 0.42)	0.19
Ra	र	17	0.34	(0.21, 0.49)	0.46

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	22	0.44	(0.3, 0.59)	0.34
Ba	ब	25	0.50	(0.36, 0.64)	0.32
Sa	स	3	0.06	(0.01, 0.17)	0.33

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	45	0.90	(0.78, 0.97)	0.49
Ba	ब	2	0.04	(0.0, 0.14)	0.21
Ssa	ष	3	0.06	(0.01, 0.17)	0.30

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	33	0.66	(0.51, 0.79)	0.44
Ba	ब	9	0.18	(0.09, 0.31)	0.26
Ta	त	8	0.16	(0.07, 0.29)	0.31

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	24	0.48	(0.34, 0.63)	0.37
Gha	घ	8	0.16	(0.07, 0.29)	0.22
Ka	ख	18	0.36	(0.23, 0.51)	0.41

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	14	0.28	(0.16, 0.42)	0.26
Gha	घ	14	0.28	(0.16, 0.42)	0.15
Ra	र	22	0.44	(0.3, 0.59)	0.59

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	23	0.46	(0.32, 0.61)	0.26
Gha	घ	18	0.36	(0.23, 0.51)	0.28
Sa	स	9	0.18	(0.09, 0.31)	0.46

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	31	0.62	(0.47, 0.75)	0.32
Gha	घ	2	0.04	(0.0, 0.14)	0.20
Ssa	ष	17	0.34	(0.21, 0.49)	0.48

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	21	0.42	(0.28, 0.57)	0.34
Gha	घ	10	0.20	(0.1, 0.34)	0.21
Ta	त	19	0.38	(0.25, 0.53)	0.45

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	29	0.58	(0.43, 0.72)	0.56
Ka	ख	10	0.20	(0.1, 0.34)	0.14
Ra	र	11	0.22	(0.12, 0.36)	0.30

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	35	0.70	(0.55, 0.82)	0.55
Ka	ख	11	0.22	(0.12, 0.36)	0.24
Sa	स	4	0.08	(0.02, 0.19)	0.21

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	19	0.38	(0.25, 0.53)	0.28
Ka	ख	9	0.18	(0.09, 0.31)	0.35
Ssa	ष	22	0.44	(0.3, 0.59)	0.37

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	26	0.52	(0.37, 0.66)	0.24
Ka	ख	11	0.22	(0.12, 0.36)	0.38
Ta	त	13	0.26	(0.15, 0.4)	0.38

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	35	0.70	(0.55, 0.82)	0.71
Ra	र	13	0.26	(0.15, 0.4)	0.26
Sa	स	2	0.04	(0.0, 0.14)	0.03

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	22	0.44	(0.3, 0.59)	0.34
Ra	र	14	0.28	(0.16, 0.42)	0.41
Ssa	ष	14	0.28	(0.16, 0.42)	0.25

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	31	0.62	(0.47, 0.75)	0.34
Ra	र	11	0.22	(0.12, 0.36)	0.44
Ta	त	8	0.16	(0.07, 0.29)	0.22

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	24	0.48	(0.34, 0.63)	0.32
Sa	स	3	0.06	(0.01, 0.17)	0.30
Ssa	ष	23	0.46	(0.32, 0.61)	0.37

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	30	0.60	(0.45, 0.74)	0.34
Sa	स	2	0.04	(0.0, 0.14)	0.32
Ta	त	18	0.36	(0.23, 0.51)	0.35

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
A	अ	26	0.52	(0.37, 0.66)	0.41
Ssa	ष	12	0.24	(0.13, 0.38)	0.31
Ta	त	12	0.24	(0.13, 0.38)	0.27

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	11	0.22	(0.12, 0.36)	0.18
Gha	घ	20	0.40	(0.26, 0.55)	0.32
Ka	ख	19	0.38	(0.25, 0.53)	0.50

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	11	0.22	(0.12, 0.36)	0.08
Gha	घ	8	0.16	(0.07, 0.29)	0.13
Ra	र	31	0.62	(0.47, 0.75)	0.79

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	23	0.46	(0.32, 0.61)	0.14
Gha	घ	9	0.18	(0.09, 0.31)	0.19
Sa	स	18	0.36	(0.23, 0.51)	0.67

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	9	0.18	(0.09, 0.31)	0.10
Gha	घ	36	0.72	(0.58, 0.84)	0.27
Ssa	ष	5	0.10	(0.03, 0.22)	0.62

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	14	0.28	(0.16, 0.42)	0.16
Gha	घ	16	0.32	(0.2, 0.47)	0.21
Ta	त	20	0.40	(0.26, 0.55)	0.63

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	28	0.56	(0.41, 0.7)	0.45
Ka	ख	13	0.26	(0.15, 0.4)	0.11
Ra	र	9	0.18	(0.09, 0.31)	0.44

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	35	0.70	(0.55, 0.82)	0.48
Ka	ख	8	0.16	(0.07, 0.29)	0.15
Sa	स	7	0.14	(0.06, 0.27)	0.37

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	5	0.10	(0.03, 0.22)	0.10
Ka	ख	41	0.82	(0.69, 0.91)	0.40
Ssa	ष	4	0.08	(0.02, 0.19)	0.50

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	9	0.18	(0.09, 0.31)	0.08
Ka	ख	20	0.40	(0.26, 0.55)	0.39
Ta	त	21	0.42	(0.28, 0.57)	0.53

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	34	0.68	(0.53, 0.8)	0.70
Ra	र	11	0.22	(0.12, 0.36)	0.21
Sa	स	5	0.10	(0.03, 0.22)	0.09

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	3	0.06	(0.01, 0.17)	0.18
Ra	र	46	0.92	(0.81, 0.98)	0.56
Ssa	ष	1	0.02	(0.0, 0.11)	0.26

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	30	0.60	(0.45, 0.74)	0.22
Ra	र	17	0.34	(0.21, 0.49)	0.52
Ta	त	3	0.06	(0.01, 0.17)	0.26

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	8	0.16	(0.07, 0.29)	0.22
Sa	स	41	0.82	(0.69, 0.91)	0.47
Ssa	ष	1	0.02	(0.0, 0.11)	0.30

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	29	0.58	(0.43, 0.72)	0.27
Sa	स	12	0.24	(0.13, 0.38)	0.42
Ta	त	9	0.18	(0.09, 0.31)	0.31

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	5	0.10	(0.03, 0.22)	0.25
Ssa	ष	5	0.10	(0.03, 0.22)	0.34
Ta	त	40	0.80	(0.66, 0.9)	0.41

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Gha	घ	27	0.54	(0.39, 0.68)	0.50
Ka	ख	11	0.22	(0.12, 0.36)	0.13
Ra	र	12	0.24	(0.13, 0.38)	0.37

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Gha	घ	39	0.78	(0.64, 0.88)	0.52
Ka	ख	7	0.14	(0.06, 0.27)	0.17
Sa	स	4	0.08	(0.02, 0.19)	0.30

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Gha	घ	6	0.12	(0.05, 0.24)	0.19
Ka	ख	27	0.54	(0.39, 0.68)	0.36
Ssa	ष	17	0.34	(0.21, 0.49)	0.45

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Gha	घ	15	0.30	(0.18, 0.45)	0.15
Ka	ख	18	0.36	(0.23, 0.51)	0.40
Ta	त	17	0.34	(0.21, 0.49)	0.45

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Gha	घ	27	0.54	(0.39, 0.68)	0.71
Ra	र	15	0.30	(0.18, 0.45)	0.20
Sa	स	8	0.16	(0.07, 0.29)	0.08

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Gha	घ	3	0.06	(0.01, 0.17)	0.26
Ra	र	31	0.62	(0.47, 0.75)	0.46
Ssa	ष	16	0.32	(0.2, 0.47)	0.28

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Gha	घ	31	0.62	(0.47, 0.75)	0.26
Ra	र	14	0.28	(0.16, 0.42)	0.50
Ta	त	5	0.10	(0.03, 0.22)	0.24

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Gha	घ	15	0.30	(0.18, 0.45)	0.29
Sa	स	23	0.46	(0.32, 0.61)	0.39
Ssa	ष	12	0.24	(0.13, 0.38)	0.32

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Gha	घ	20	0.40	(0.26, 0.55)	0.31
Sa	स	14	0.28	(0.16, 0.42)	0.40
Ta	त	16	0.32	(0.2, 0.47)	0.29

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Gha	घ	12	0.24	(0.13, 0.38)	0.33
Ssa	ष	16	0.32	(0.2, 0.47)	0.36
Ta	त	22	0.44	(0.3, 0.59)	0.31

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	ख	18	0.36	(0.23, 0.51)	0.42
Ra	र	22	0.44	(0.3, 0.59)	0.35
Sa	स	10	0.20	(0.1, 0.34)	0.23

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	ख	7	0.14	(0.06, 0.27)	0.21
Ra	र	7	0.14	(0.06, 0.27)	0.20
Ssa	ष	36	0.72	(0.58, 0.84)	0.59

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	ख	21	0.42	(0.28, 0.57)	0.24
Ra	र	7	0.14	(0.06, 0.27)	0.17
Ta	त	22	0.44	(0.3, 0.59)	0.60

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	ख	14	0.28	(0.16, 0.42)	0.24
Sa	स	2	0.04	(0.0, 0.14)	0.16
Ssa	ष	34	0.68	(0.53, 0.8)	0.61

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	ख	12	0.24	(0.13, 0.38)	0.27
Sa	स	4	0.08	(0.02, 0.19)	0.12
Ta	त	34	0.68	(0.53, 0.8)	0.61

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	ख	23	0.46	(0.32, 0.61)	0.47
Ssa	ष	12	0.24	(0.13, 0.38)	0.24
Ta	त	15	0.30	(0.18, 0.45)	0.29

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ra	र	17	0.34	(0.21, 0.49)	0.18
Sa	स	4	0.08	(0.02, 0.19)	0.09
Ssa	ष	29	0.58	(0.43, 0.72)	0.73

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ra	र	12	0.24	(0.13, 0.38)	0.21
Sa	स	15	0.30	(0.18, 0.45)	0.09
Ta	त	23	0.46	(0.32, 0.61)	0.70

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ra	र	22	0.44	(0.3, 0.59)	0.47
Ssa	ष	23	0.46	(0.32, 0.61)	0.28
Ta	त	5	0.10	(0.03, 0.22)	0.25

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Sa	स	12	0.24	(0.13, 0.38)	0.39
Ssa	ष	24	0.48	(0.34, 0.63)	0.33
Ta	त	14	0.28	(0.16, 0.42)	0.28

Adobe Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	25	0.45	(0.32, 0.59)	0.07
Cha	छ	21	0.38	(0.25, 0.52)	0.40
Dda	ड	9	0.16	(0.08, 0.29)	0.53

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	39	0.71	(0.57, 0.82)	0.43
Cha	छ	10	0.18	(0.09, 0.31)	0.41
Dha	ध	6	0.11	(0.04, 0.22)	0.17

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	37	0.67	(0.53, 0.79)	0.42
Cha	छ	13	0.24	(0.13, 0.37)	0.37
Gha	घ	5	0.09	(0.03, 0.2)	0.21

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	28	0.51	(0.37, 0.65)	0.08
Cha	छ	15	0.27	(0.16, 0.41)	0.42
Ha	ह	12	0.22	(0.12, 0.35)	0.51

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	21	0.38	(0.25, 0.52)	0.06
Cha	छ	17	0.31	(0.19, 0.45)	0.37
Tta	ट	17	0.31	(0.19, 0.45)	0.56

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	35	0.64	(0.5, 0.76)	0.32
Cha	छ	13	0.24	(0.13, 0.37)	0.20
U	उ	7	0.13	(0.05, 0.24)	0.48

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	20	0.36	(0.24, 0.5)	0.08
Dda	ड	17	0.31	(0.19, 0.45)	0.65
Dha	ध	18	0.33	(0.21, 0.47)	0.26

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	13	0.24	(0.13, 0.37)	0.12
Dda	ड	32	0.58	(0.44, 0.71)	0.60
Gha	घ	10	0.18	(0.09, 0.31)	0.28

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	45	0.82	(0.69, 0.91)	0.47
Dda	ड	2	0.04	(0.0, 0.13)	0.29
Ha	ह	8	0.15	(0.06, 0.27)	0.24

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	42	0.76	(0.63, 0.87)	0.43
Dda	ड	5	0.09	(0.03, 0.2)	0.27
Tta	ट	8	0.15	(0.06, 0.27)	0.31

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	40	0.73	(0.59, 0.84)	0.46
Dda	ड	3	0.05	(0.01, 0.15)	0.19
U	उ	12	0.22	(0.12, 0.35)	0.35

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	35	0.64	(0.5, 0.76)	0.33
Dha	ध	13	0.24	(0.13, 0.37)	0.29
Gha	घ	7	0.13	(0.05, 0.24)	0.37

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	20	0.36	(0.24, 0.5)	0.09
Dha	ध	15	0.27	(0.16, 0.41)	0.28
Ha	ह	20	0.36	(0.24, 0.5)	0.63

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	17	0.31	(0.19, 0.45)	0.08
Dha	ध	19	0.35	(0.22, 0.49)	0.23
Tta	ट	19	0.35	(0.22, 0.49)	0.69

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	22	0.40	(0.27, 0.54)	0.17
Dha	ध	9	0.16	(0.08, 0.29)	0.12
U	उ	24	0.44	(0.3, 0.58)	0.71

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	21	0.38	(0.25, 0.52)	0.12
Gha	घ	7	0.13	(0.05, 0.24)	0.30
Ha	ह	27	0.49	(0.35, 0.63)	0.58

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	11	0.20	(0.1, 0.33)	0.13
Gha	घ	9	0.16	(0.08, 0.29)	0.25
Tta	ट	35	0.64	(0.5, 0.76)	0.62

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	18	0.33	(0.21, 0.47)	0.39
Gha	घ	3	0.05	(0.01, 0.15)	0.05
U	उ	34	0.62	(0.48, 0.75)	0.56

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	33	0.60	(0.46, 0.73)	0.45
Ha	ह	12	0.22	(0.12, 0.35)	0.23
Tta	ट	10	0.18	(0.09, 0.31)	0.32

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	33	0.60	(0.46, 0.73)	0.28
Ha	ह	6	0.11	(0.04, 0.22)	0.24
U	उ	16	0.29	(0.18, 0.43)	0.48

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	39	0.71	(0.57, 0.82)	0.26
Tta	ट	7	0.13	(0.05, 0.24)	0.31
U	उ	9	0.16	(0.08, 0.29)	0.42

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	6	0.11	(0.04, 0.22)	0.22
Dda	ड	42	0.76	(0.63, 0.87)	0.71
Dha	ध	7	0.13	(0.05, 0.24)	0.07

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	11	0.20	(0.1, 0.33)	0.24
Dda	ड	38	0.69	(0.55, 0.81)	0.69
Gha	घ	6	0.11	(0.04, 0.22)	0.07

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	44	0.80	(0.67, 0.9)	0.66
Dda	ड	7	0.13	(0.05, 0.24)	0.19
Ha	ह	4	0.07	(0.02, 0.18)	0.15

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	47	0.85	(0.73, 0.94)	0.62
Dda	ड	1	0.02	(0.0, 0.1)	0.18
Tta	ट	7	0.13	(0.05, 0.24)	0.20

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	29	0.53	(0.39, 0.66)	0.47
Dda	ड	13	0.24	(0.13, 0.37)	0.38
U	उ	13	0.24	(0.13, 0.37)	0.15

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	39	0.71	(0.57, 0.82)	0.30
Dha	ध	8	0.15	(0.06, 0.27)	0.33
Gha	घ	8	0.15	(0.06, 0.27)	0.37

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	4	0.07	(0.02, 0.18)	0.23
Dha	ध	12	0.22	(0.12, 0.35)	0.07
Ha	ह	39	0.71	(0.57, 0.82)	0.70

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	7	0.13	(0.05, 0.24)	0.22
Dha	ध	6	0.11	(0.04, 0.22)	0.06
Tta	ट	42	0.76	(0.63, 0.87)	0.73

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	10	0.18	(0.09, 0.31)	0.15
Dha	ध	7	0.13	(0.05, 0.24)	0.23
U	उ	38	0.69	(0.55, 0.81)	0.63

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	10	0.18	(0.09, 0.31)	0.25
Gha	घ	9	0.16	(0.08, 0.29)	0.07
Ha	ह	36	0.65	(0.51, 0.78)	0.68

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	9	0.16	(0.08, 0.29)	0.24
Gha	घ	8	0.15	(0.06, 0.27)	0.06
Tta	ट	38	0.69	(0.55, 0.81)	0.70

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	17	0.31	(0.19, 0.45)	0.34
Gha	घ	6	0.11	(0.04, 0.22)	0.13
U	उ	32	0.58	(0.44, 0.71)	0.54

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	29	0.53	(0.39, 0.66)	0.64
Ha	ह	2	0.04	(0.0, 0.13)	0.15
Tta	ट	24	0.44	(0.3, 0.58)	0.21

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	20	0.36	(0.24, 0.5)	0.34
Ha	ह	17	0.31	(0.19, 0.45)	0.42
U	उ	18	0.33	(0.21, 0.47)	0.24

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	24	0.44	(0.3, 0.58)	0.32
Tta	ट	21	0.38	(0.25, 0.52)	0.47
U	उ	10	0.18	(0.09, 0.31)	0.21

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	47	0.85	(0.73, 0.94)	0.67
Dha	ध	7	0.13	(0.05, 0.24)	0.18
Gha	घ	1	0.02	(0.0, 0.1)	0.16

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	8	0.15	(0.06, 0.27)	0.22
Dha	ध	38	0.69	(0.55, 0.81)	0.61
Ha	ह	9	0.16	(0.08, 0.29)	0.17

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	4	0.07	(0.02, 0.18)	0.21
Dha	ध	40	0.73	(0.59, 0.84)	0.55
Tta	ट	11	0.20	(0.1, 0.33)	0.24

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	15	0.27	(0.16, 0.41)	0.29
Dha	ध	29	0.53	(0.39, 0.66)	0.50
U	उ	11	0.20	(0.1, 0.33)	0.21

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	9	0.16	(0.08, 0.29)	0.22
Gha	घ	34	0.62	(0.48, 0.75)	0.61
Ha	ह	12	0.22	(0.12, 0.35)	0.17

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	2	0.04	(0.0, 0.13)	0.22
Gha	घ	38	0.69	(0.55, 0.81)	0.54
Tta	ट	15	0.27	(0.16, 0.41)	0.24

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	17	0.31	(0.19, 0.45)	0.48
Gha	घ	29	0.53	(0.39, 0.66)	0.40
U	उ	9	0.16	(0.08, 0.29)	0.12

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	4	0.07	(0.02, 0.18)	0.33
Ha	ह	25	0.45	(0.32, 0.59)	0.28
Tta	ट	26	0.47	(0.34, 0.61)	0.39

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	4	0.07	(0.02, 0.18)	0.12
Ha	ह	20	0.36	(0.24, 0.5)	0.31
U	उ	31	0.56	(0.42, 0.7)	0.58

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	4	0.07	(0.02, 0.18)	0.11
Tta	ट	27	0.49	(0.35, 0.63)	0.38
U	उ	24	0.44	(0.3, 0.58)	0.51

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dha	ध	6	0.11	(0.04, 0.22)	0.18
Gha	घ	4	0.07	(0.02, 0.18)	0.17
Ha	ह	45	0.82	(0.69, 0.91)	0.65

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dha	ध	6	0.11	(0.04, 0.22)	0.18
Gha	घ	4	0.07	(0.02, 0.18)	0.15
Tta	ट	45	0.82	(0.69, 0.91)	0.67

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dha	ध	10	0.18	(0.09, 0.31)	0.36
Gha	घ	3	0.05	(0.01, 0.15)	0.07
U	उ	42	0.76	(0.63, 0.87)	0.58

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dha	ध	28	0.51	(0.37, 0.65)	0.58
Ha	ह	13	0.24	(0.13, 0.37)	0.17
Tta	ट	14	0.25	(0.15, 0.39)	0.25

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dha	ध	19	0.35	(0.22, 0.49)	0.33
Ha	ह	24	0.44	(0.3, 0.58)	0.35
U	उ	12	0.22	(0.12, 0.35)	0.32

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dha	ध	16	0.29	(0.18, 0.43)	0.30
Tta	ट	29	0.53	(0.39, 0.66)	0.40
U	उ	10	0.18	(0.09, 0.31)	0.30

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Gha	घ	30	0.55	(0.41, 0.68)	0.57
Ha	ह़	8	0.15	(0.06, 0.27)	0.18
Tta	ट	17	0.31	(0.19, 0.45)	0.25

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Gha	घ	12	0.22	(0.12, 0.35)	0.22
Ha	ह़	24	0.44	(0.3, 0.58)	0.55
U	उ	19	0.35	(0.22, 0.49)	0.23

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Gha	घ	17	0.31	(0.19, 0.45)	0.20
Tta	ट	28	0.51	(0.37, 0.65)	0.58
U	उ	10	0.18	(0.09, 0.31)	0.22

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ha	ह़	12	0.22	(0.12, 0.35)	0.15
Tta	ट	13	0.24	(0.13, 0.37)	0.23
U	उ	30	0.55	(0.41, 0.68)	0.62

Devanagari MT (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	13	0.22	(0.13, 0.35)	0.17
Ddha	ढ	21	0.36	(0.24, 0.5)	0.61
E	ए	24	0.41	(0.29, 0.55)	0.22

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	12	0.21	(0.11, 0.33)	0.36
Ddha	ढ	36	0.62	(0.48, 0.74)	0.50
Ja	ज	10	0.17	(0.09, 0.29)	0.14

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	25	0.43	(0.3, 0.57)	0.35
Ddha	ढ	19	0.33	(0.21, 0.46)	0.32
Ka	क	14	0.24	(0.14, 0.37)	0.33

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	3	0.05	(0.01, 0.14)	0.17
Ddha	ढ	51	0.88	(0.77, 0.95)	0.68
Ma	म	4	0.07	(0.02, 0.17)	0.16

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	9	0.16	(0.07, 0.27)	0.10
Ddha	ढ	41	0.71	(0.57, 0.82)	0.74
Na	न	8	0.14	(0.06, 0.25)	0.17

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	8	0.14	(0.06, 0.25)	0.28
Ddha	ढ	38	0.66	(0.52, 0.78)	0.48
Sha	श	12	0.21	(0.11, 0.33)	0.24

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	10	0.17	(0.09, 0.29)	0.23
E	ए	40	0.69	(0.55, 0.8)	0.35
Ja	ज	8	0.14	(0.06, 0.25)	0.42

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	9	0.16	(0.07, 0.27)	0.31
E	ए	36	0.62	(0.48, 0.74)	0.12
Ka	क	13	0.22	(0.13, 0.35)	0.57

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	5	0.09	(0.03, 0.19)	0.16
E	ए	51	0.88	(0.77, 0.95)	0.50
Ma	म	2	0.03	(0.0, 0.12)	0.34

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	3	0.05	(0.01, 0.14)	0.18
E	ए	46	0.79	(0.67, 0.89)	0.51
Na	न	9	0.16	(0.07, 0.27)	0.31

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	11	0.19	(0.1, 0.31)	0.25
E	ए	41	0.71	(0.57, 0.82)	0.29
Sha	श	6	0.10	(0.04, 0.21)	0.46

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	27	0.47	(0.33, 0.6)	0.36
Ja	ज	3	0.05	(0.01, 0.14)	0.13
Ka	क	28	0.48	(0.35, 0.62)	0.52

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	10	0.17	(0.09, 0.29)	0.16
Ja	ज	45	0.78	(0.65, 0.87)	0.53
Ma	म	3	0.05	(0.01, 0.14)	0.31

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	25	0.43	(0.3, 0.57)	0.17
Ja	ज	27	0.47	(0.33, 0.6)	0.54
Na	न	6	0.10	(0.04, 0.21)	0.28

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	6	0.10	(0.04, 0.21)	0.29
Ja	ज	24	0.41	(0.29, 0.55)	0.30
Sha	श	28	0.48	(0.35, 0.62)	0.41

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	8	0.14	(0.06, 0.25)	0.15
Ka	क	47	0.81	(0.69, 0.9)	0.70
Ma	म	3	0.05	(0.01, 0.14)	0.14

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	9	0.16	(0.07, 0.27)	0.08
Ka	क	41	0.71	(0.57, 0.82)	0.76
Na	न	8	0.14	(0.06, 0.25)	0.16

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	11	0.19	(0.1, 0.31)	0.37
Ka	क	28	0.48	(0.35, 0.62)	0.46
Sha	श	19	0.33	(0.21, 0.46)	0.17

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	23	0.40	(0.27, 0.53)	0.40
Ma	म	2	0.03	(0.0, 0.12)	0.32
Na	न	33	0.57	(0.43, 0.7)	0.29

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	2	0.03	(0.0, 0.12)	0.11
Ma	म	10	0.17	(0.09, 0.29)	0.28
Sha	श	46	0.79	(0.67, 0.89)	0.61

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	6	0.10	(0.04, 0.21)	0.12
Na	न	18	0.31	(0.2, 0.45)	0.25
Sha	श	34	0.59	(0.45, 0.71)	0.63

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	13	0.22	(0.13, 0.35)	0.47
E	ए	39	0.67	(0.54, 0.79)	0.40
Ja	ज	6	0.10	(0.04, 0.21)	0.13

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	4	0.07	(0.02, 0.17)	0.45
E	ए	47	0.81	(0.69, 0.9)	0.32
Ka	क	7	0.12	(0.05, 0.23)	0.24

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	31	0.53	(0.4, 0.67)	0.52
E	ए	17	0.29	(0.18, 0.43)	0.26
Ma	म	10	0.17	(0.09, 0.29)	0.22

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	19	0.33	(0.21, 0.46)	0.58
E	ए	26	0.45	(0.32, 0.58)	0.19
Na	न	13	0.22	(0.13, 0.35)	0.23

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	15	0.26	(0.15, 0.39)	0.50
E	ए	28	0.48	(0.35, 0.62)	0.31
Sha	श	15	0.26	(0.15, 0.39)	0.20

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	34	0.59	(0.45, 0.71)	0.40
Ja	ज	11	0.19	(0.1, 0.31)	0.22
Ka	क	13	0.22	(0.13, 0.35)	0.39

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	40	0.69	(0.55, 0.8)	0.41
Ja	ज	4	0.07	(0.02, 0.17)	0.19
Ma	म	14	0.24	(0.14, 0.37)	0.40

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	44	0.76	(0.63, 0.86)	0.48
Ja	ज	4	0.07	(0.02, 0.17)	0.11
Na	न	10	0.17	(0.09, 0.29)	0.41

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	19	0.33	(0.21, 0.46)	0.44
Ja	ज	11	0.19	(0.1, 0.31)	0.20
Sha	श	28	0.48	(0.35, 0.62)	0.36

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	26	0.45	(0.32, 0.58)	0.31
Ka	क	5	0.09	(0.03, 0.19)	0.32
Ma	म	27	0.47	(0.33, 0.6)	0.36

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	22	0.38	(0.26, 0.52)	0.29
Ka	क	10	0.17	(0.09, 0.29)	0.29
Na	न	26	0.45	(0.32, 0.58)	0.41

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	16	0.28	(0.17, 0.41)	0.43
Ka	क	7	0.12	(0.05, 0.23)	0.30
Sha	श	35	0.60	(0.47, 0.73)	0.27

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	47	0.81	(0.69, 0.9)	0.76
Ma	म	7	0.12	(0.05, 0.23)	0.08
Na	न	4	0.07	(0.02, 0.17)	0.16

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	25	0.43	(0.3, 0.57)	0.38
Ma	म	10	0.17	(0.09, 0.29)	0.34
Sha	श	23	0.40	(0.27, 0.53)	0.28

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	22	0.38	(0.26, 0.52)	0.44
Na	न	16	0.28	(0.17, 0.41)	0.35
Sha	श	20	0.34	(0.22, 0.48)	0.20

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
E	ए	51	0.88	(0.77, 0.95)	0.33
Ja	ज	3	0.05	(0.01, 0.14)	0.28
Ka	क	4	0.07	(0.02, 0.17)	0.39

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
E	ए	36	0.62	(0.48, 0.74)	0.30
Ja	ज	15	0.26	(0.15, 0.39)	0.34
Ma	म	7	0.12	(0.05, 0.23)	0.36

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
E	ए	54	0.93	(0.83, 0.98)	0.32
Ja	ज	1	0.02	(0.0, 0.09)	0.36
Na	न	3	0.05	(0.01, 0.14)	0.33

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
E	ए	31	0.53	(0.4, 0.67)	0.35
Ja	ज	14	0.24	(0.14, 0.37)	0.31
Sha	श	13	0.22	(0.13, 0.35)	0.34

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
E	ए	27	0.47	(0.33, 0.6)	0.18
Ka	क	17	0.29	(0.18, 0.43)	0.47
Ma	म	14	0.24	(0.14, 0.37)	0.35

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
E	ए	40	0.69	(0.55, 0.8)	0.10
Ka	क	9	0.16	(0.07, 0.27)	0.53
Na	न	9	0.16	(0.07, 0.27)	0.37

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
E	ए	36	0.62	(0.48, 0.74)	0.32
Ka	क	10	0.17	(0.09, 0.29)	0.38
Sha	श	12	0.21	(0.11, 0.33)	0.30

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
E	ए	50	0.86	(0.75, 0.94)	0.62
Ma	म	4	0.07	(0.02, 0.17)	0.21
Na	न	4	0.07	(0.02, 0.17)	0.17

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
E	ए	25	0.43	(0.3, 0.57)	0.22
Ma	म	10	0.17	(0.09, 0.29)	0.38
Sha	श	23	0.40	(0.27, 0.53)	0.40

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
E	ए	30	0.52	(0.38, 0.65)	0.23
Na	न	12	0.21	(0.11, 0.33)	0.35
Sha	श	16	0.28	(0.17, 0.41)	0.42

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ja	ज	9	0.16	(0.07, 0.27)	0.18
Ka	क	22	0.38	(0.26, 0.52)	0.43
Ma	म	27	0.47	(0.33, 0.6)	0.39

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ja	ज	5	0.09	(0.03, 0.19)	0.10
Ka	क	38	0.66	(0.52, 0.78)	0.49
Na	न	15	0.26	(0.15, 0.39)	0.41

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ja	ज	10	0.17	(0.09, 0.29)	0.29
Ka	क	10	0.17	(0.09, 0.29)	0.39
Sha	श	38	0.66	(0.52, 0.78)	0.32

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ja	ज	46	0.79	(0.67, 0.89)	0.64
Ma	म	9	0.16	(0.07, 0.27)	0.20
Na	न	3	0.05	(0.01, 0.14)	0.16

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ja	ज	7	0.12	(0.05, 0.23)	0.22
Ma	म	13	0.22	(0.13, 0.35)	0.41
Sha	श	38	0.66	(0.52, 0.78)	0.36

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ja	ज	-	-	-	0.24
Na	न	13	0.22	-	0.38
Sha	श	45	0.78	-	0.38

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	50	0.86	(0.75, 0.94)	0.78
Ma	म	5	0.09	(0.03, 0.19)	0.07
Na	न	3	0.05	(0.01, 0.14)	0.15

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	12	0.21	(0.11, 0.33)	0.35
Ma	म	18	0.31	(0.2, 0.45)	0.42
Sha	श	28	0.48	(0.35, 0.62)	0.23

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	10	0.17	(0.09, 0.29)	0.41
Na	न	19	0.33	(0.21, 0.46)	0.44
Sha	श	29	0.50	(0.37, 0.63)	0.15

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ma	म	3	0.05	(0.01, 0.14)	0.15
Na	न	5	0.09	(0.03, 0.19)	0.11
Sha	श	50	0.86	(0.75, 0.94)	0.73

Ek Mukta (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	11	0.19	(0.1, 0.32)	0.32
Ca	च	17	0.30	(0.18, 0.43)	0.32
Jha	झ	29	0.51	(0.37, 0.64)	0.36

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	6	0.11	(0.04, 0.22)	0.12
Ca	च	45	0.79	(0.66, 0.89)	0.67
Ma	म	6	0.11	(0.04, 0.22)	0.20

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	46	0.81	(0.68, 0.9)	0.35
Ca	च	2	0.04	(0.0, 0.12)	0.20
Nya	ञ	9	0.16	(0.07, 0.28)	0.45

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	12	0.21	(0.11, 0.34)	0.32
Ca	च	10	0.18	(0.09, 0.3)	0.20
Ra	र	35	0.61	(0.48, 0.74)	0.48

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	30	0.53	(0.39, 0.66)	0.37
Ca	च	5	0.09	(0.03, 0.19)	0.21
Ttha	ठ	22	0.39	(0.26, 0.52)	0.42

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	41	0.72	(0.58, 0.83)	0.66
Ca	च	3	0.05	(0.01, 0.15)	0.09
Va	व	13	0.23	(0.13, 0.36)	0.25

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	5	0.09	(0.03, 0.19)	0.14
Jha	झ	49	0.86	(0.74, 0.94)	0.67
Ma	म	3	0.05	(0.01, 0.15)	0.20

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	21	0.37	(0.24, 0.51)	0.49
Jha	झ	13	0.23	(0.13, 0.36)	0.15
Nya	ञ	23	0.40	(0.28, 0.54)	0.36

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	4	0.07	(0.02, 0.17)	0.23
Jha	झ	21	0.37	(0.24, 0.51)	0.27
Ra	र	32	0.56	(0.42, 0.69)	0.50

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	9	0.16	(0.07, 0.28)	0.37
Jha	झ	17	0.30	(0.18, 0.43)	0.25
Ttha	ठ	31	0.54	(0.41, 0.68)	0.39

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	10	0.18	(0.09, 0.3)	0.26
Jha	झ	20	0.35	(0.23, 0.49)	0.30
Va	व	27	0.47	(0.34, 0.61)	0.44

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	4	0.07	(0.02, 0.17)	0.06
Ma	म	6	0.11	(0.04, 0.22)	0.11
Nya	ञ	47	0.82	(0.7, 0.91)	0.83

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	1	0.02	(0.0, 0.09)	0.03
Ma	म	5	0.09	(0.03, 0.19)	0.16
Ra	र	51	0.89	(0.78, 0.96)	0.81

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	5	0.09	(0.03, 0.19)	0.13
Ma	म	6	0.11	(0.04, 0.22)	0.10
Ttha	ठ	46	0.81	(0.68, 0.9)	0.76

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	5	0.09	(0.03, 0.19)	0.07
Ma	म	3	0.05	(0.01, 0.15)	0.14
Va	व	49	0.86	(0.74, 0.94)	0.79

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	7	0.12	(0.05, 0.24)	0.40
Nya	ञ	19	0.33	(0.21, 0.47)	0.29
Ra	र	31	0.54	(0.41, 0.68)	0.30

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	26	0.46	(0.32, 0.59)	0.31
Nya	ञ	13	0.23	(0.13, 0.36)	0.38
Ttha	ठ	18	0.32	(0.2, 0.45)	0.31

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	35	0.61	(0.48, 0.74)	0.47
Nya	ञ	13	0.23	(0.13, 0.36)	0.32
Va	व	9	0.16	(0.07, 0.28)	0.21

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	18	0.32	(0.2, 0.45)	0.30
Ra	र	12	0.21	(0.11, 0.34)	0.39
Ttha	ठ	27	0.47	(0.34, 0.61)	0.31

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	19	0.33	(0.21, 0.47)	0.45
Ra	र	25	0.44	(0.31, 0.58)	0.34
Va	व	13	0.23	(0.13, 0.36)	0.21

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Bha	भ	46	0.81	(0.68, 0.9)	0.34
Ttha	ठ	5	0.09	(0.03, 0.19)	0.35
Va	व	6	0.11	(0.04, 0.22)	0.31

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	15	0.26	(0.16, 0.4)	0.30
Jha	झ	24	0.42	(0.29, 0.56)	0.34
Ma	म	18	0.32	(0.2, 0.45)	0.36

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	5	0.09	(0.03, 0.19)	0.48
Jha	झ	43	0.75	(0.62, 0.86)	0.24
Nya	ञ	9	0.16	(0.07, 0.28)	0.29

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	3	0.05	(0.01, 0.15)	0.19
Jha	झ	17	0.30	(0.18, 0.43)	0.36
Ra	र	37	0.65	(0.51, 0.77)	0.45

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	4	0.07	(0.02, 0.17)	0.33
Jha	झ	32	0.56	(0.42, 0.69)	0.33
Ttha	ठ	21	0.37	(0.24, 0.51)	0.34

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	–	–	–	0.10
Jha	झ	48	0.84	–	0.65
Va	व	9	0.16	–	0.24

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	2	0.04	(0.0, 0.12)	0.18
Ma	म	44	0.77	(0.64, 0.87)	0.39
Nya	ञ	11	0.19	(0.1, 0.32)	0.43

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	10	0.18	(0.09, 0.3)	0.16
Ma	म	12	0.21	(0.11, 0.34)	0.39
Ra	र	35	0.61	(0.48, 0.74)	0.45

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	4	0.07	(0.02, 0.17)	0.26
Ma	म	33	0.58	(0.44, 0.71)	0.38
Ttha	ठ	20	0.35	(0.23, 0.49)	0.36

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	6	0.11	(0.04, 0.22)	0.07
Ma	म	44	0.77	(0.64, 0.87)	0.70
Va	व	7	0.12	(0.05, 0.24)	0.24

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	1	0.02	(0.0, 0.09)	0.31
Nya	ञ	12	0.21	(0.11, 0.34)	0.32
Ra	र	44	0.77	(0.64, 0.87)	0.37

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	5	0.09	(0.03, 0.19)	0.22
Nya	ञ	10	0.18	(0.09, 0.3)	0.40
Ttha	ठ	42	0.74	(0.6, 0.84)	0.38

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	8	0.14	(0.06, 0.26)	0.23
Nya	ञ	31	0.54	(0.41, 0.68)	0.65
Va	व	18	0.32	(0.2, 0.45)	0.12

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	7	0.12	(0.05, 0.24)	0.21
Ra	र	36	0.63	(0.49, 0.76)	0.44
Ttha	ठ	14	0.25	(0.14, 0.38)	0.36

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	8	0.14	(0.06, 0.26)	0.20
Ra	र	45	0.79	(0.66, 0.89)	0.73
Va	व	4	0.07	(0.02, 0.17)	0.08

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ca	च	9	0.16	(0.07, 0.28)	0.09
Ttha	ठ	42	0.74	(0.6, 0.84)	0.72
Va	व	6	0.11	(0.04, 0.22)	0.19

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Jha	झ	13	0.23	(0.13, 0.36)	0.14
Ma	म	23	0.40	(0.28, 0.54)	0.53
Nya	ञ	21	0.37	(0.24, 0.51)	0.33

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Jha	झ	19	0.33	(0.21, 0.47)	0.25
Ma	म	4	0.07	(0.02, 0.17)	0.27
Ra	र	34	0.60	(0.46, 0.72)	0.48

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Jha	झ	16	0.28	(0.17, 0.42)	0.28
Ma	म	10	0.18	(0.09, 0.3)	0.39
Ttha	ठ	31	0.54	(0.41, 0.68)	0.33

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Jha	झ	22	0.39	(0.26, 0.52)	0.28
Ma	म	8	0.14	(0.06, 0.26)	0.32
Va	व	27	0.47	(0.34, 0.61)	0.40

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Jha	झ	12	0.21	(0.11, 0.34)	0.29
Nya	ञ	7	0.12	(0.05, 0.24)	0.14
Ra	र	38	0.67	(0.53, 0.79)	0.58

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Jha	झ	17	0.30	(0.18, 0.43)	0.16
Nya	ञ	4	0.07	(0.02, 0.17)	0.33
Ttha	ठ	36	0.63	(0.49, 0.76)	0.50

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Jha	झ	35	0.61	(0.48, 0.74)	0.34
Nya	ञ	10	0.18	(0.09, 0.3)	0.17
Va	व	12	0.21	(0.11, 0.34)	0.49

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Jha	झ	22	0.39	(0.26, 0.52)	0.28
Ra	र	21	0.37	(0.24, 0.51)	0.49
Ttha	ठ	14	0.25	(0.14, 0.38)	0.24

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Jha	झ	28	0.49	(0.36, 0.63)	0.47
Ra	र	21	0.37	(0.24, 0.51)	0.33
Va	व	8	0.14	(0.06, 0.26)	0.20

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Jha	झ	47	0.82	(0.7, 0.91)	0.31
Ttha	ठ	5	0.09	(0.03, 0.19)	0.28
Va	व	5	0.09	(0.03, 0.19)	0.42

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ma	म	8	0.14	(0.06, 0.26)	0.45
Nya	ञ	24	0.42	(0.29, 0.56)	0.26
Ra	र	25	0.44	(0.31, 0.58)	0.29

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ma	म	21	0.37	(0.24, 0.51)	0.34
Nya	ञ	13	0.23	(0.13, 0.36)	0.40
Ttha	ठ	23	0.40	(0.28, 0.54)	0.26

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ma	म	40	0.70	(0.57, 0.82)	0.52
Nya	ञ	11	0.19	(0.1, 0.32)	0.29
Va	व	6	0.11	(0.04, 0.22)	0.19

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ma	म	15	0.26	(0.16, 0.4)	0.32
Ra	र	18	0.32	(0.2, 0.45)	0.43
Ttha	ठ	24	0.42	(0.29, 0.56)	0.25

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ma	म	16	0.28	(0.17, 0.42)	0.51
Ra	र	29	0.51	(0.37, 0.64)	0.31
Va	व	12	0.21	(0.11, 0.34)	0.17

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ma	म	42	0.74	(0.6, 0.84)	0.37
Ttha	ठ	10	0.18	(0.09, 0.3)	0.29
Va	व	5	0.09	(0.03, 0.19)	0.34

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nya	ञ	16	0.28	(0.17, 0.42)	0.28
Ra	र	32	0.56	(0.42, 0.69)	0.31
Ttha	ठ	9	0.16	(0.07, 0.28)	0.41

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nya	ञ	7	0.12	(0.05, 0.24)	0.36
Ra	र	42	0.74	(0.6, 0.84)	0.42
Va	व	8	0.14	(0.06, 0.26)	0.22

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nya	ञ	29	0.51	(0.37, 0.64)	0.31
Ttha	ठ	24	0.42	(0.29, 0.56)	0.48
Va	व	4	0.07	(0.02, 0.17)	0.22

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ra	र	44	0.77	(0.64, 0.87)	0.34
Ttha	ठ	6	0.11	(0.04, 0.22)	0.46
Va	व	7	0.12	(0.05, 0.24)	0.20

ITF Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	14	0.25	(0.14, 0.38)	0.14
Dda	ड	20	0.35	(0.23, 0.49)	0.44
Ga	ग	23	0.40	(0.28, 0.54)	0.42

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	53	0.93	(0.83, 0.98)	0.61
Dda	ड	1	0.02	(0.0, 0.09)	0.17
Ha	ह	3	0.05	(0.01, 0.15)	0.22

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	51	0.89	(0.78, 0.96)	0.70
Dda	ड	2	0.04	(0.0, 0.12)	0.12
I	इ	4	0.07	(0.02, 0.17)	0.18

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	9	0.16	(0.07, 0.28)	0.29
Dda	ड	22	0.39	(0.26, 0.52)	0.32
Na	न	26	0.46	(0.32, 0.59)	0.39

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	16	0.28	(0.17, 0.42)	0.18
Dda	ड	37	0.65	(0.51, 0.77)	0.50
Pa	प	4	0.07	(0.02, 0.17)	0.32

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	16	0.28	(0.17, 0.42)	0.19
Dda	ड	23	0.40	(0.28, 0.54)	0.50
Sa	स	18	0.32	(0.2, 0.45)	0.31

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	10	0.18	(0.09, 0.3)	0.14
Ga	ग	26	0.46	(0.32, 0.59)	0.39
Ha	ह	21	0.37	(0.24, 0.51)	0.48

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	12	0.21	(0.11, 0.34)	0.26
Ga	ग	24	0.42	(0.29, 0.56)	0.35
I	इ	21	0.37	(0.24, 0.51)	0.40

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	49	0.86	(0.74, 0.94)	0.69
Ga	ग	6	0.11	(0.04, 0.22)	0.10
Na	न	2	0.04	(0.0, 0.12)	0.21

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	39	0.68	(0.55, 0.8)	0.51
Ga	ग	17	0.30	(0.18, 0.43)	0.33
Pa	प	1	0.02	(0.0, 0.09)	0.16

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	41	0.72	(0.58, 0.83)	0.21
Ga	ग	10	0.18	(0.09, 0.3)	0.50
Sa	स	6	0.11	(0.04, 0.22)	0.30

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	52	0.91	(0.81, 0.97)	0.56
Ha	ह	1	0.02	(0.0, 0.09)	0.21
I	इ	4	0.07	(0.02, 0.17)	0.23

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	16	0.28	(0.17, 0.42)	0.28
Ha	ह	23	0.40	(0.28, 0.54)	0.35
Na	न	18	0.32	(0.2, 0.45)	0.36

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	7	0.12	(0.05, 0.24)	0.17
Ha	ह	38	0.67	(0.53, 0.79)	0.53
Pa	प	12	0.21	(0.11, 0.34)	0.30

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	21	0.37	(0.24, 0.51)	0.18
Ha	ह	25	0.44	(0.31, 0.58)	0.53
Sa	स	11	0.19	(0.1, 0.32)	0.29

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	8	0.14	(0.06, 0.26)	0.40
I	इ	27	0.47	(0.34, 0.61)	0.28
Na	न	22	0.39	(0.26, 0.52)	0.32

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	10	0.18	(0.09, 0.3)	0.17
I	इ	39	0.68	(0.55, 0.8)	0.53
Pa	प	8	0.14	(0.06, 0.26)	0.30

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	15	0.26	(0.16, 0.4)	0.19
I	इ	29	0.51	(0.37, 0.64)	0.53
Sa	स	13	0.23	(0.13, 0.36)	0.28

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	28	0.49	(0.36, 0.63)	0.44
Na	न	24	0.42	(0.29, 0.56)	0.41
Pa	प	5	0.09	(0.03, 0.19)	0.15

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	30	0.53	(0.39, 0.66)	0.18
Na	न	14	0.25	(0.14, 0.38)	0.56
Sa	स	13	0.23	(0.13, 0.36)	0.27

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ba	ब	28	0.49	(0.36, 0.63)	0.27
Pa	प	5	0.09	(0.03, 0.19)	0.38
Sa	स	24	0.42	(0.29, 0.56)	0.36

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	1	0.02	(0.0, 0.09)	0.12
Ga	ग	52	0.91	(0.81, 0.97)	0.73
Ha	ह	4	0.07	(0.02, 0.17)	0.16

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	2	0.04	(0.0, 0.12)	0.18
Ga	ग	51	0.89	(0.78, 0.96)	0.76
I	इ	4	0.07	(0.02, 0.17)	0.06

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	49	0.86	(0.74, 0.94)	0.74
Ga	ग	6	0.11	(0.04, 0.22)	0.21
Na	न	2	0.04	(0.0, 0.12)	0.05

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	42	0.74	(0.6, 0.84)	0.68
Ga	ग	13	0.23	(0.13, 0.36)	0.21
Pa	प	2	0.04	(0.0, 0.12)	0.10

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	37	0.65	(0.51, 0.77)	0.43
Ga	ग	15	0.26	(0.16, 0.4)	0.35
Sa	स	5	0.09	(0.03, 0.19)	0.22

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	8	0.14	(0.06, 0.26)	0.24
Ha	ह	40	0.70	(0.57, 0.82)	0.44
I	इ	9	0.16	(0.07, 0.28)	0.31

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	2	0.04	(0.0, 0.12)	0.17
Ha	ह	6	0.11	(0.04, 0.22)	0.21
Na	न	49	0.86	(0.74, 0.94)	0.61

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	3	0.05	(0.01, 0.15)	0.15
Ha	ह	5	0.09	(0.03, 0.19)	0.19
Pa	प	49	0.86	(0.74, 0.94)	0.67

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	8	0.14	(0.06, 0.26)	0.15
Ha	ह	3	0.05	(0.01, 0.15)	0.19
Sa	स	46	0.81	(0.68, 0.9)	0.66

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	2	0.04	(0.0, 0.12)	0.24
I	इ	3	0.05	(0.01, 0.15)	0.11
Na	न	52	0.91	(0.81, 0.97)	0.65

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	1	0.02	(0.0, 0.09)	0.09
I	इ	5	0.09	(0.03, 0.19)	0.15
Pa	प	51	0.89	(0.78, 0.96)	0.76

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	3	0.05	(0.01, 0.15)	0.11
I	इ	5	0.09	(0.03, 0.19)	0.15
Sa	स	49	0.86	(0.74, 0.94)	0.74

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	44	0.77	(0.64, 0.87)	0.54
Na	न	11	0.19	(0.1, 0.32)	0.20
Pa	प	2	0.04	(0.0, 0.12)	0.25

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	35	0.61	(0.48, 0.74)	0.30
Na	न	10	0.18	(0.09, 0.3)	0.34
Sa	स	12	0.21	(0.11, 0.34)	0.37

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dda	ड	39	0.68	(0.55, 0.8)	0.47
Pa	प	6	0.11	(0.04, 0.22)	0.27
Sa	स	12	0.21	(0.11, 0.34)	0.26

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	50	0.88	(0.76, 0.95)	0.62
Ha	ह	3	0.05	(0.01, 0.15)	0.26
I	इ	4	0.07	(0.02, 0.17)	0.11

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	7	0.12	(0.05, 0.24)	0.20
Ha	ह	45	0.79	(0.66, 0.89)	0.75
Na	न	5	0.09	(0.03, 0.19)	0.05

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	6	0.11	(0.04, 0.22)	0.20
Ha	ह	49	0.86	(0.74, 0.94)	0.71
Pa	प	2	0.04	(0.0, 0.12)	0.10

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	14	0.25	(0.14, 0.38)	0.33
Ha	ह	37	0.65	(0.51, 0.77)	0.46
Sa	स	6	0.11	(0.04, 0.22)	0.21

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	6	0.11	(0.04, 0.22)	0.26
I	इ	49	0.86	(0.74, 0.94)	0.65
Na	न	2	0.04	(0.0, 0.12)	0.09

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	8	0.14	(0.06, 0.26)	0.16
I	इ	44	0.77	(0.64, 0.87)	0.62
Pa	प	5	0.09	(0.03, 0.19)	0.22

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	13	0.23	(0.13, 0.36)	0.30
I	इ	40	0.70	(0.57, 0.82)	0.38
Sa	स	4	0.07	(0.02, 0.17)	0.32

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	11	0.19	(0.1, 0.32)	0.19
Na	न	13	0.23	(0.13, 0.36)	0.32
Pa	प	33	0.58	(0.44, 0.71)	0.49

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	6	0.11	(0.04, 0.22)	0.08
Na	न	4	0.07	(0.02, 0.17)	0.20
Sa	स	47	0.82	(0.7, 0.91)	0.71

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	16	0.28	(0.17, 0.42)	0.33
Pa	प	14	0.25	(0.14, 0.38)	0.14
Sa	स	27	0.47	(0.34, 0.61)	0.52

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ha	ह	3	0.05	(0.01, 0.15)	0.32
I	इ	5	0.09	(0.03, 0.19)	0.16
Na	न	49	0.86	(0.74, 0.94)	0.52

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ha	ह	4	0.07	(0.02, 0.17)	0.18
I	इ	7	0.12	(0.05, 0.24)	0.20
Pa	प	46	0.81	(0.68, 0.9)	0.62

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ha	ह	3	0.05	(0.01, 0.15)	0.20
I	इ	4	0.07	(0.02, 0.17)	0.19
Sa	स	50	0.88	(0.76, 0.95)	0.61

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ha	ह	40	0.70	(0.57, 0.82)	0.57
Na	न	12	0.21	(0.11, 0.34)	0.19
Pa	प	5	0.09	(0.03, 0.19)	0.25

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ha	ह	35	0.61	(0.48, 0.74)	0.33
Na	न	18	0.32	(0.2, 0.45)	0.31
Sa	स	4	0.07	(0.02, 0.17)	0.35

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ha	ह	33	0.58	(0.44, 0.71)	0.50
Pa	प	14	0.25	(0.14, 0.38)	0.26
Sa	स	10	0.18	(0.09, 0.3)	0.25

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
I	इ	42	0.74	(0.6, 0.84)	0.49
Na	न	9	0.16	(0.07, 0.28)	0.15
Pa	प	6	0.11	(0.04, 0.22)	0.36

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
I	इ	32	0.56	(0.42, 0.69)	0.26
Na	न	15	0.26	(0.16, 0.4)	0.29
Sa	स	10	0.18	(0.09, 0.3)	0.45

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
I	इ	41	0.72	(0.58, 0.83)	0.49
Pa	प	9	0.16	(0.07, 0.28)	0.27
Sa	स	7	0.12	(0.05, 0.24)	0.24

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Na	न	24	0.42	(0.29, 0.56)	0.41
Pa	प	9	0.16	(0.07, 0.28)	0.12
Sa	स	24	0.42	(0.29, 0.56)	0.47

Kohinoor Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	48	0.76	(0.64, 0.86)	0.49
Ka	क	6	0.10	(0.04, 0.2)	0.28
La	ल	9	0.14	(0.07, 0.25)	0.24

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	18	0.29	(0.18, 0.41)	0.28
Ka	क	41	0.65	(0.52, 0.77)	0.58
Nna	ण	4	0.06	(0.02, 0.15)	0.14

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	23	0.37	(0.25, 0.5)	0.27
Ka	क	35	0.56	(0.42, 0.68)	0.58
Pa	प	5	0.08	(0.03, 0.18)	0.15

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	54	0.86	(0.75, 0.93)	0.47
Ka	क	3	0.05	(0.01, 0.13)	0.41
Pha	फ	6	0.10	(0.04, 0.2)	0.12

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	31	0.49	(0.36, 0.62)	0.44
Ka	क	19	0.30	(0.19, 0.43)	0.23
Tha	थ	13	0.21	(0.11, 0.33)	0.33

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	25	0.40	(0.28, 0.53)	0.44
Ka	क	31	0.49	(0.36, 0.62)	0.27
Ya	य	7	0.11	(0.05, 0.22)	0.30

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	16	0.25	(0.15, 0.38)	0.31
La	ल	41	0.65	(0.52, 0.77)	0.55
Nna	ण	6	0.10	(0.04, 0.2)	0.14

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	24	0.38	(0.26, 0.51)	0.29
La	ल	37	0.59	(0.46, 0.71)	0.55
Pa	प	2	0.03	(0.0, 0.11)	0.15

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	37	0.59	(0.46, 0.71)	0.32
La	ल	21	0.33	(0.22, 0.46)	0.47
Pha	फ	5	0.08	(0.03, 0.18)	0.21

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	32	0.51	(0.38, 0.64)	0.45
La	ल	15	0.24	(0.14, 0.36)	0.21
Tha	थ	16	0.25	(0.15, 0.38)	0.35

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	26	0.41	(0.29, 0.54)	0.47
La	ल	30	0.48	(0.35, 0.61)	0.23
Ya	य	7	0.11	(0.05, 0.22)	0.30

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	45	0.71	(0.59, 0.82)	0.58
Nna	ण	6	0.10	(0.04, 0.2)	0.20
Pa	प	12	0.19	(0.1, 0.31)	0.22

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	36	0.57	(0.44, 0.7)	0.51
Nna	ण	9	0.14	(0.07, 0.25)	0.19
Pha	फ	18	0.29	(0.18, 0.41)	0.30

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	15	0.24	(0.14, 0.36)	0.35
Nna	ण	9	0.14	(0.07, 0.25)	0.07
Tha	थ	39	0.62	(0.49, 0.74)	0.59

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	26	0.41	(0.29, 0.54)	0.39
Nna	ण	9	0.14	(0.07, 0.25)	0.08
Ya	य	28	0.44	(0.32, 0.58)	0.52

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	50	0.79	(0.67, 0.89)	0.50
Pa	प	1	0.02	(0.0, 0.09)	0.20
Pha	फ	12	0.19	(0.1, 0.31)	0.30

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	29	0.46	(0.33, 0.59)	0.34
Pa	प	4	0.06	(0.02, 0.15)	0.07
Tha	थ	30	0.48	(0.35, 0.61)	0.59

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	52	0.83	(0.71, 0.91)	0.38
Pa	प	5	0.08	(0.03, 0.18)	0.09
Ya	य	6	0.10	(0.04, 0.2)	0.52

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	36	0.57	(0.44, 0.7)	0.35
Pha	फ	11	0.17	(0.09, 0.29)	0.14
Tha	थ	16	0.25	(0.15, 0.38)	0.51

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	42	0.67	(0.54, 0.78)	0.37
Pha	फ	19	0.30	(0.19, 0.43)	0.16
Ya	य	2	0.03	(0.0, 0.11)	0.47

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ga	ग	47	0.75	(0.62, 0.85)	0.79
Tha	थ	13	0.21	(0.11, 0.33)	0.14
Ya	य	3	0.05	(0.01, 0.13)	0.07

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	10	0.16	(0.08, 0.27)	0.33
La	ल	14	0.22	(0.13, 0.34)	0.29
Nna	ण	39	0.62	(0.49, 0.74)	0.38

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	10	0.16	(0.08, 0.27)	0.33
La	ल	18	0.29	(0.18, 0.41)	0.28
Pa	प	35	0.56	(0.42, 0.68)	0.39

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	4	0.06	(0.02, 0.15)	0.22
La	ल	45	0.71	(0.59, 0.82)	0.46
Pha	फ	14	0.22	(0.13, 0.34)	0.32

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	11	0.17	(0.09, 0.29)	0.31
La	ल	11	0.17	(0.09, 0.29)	0.28
Tha	थ	41	0.65	(0.52, 0.77)	0.41

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	14	0.22	(0.13, 0.34)	0.33
La	ल	17	0.27	(0.17, 0.4)	0.29
Ya	य	32	0.51	(0.38, 0.64)	0.38

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	54	0.86	(0.75, 0.93)	0.72
Nna	ण	7	0.11	(0.05, 0.22)	0.13
Pa	प	2	0.03	(0.0, 0.11)	0.15

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	15	0.24	(0.14, 0.36)	0.55
Nna	ण	44	0.70	(0.57, 0.81)	0.33
Pha	फ	4	0.06	(0.02, 0.15)	0.12

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	18	0.29	(0.18, 0.41)	0.37
Nna	ण	26	0.41	(0.29, 0.54)	0.28
Tha	थ	19	0.30	(0.19, 0.43)	0.34

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	31	0.49	(0.36, 0.62)	0.41
Nna	ण	19	0.30	(0.19, 0.43)	0.29
Ya	य	13	0.21	(0.11, 0.33)	0.30

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	27	0.43	(0.3, 0.56)	0.55
Pa	प	32	0.51	(0.38, 0.64)	0.34
Pha	फ	4	0.06	(0.02, 0.15)	0.11

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	30	0.48	(0.35, 0.61)	0.37
Pa	प	17	0.27	(0.17, 0.4)	0.29
Tha	थ	16	0.25	(0.15, 0.38)	0.33

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	51	0.81	(0.69, 0.9)	0.41
Pa	प	6	0.10	(0.04, 0.2)	0.30
Ya	य	6	0.10	(0.04, 0.2)	0.30

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	11	0.17	(0.09, 0.29)	0.27
Pha	फ	5	0.08	(0.03, 0.18)	0.24
Tha	थ	47	0.75	(0.62, 0.85)	0.49

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	15	0.24	(0.14, 0.36)	0.29
Pha	फ	4	0.06	(0.02, 0.15)	0.25
Ya	य	44	0.70	(0.57, 0.81)	0.46

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ka	क	44	0.70	(0.57, 0.81)	0.67
Tha	थ	15	0.24	(0.14, 0.36)	0.18
Ya	य	4	0.06	(0.02, 0.15)	0.14

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
La	ल	42	0.67	(0.54, 0.78)	0.70
Nna	ण	15	0.24	(0.14, 0.36)	0.14
Pa	प	6	0.10	(0.04, 0.2)	0.16

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
La	ल	32	0.51	(0.38, 0.64)	0.63
Nna	ण	26	0.41	(0.29, 0.54)	0.16
Pha	फ	5	0.08	(0.03, 0.18)	0.22

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
La	ल	24	0.38	(0.26, 0.51)	0.35
Nna	ण	20	0.32	(0.21, 0.45)	0.29
Tha	थ	19	0.30	(0.19, 0.43)	0.36

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
La	ल	35	0.56	(0.42, 0.68)	0.38
Nna	ण	21	0.33	(0.22, 0.46)	0.30
Ya	य	7	0.11	(0.05, 0.22)	0.31

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
La	ल	49	0.78	(0.66, 0.87)	0.62
Pa	प	8	0.13	(0.06, 0.23)	0.17
Pha	फ	6	0.10	(0.04, 0.2)	0.21

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
La	ल	32	0.51	(0.38, 0.64)	0.35
Pa	प	14	0.22	(0.13, 0.34)	0.30
Tha	थ	17	0.27	(0.17, 0.4)	0.35

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
La	ल	53	0.84	(0.73, 0.92)	0.38
Pa	प	3	0.05	(0.01, 0.13)	0.31
Ya	य	7	0.11	(0.05, 0.22)	0.31

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
La	ल	23	0.37	(0.25, 0.5)	0.33
Pha	फ	11	0.17	(0.09, 0.29)	0.33
Tha	थ	29	0.46	(0.33, 0.59)	0.33

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
La	ल	31	0.49	(0.36, 0.62)	0.35
Pha	फ	10	0.16	(0.08, 0.27)	0.35
Ya	य	22	0.35	(0.23, 0.48)	0.30

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
La	ल	48	0.76	(0.64, 0.86)	0.66
Tha	थ	12	0.19	(0.1, 0.31)	0.20
Ya	य	3	0.05	(0.01, 0.13)	0.14

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nna	ण	31	0.49	(0.36, 0.62)	0.29
Pa	प	7	0.11	(0.05, 0.22)	0.31
Pha	फ	25	0.40	(0.28, 0.53)	0.40

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nna	ण	17	0.27	(0.17, 0.4)	0.15
Pa	प	6	0.10	(0.04, 0.2)	0.16
Tha	थ	40	0.63	(0.5, 0.75)	0.70

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nna	ण	36	0.57	(0.44, 0.7)	0.17
Pa	प	4	0.06	(0.02, 0.15)	0.19
Ya	य	23	0.37	(0.25, 0.5)	0.64

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nna	ण	22	0.35	(0.23, 0.48)	0.16
Pha	फ	5	0.08	(0.03, 0.18)	0.22
Tha	थ	36	0.57	(0.44, 0.7)	0.62

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nna	ण	27	0.43	(0.3, 0.56)	0.17
Pha	फ	17	0.27	(0.17, 0.4)	0.24
Ya	य	19	0.30	(0.19, 0.43)	0.58

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nna	ण	44	0.70	(0.57, 0.81)	0.64
Tha	थ	16	0.25	(0.15, 0.38)	0.22
Ya	य	3	0.05	(0.01, 0.13)	0.15

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Pa	प	6	0.10	(0.04, 0.2)	0.17
Pha	फ	10	0.16	(0.08, 0.27)	0.22
Tha	थ	47	0.75	(0.62, 0.85)	0.62

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Pa	प	7	0.11	(0.05, 0.22)	0.18
Pha	फ	30	0.48	(0.35, 0.61)	0.24
Ya	य	26	0.41	(0.29, 0.54)	0.58

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Pa	प	34	0.54	(0.41, 0.67)	0.64
Tha	थ	24	0.38	(0.26, 0.51)	0.22
Ya	य	5	0.08	(0.03, 0.18)	0.15

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Pha	फ	43	0.68	(0.55, 0.79)	0.64
Tha	थ	19	0.30	(0.19, 0.43)	0.20
Ya	य	1	0.02	(0.0, 0.09)	0.15

Lohit Devanagari (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	3	0.05	(0.01, 0.13)	0.15
Ddha	ढ	3	0.05	(0.01, 0.13)	0.10
Nna	ण	56	0.90	(0.8, 0.96)	0.75

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	6	0.10	(0.04, 0.2)	0.22
Ddha	ढ	5	0.08	(0.03, 0.18)	0.16
Sha	श	51	0.82	(0.7, 0.91)	0.62

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	5	0.08	(0.03, 0.18)	0.16
Ddha	ढ	2	0.03	(0.0, 0.11)	0.11
Ssa	ष	55	0.89	(0.78, 0.95)	0.73

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	4	0.06	(0.02, 0.16)	0.18
Ddha	ढ	10	0.16	(0.08, 0.28)	0.21
Ta	त	48	0.77	(0.65, 0.87)	0.60

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	4	0.06	(0.02, 0.16)	0.27
Ddha	ढ	9	0.15	(0.07, 0.26)	0.06
Tha	थ	49	0.79	(0.67, 0.88)	0.67

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	3	0.05	(0.01, 0.13)	0.17
Ddha	ढ	9	0.15	(0.07, 0.26)	0.13
Ya	य	50	0.81	(0.69, 0.9)	0.70

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	40	0.65	(0.51, 0.76)	0.57
Nna	ण	12	0.19	(0.1, 0.31)	0.32
Sha	श	10	0.16	(0.08, 0.28)	0.11

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	47	0.76	(0.63, 0.86)	0.80
Nna	ण	9	0.15	(0.07, 0.26)	0.14
Ssa	ष	6	0.10	(0.04, 0.2)	0.06

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	15	0.24	(0.14, 0.37)	0.34
Nna	ण	46	0.74	(0.62, 0.84)	0.52
Ta	त	1	0.02	(0.0, 0.09)	0.14

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	23	0.37	(0.25, 0.5)	0.46
Nna	ण	36	0.58	(0.45, 0.7)	0.31
Tha	थ	3	0.05	(0.01, 0.13)	0.24

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	29	0.47	(0.34, 0.6)	0.49
Nna	ण	29	0.47	(0.34, 0.6)	0.33
Ya	य	4	0.06	(0.02, 0.16)	0.18

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	32	0.52	(0.39, 0.65)	0.44
Sha	श	14	0.23	(0.13, 0.35)	0.21
Ssa	ष	16	0.26	(0.16, 0.38)	0.35

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	11	0.18	(0.09, 0.3)	0.31
Sha	श	46	0.74	(0.62, 0.84)	0.42
Ta	त	5	0.08	(0.03, 0.18)	0.27

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	34	0.55	(0.42, 0.68)	0.56
Sha	श	23	0.37	(0.25, 0.5)	0.14
Tha	थ	5	0.08	(0.03, 0.18)	0.30

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	32	0.52	(0.39, 0.65)	0.47
Sha	श	21	0.34	(0.22, 0.47)	0.22
Ya	य	9	0.15	(0.07, 0.26)	0.31

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	20	0.32	(0.21, 0.45)	0.37
Ssa	ष	38	0.61	(0.48, 0.73)	0.47
Ta	त	4	0.06	(0.02, 0.16)	0.15

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	37	0.60	(0.46, 0.72)	0.50
Ssa	ष	20	0.32	(0.21, 0.45)	0.24
Tha	थ	5	0.08	(0.03, 0.18)	0.26

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	42	0.68	(0.55, 0.79)	0.53
Ssa	ष	16	0.26	(0.16, 0.38)	0.27
Ya	य	4	0.06	(0.02, 0.16)	0.20

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	16	0.26	(0.16, 0.38)	0.37
Ta	त	14	0.23	(0.13, 0.35)	0.15
Tha	थ	32	0.52	(0.39, 0.65)	0.48

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	24	0.39	(0.27, 0.52)	0.40
Ta	त	11	0.18	(0.09, 0.3)	0.17
Ya	य	27	0.44	(0.31, 0.57)	0.43

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	51	0.82	(0.7, 0.91)	0.80
Tha	थ	5	0.08	(0.03, 0.18)	0.14
Ya	य	6	0.10	(0.04, 0.2)	0.06

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	39	0.63	(0.5, 0.75)	0.54
Nna	ण	15	0.24	(0.14, 0.37)	0.35
Sha	श	8	0.13	(0.06, 0.24)	0.12

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	52	0.84	(0.72, 0.92)	0.77
Nna	ण	3	0.05	(0.01, 0.13)	0.16
Ssa	ष	7	0.11	(0.05, 0.22)	0.07

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	18	0.29	(0.18, 0.42)	0.33
Nna	ण	36	0.58	(0.45, 0.7)	0.48
Ta	त	8	0.13	(0.06, 0.24)	0.19

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	20	0.32	(0.21, 0.45)	0.36
Nna	ण	35	0.56	(0.43, 0.69)	0.44
Tha	थ	7	0.11	(0.05, 0.22)	0.19

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	34	0.55	(0.42, 0.68)	0.45
Nna	ण	22	0.35	(0.24, 0.49)	0.35
Ya	य	6	0.10	(0.04, 0.2)	0.20

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	37	0.60	(0.46, 0.72)	0.40
Sha	श	14	0.23	(0.13, 0.35)	0.22
Ssa	ष	11	0.18	(0.09, 0.3)	0.38

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	12	0.19	(0.1, 0.31)	0.30
Sha	श	38	0.61	(0.48, 0.73)	0.37
Ta	त	12	0.19	(0.1, 0.31)	0.33

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	38	0.61	(0.48, 0.73)	0.46
Sha	श	18	0.29	(0.18, 0.42)	0.28
Tha	थ	6	0.10	(0.04, 0.2)	0.26

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	30	0.48	(0.35, 0.61)	0.43
Sha	श	20	0.32	(0.21, 0.45)	0.23
Ya	य	12	0.19	(0.1, 0.31)	0.34

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	30	0.48	(0.35, 0.61)	0.37
Ssa	ष	28	0.45	(0.32, 0.58)	0.43
Ta	त	4	0.06	(0.02, 0.16)	0.21

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	32	0.52	(0.39, 0.65)	0.40
Ssa	ष	23	0.37	(0.25, 0.5)	0.39
Tha	थ	7	0.11	(0.05, 0.22)	0.21

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	43	0.69	(0.56, 0.8)	0.50
Ssa	ष	11	0.18	(0.09, 0.3)	0.28
Ya	य	8	0.13	(0.06, 0.24)	0.22

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	26	0.42	(0.3, 0.55)	0.29
Ta	त	13	0.21	(0.12, 0.33)	0.34
Tha	थ	23	0.37	(0.25, 0.5)	0.37

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	37	0.60	(0.46, 0.72)	0.39
Ta	त	5	0.08	(0.03, 0.18)	0.23
Ya	य	20	0.32	(0.21, 0.45)	0.39

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ddha	ढ	49	0.79	(0.67, 0.88)	0.69
Tha	थ	5	0.08	(0.03, 0.18)	0.09
Ya	य	8	0.13	(0.06, 0.24)	0.22

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nna	ण	8	0.13	(0.06, 0.24)	0.10
Sha	श	45	0.73	(0.6, 0.83)	0.59
Ssa	ष	9	0.15	(0.07, 0.26)	0.30

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nna	ण	7	0.11	(0.05, 0.22)	0.27
Sha	श	21	0.34	(0.22, 0.47)	0.21
Ta	त	34	0.55	(0.42, 0.68)	0.52

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nna	ण	30	0.48	(0.35, 0.61)	0.41
Sha	श	10	0.16	(0.08, 0.28)	0.15
Tha	थ	22	0.35	(0.24, 0.49)	0.44

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nna	ण	25	0.40	(0.28, 0.54)	0.31
Sha	श	18	0.29	(0.18, 0.42)	0.23
Ya	य	19	0.31	(0.2, 0.44)	0.45

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nna	ण	12	0.19	(0.1, 0.31)	0.19
Ssa	ष	12	0.19	(0.1, 0.31)	0.10
Ta	त	38	0.61	(0.48, 0.73)	0.72

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nna	ण	13	0.21	(0.12, 0.33)	0.19
Ssa	ष	6	0.10	(0.04, 0.2)	0.10
Tha	थ	43	0.69	(0.56, 0.8)	0.71

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nna	ण	25	0.40	(0.28, 0.54)	0.22
Ssa	ष	4	0.06	(0.02, 0.16)	0.12
Ya	य	33	0.53	(0.4, 0.66)	0.66

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nna	ण	38	0.61	(0.48, 0.73)	0.37
Ta	त	14	0.23	(0.13, 0.35)	0.33
Tha	थ	10	0.16	(0.08, 0.28)	0.30

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nna	ण	37	0.60	(0.46, 0.72)	0.40
Ta	त	20	0.32	(0.21, 0.45)	0.36
Ya	य	5	0.08	(0.03, 0.18)	0.24

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nna	ण	52	0.84	(0.72, 0.92)	0.71
Tha	थ	5	0.08	(0.03, 0.18)	0.19
Ya	य	5	0.08	(0.03, 0.18)	0.10

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Sha	श	27	0.44	(0.31, 0.57)	0.33
Ssa	ष	10	0.16	(0.08, 0.28)	0.31
Ta	त	25	0.40	(0.28, 0.54)	0.36

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Sha	श	21	0.34	(0.22, 0.47)	0.27
Ssa	ष	29	0.47	(0.34, 0.6)	0.46
Tha	थ	12	0.19	(0.1, 0.31)	0.27

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Sha	श	37	0.60	(0.46, 0.72)	0.37
Ssa	ष	15	0.24	(0.14, 0.37)	0.35
Ya	य	10	0.16	(0.08, 0.28)	0.28

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Sha	श	24	0.39	(0.27, 0.52)	0.25
Ta	त	33	0.53	(0.4, 0.66)	0.50
Tha	थ	5	0.08	(0.03, 0.18)	0.25

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Sha	श	31	0.50	(0.37, 0.63)	0.35
Ta	त	24	0.39	(0.27, 0.52)	0.39
Ya	य	7	0.11	(0.05, 0.22)	0.26

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Sha	श	45	0.73	(0.6, 0.83)	0.59
Tha	थ	5	0.08	(0.03, 0.18)	0.14
Ya	य	12	0.19	(0.1, 0.31)	0.27

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ssa	ष	21	0.34	(0.22, 0.47)	0.30
Ta	त	21	0.34	(0.22, 0.47)	0.37
Tha	थ	20	0.32	(0.21, 0.45)	0.33

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ssa	ष	15	0.24	(0.14, 0.37)	0.34
Ta	त	43	0.69	(0.56, 0.8)	0.40
Ya	य	4	0.06	(0.02, 0.16)	0.27

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ssa	ष	48	0.77	(0.65, 0.87)	0.67
Tha	थ	11	0.18	(0.09, 0.3)	0.21
Ya	य	3	0.05	(0.01, 0.13)	0.12

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Ta	त	48	0.77	(0.65, 0.87)	0.70
Tha	थ	9	0.15	(0.07, 0.26)	0.19
Ya	य	5	0.08	(0.03, 0.18)	0.10

Murty Hindi (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	10	0.20	(0.1, 0.33)	0.18
Da	द	33	0.65	(0.5, 0.78)	0.72
Dha	ध	8	0.16	(0.07, 0.29)	0.10

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	14	0.27	(0.16, 0.42)	0.29
Da	द	30	0.59	(0.44, 0.72)	0.50
La	ल	7	0.14	(0.06, 0.26)	0.21

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	22	0.43	(0.29, 0.58)	0.54
Da	द	10	0.20	(0.1, 0.33)	0.21
Nga	ङ	19	0.37	(0.24, 0.52)	0.24

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	42	0.82	(0.69, 0.92)	0.77
Da	द	3	0.06	(0.01, 0.16)	0.14
Tta	ट	6	0.12	(0.04, 0.24)	0.08

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	37	0.73	(0.58, 0.84)	0.71
Da	द	4	0.08	(0.02, 0.19)	0.22
Ttha	ठ	10	0.20	(0.1, 0.33)	0.07

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	13	0.25	(0.14, 0.4)	0.24
Da	द	25	0.49	(0.35, 0.63)	0.57
U	उ	13	0.25	(0.14, 0.4)	0.18

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	14	0.27	(0.16, 0.42)	0.31
Dha	ध	8	0.16	(0.07, 0.29)	0.13
La	ल	29	0.57	(0.42, 0.71)	0.56

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	5	0.10	(0.03, 0.21)	0.18
Dha	ध	13	0.25	(0.14, 0.4)	0.11
Nga	ङ	33	0.65	(0.5, 0.78)	0.72

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	13	0.25	(0.14, 0.4)	0.20
Dha	ध	4	0.08	(0.02, 0.19)	0.11
Tta	ट	34	0.67	(0.52, 0.79)	0.69

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	3	0.06	(0.01, 0.16)	0.15
Dha	ध	9	0.18	(0.08, 0.31)	0.21
Ttha	ठ	39	0.76	(0.63, 0.87)	0.64

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	9	0.18	(0.08, 0.31)	0.30
Dha	ध	10	0.20	(0.1, 0.33)	0.20
U	उ	32	0.63	(0.48, 0.76)	0.50

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	9	0.18	(0.08, 0.31)	0.27
La	ल	10	0.20	(0.1, 0.33)	0.22
Nga	ङ	32	0.63	(0.48, 0.76)	0.52

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	19	0.37	(0.24, 0.52)	0.31
La	ल	6	0.12	(0.04, 0.24)	0.23
Tta	ट	26	0.51	(0.37, 0.65)	0.46

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	7	0.14	(0.06, 0.26)	0.25
La	ल	7	0.14	(0.06, 0.26)	0.32
Ttha	ठ	37	0.73	(0.58, 0.84)	0.42

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	26	0.51	(0.37, 0.65)	0.38
La	ल	15	0.29	(0.17, 0.44)	0.42
U	उ	10	0.20	(0.1, 0.33)	0.19

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	26	0.51	(0.37, 0.65)	0.57
Nga	ङ	14	0.27	(0.16, 0.42)	0.26
Tta	ट	11	0.22	(0.11, 0.35)	0.17

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	14	0.27	(0.16, 0.42)	0.52
Nga	ङ	19	0.37	(0.24, 0.52)	0.32
Ttha	ठ	18	0.35	(0.22, 0.5)	0.16

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	9	0.18	(0.08, 0.31)	0.26
Nga	ङ	32	0.63	(0.48, 0.76)	0.56
U	उ	10	0.20	(0.1, 0.33)	0.18

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	42	0.82	(0.69, 0.92)	0.75
Tta	ट	4	0.08	(0.02, 0.19)	0.17
Ttha	ठ	5	0.10	(0.03, 0.21)	0.08

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	15	0.29	(0.17, 0.44)	0.27
Tta	ट	30	0.59	(0.44, 0.72)	0.54
U	उ	6	0.12	(0.04, 0.24)	0.20

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Cha	छ	10	0.20	(0.1, 0.33)	0.21
Ttha	ठ	33	0.65	(0.5, 0.78)	0.50
U	उ	8	0.16	(0.07, 0.29)	0.29

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	26	0.51	(0.37, 0.65)	0.56
Dha	ध	13	0.25	(0.14, 0.4)	0.22
La	ल	12	0.24	(0.13, 0.37)	0.22

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	8	0.16	(0.07, 0.29)	0.23
Dha	ध	26	0.51	(0.37, 0.65)	0.51
Nga	ङ	17	0.33	(0.21, 0.48)	0.26

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	2	0.04	(0.0, 0.13)	0.16
Dha	ध	44	0.86	(0.74, 0.94)	0.75
Tta	ट	5	0.10	(0.03, 0.21)	0.09

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	2	0.04	(0.0, 0.13)	0.16
Dha	ध	34	0.67	(0.52, 0.79)	0.72
Ttha	ठ	15	0.29	(0.17, 0.44)	0.12

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	25	0.49	(0.35, 0.63)	0.62
Dha	ध	13	0.25	(0.14, 0.4)	0.20
U	उ	13	0.25	(0.14, 0.4)	0.18

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	10	0.20	(0.1, 0.33)	0.23
La	ल	16	0.31	(0.19, 0.46)	0.51
Nga	ङ	25	0.49	(0.35, 0.63)	0.27

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	3	0.06	(0.01, 0.16)	0.17
La	ल	41	0.80	(0.67, 0.9)	0.73
Tta	ट	7	0.14	(0.06, 0.26)	0.10

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	2	0.04	(0.0, 0.13)	0.17
La	ल	36	0.71	(0.56, 0.83)	0.70
Ttha	ठ	13	0.25	(0.14, 0.4)	0.13

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	36	0.71	(0.56, 0.83)	0.55
La	ल	3	0.06	(0.01, 0.16)	0.24
U	उ	12	0.24	(0.13, 0.37)	0.21

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	4	0.08	(0.02, 0.19)	0.27
Nga	ङ	42	0.82	(0.69, 0.92)	0.52
Tta	ट	5	0.10	(0.03, 0.21)	0.21

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	3	0.06	(0.01, 0.16)	0.27
Nga	ङ	36	0.71	(0.56, 0.83)	0.50
Ttha	ठ	12	0.24	(0.13, 0.37)	0.23

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	19	0.37	(0.24, 0.52)	0.24
Nga	ङ	13	0.25	(0.14, 0.4)	0.23
U	उ	19	0.37	(0.24, 0.52)	0.53

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	22	0.43	(0.29, 0.58)	0.37
Tta	ट	2	0.04	(0.0, 0.13)	0.30
Ttha	ठ	27	0.53	(0.38, 0.67)	0.33

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	4	0.08	(0.02, 0.19)	0.14
Tta	ट	4	0.08	(0.02, 0.19)	0.08
U	उ	43	0.84	(0.71, 0.93)	0.78

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Da	द	4	0.08	(0.02, 0.19)	0.15
Ttha	ठ	13	0.25	(0.14, 0.4)	0.11
U	उ	34	0.67	(0.52, 0.79)	0.74

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dha	ध	11	0.22	(0.11, 0.35)	0.20
La	ल	11	0.22	(0.11, 0.35)	0.22
Nga	ङ	29	0.57	(0.42, 0.71)	0.58

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dha	ध	15	0.29	(0.17, 0.44)	0.24
La	ल	7	0.14	(0.06, 0.26)	0.24
Tta	ट	29	0.57	(0.42, 0.71)	0.52

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dha	ध	10	0.20	(0.1, 0.33)	0.23
La	ल	7	0.14	(0.06, 0.26)	0.24
Ttha	ठ	34	0.67	(0.52, 0.79)	0.53

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dha	ध	28	0.55	(0.4, 0.69)	0.26
La	ल	13	0.25	(0.14, 0.4)	0.38
U	उ	10	0.20	(0.1, 0.33)	0.36

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dha	ध	27	0.53	(0.38, 0.67)	0.54
Nga	ङ	18	0.35	(0.22, 0.5)	0.29
Tta	ट	6	0.12	(0.04, 0.24)	0.18

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dha	ध	21	0.41	(0.28, 0.56)	0.52
Nga	ङ	15	0.29	(0.17, 0.44)	0.28
Ttha	ठ	15	0.29	(0.17, 0.44)	0.20

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dha	ध	20	0.39	(0.26, 0.54)	0.23
Nga	ङ	23	0.45	(0.31, 0.6)	0.61
U	उ	8	0.16	(0.07, 0.29)	0.16

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dha	ध	48	0.94	–	0.76
Tta	ट	–	–	–	0.11
Ttha	ठ	3	0.06	–	0.13

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dha	ध	9	0.18	(0.08, 0.31)	0.22
Tta	ट	28	0.55	(0.4, 0.69)	0.58
U	उ	14	0.27	(0.16, 0.42)	0.20

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Dha	ध	14	0.27	(0.16, 0.42)	0.22
Ttha	ठ	28	0.55	(0.4, 0.69)	0.59
U	उ	9	0.18	(0.08, 0.31)	0.20

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
La	ल	19	0.37	(0.24, 0.52)	0.53
Nga	ङ	18	0.35	(0.22, 0.5)	0.29
Tta	ट	14	0.27	(0.16, 0.42)	0.18

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
La	ल	17	0.33	(0.21, 0.48)	0.52
Nga	ङ	16	0.31	(0.19, 0.46)	0.28
Ttha	ठ	18	0.35	(0.22, 0.5)	0.20

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
La	ल	10	0.20	(0.1, 0.33)	0.27
Nga	ङ	36	0.71	(0.56, 0.83)	0.55
U	उ	5	0.10	(0.03, 0.21)	0.18

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
La	ल	44	0.86	(0.74, 0.94)	0.75
Tta	ट	1	0.02	(0.0, 0.1)	0.11
Ttha	ठ	6	0.12	(0.04, 0.24)	0.14

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
La	ल	8	0.16	(0.07, 0.29)	0.26
Tta	ट	30	0.59	(0.44, 0.72)	0.51
U	उ	13	0.25	(0.14, 0.4)	0.23

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
La	ल	6	0.12	(0.04, 0.24)	0.26
Ttha	ठ	34	0.67	(0.52, 0.79)	0.52
U	उ	11	0.22	(0.11, 0.35)	0.22

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nga	ङ	43	0.84	(0.71, 0.93)	0.53
Tta	ट	1	0.02	(0.0, 0.1)	0.23
Ttha	ठ	7	0.14	(0.06, 0.26)	0.24

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nga	ङ	14	0.27	(0.16, 0.42)	0.25
Tta	ट	19	0.37	(0.24, 0.52)	0.19
U	उ	18	0.35	(0.22, 0.5)	0.56

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Nga	ङ	14	0.27	(0.16, 0.42)	0.25
Ttha	ठ	27	0.53	(0.38, 0.67)	0.21
U	उ	10	0.20	(0.1, 0.33)	0.54

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
Tta	ट	3	0.06	(0.01, 0.16)	0.09
Ttha	ठ	7	0.14	(0.06, 0.26)	0.11
U	उ	41	0.80	(0.67, 0.9)	0.79

Nirmala UI (Devanagari)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	1	0.01	(0.0, 0.08)	0.05
e	e	1	0.01	(0.0, 0.08)	0.15
l	l	65	0.97	(0.9, 1.0)	0.80

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	12	0.18	(0.1, 0.29)	0.30
e	e	3	0.04	(0.01, 0.13)	0.24
p	p	52	0.78	(0.66, 0.87)	0.46

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	7	0.10	(0.04, 0.2)	0.28
e	e	12	0.18	(0.1, 0.29)	0.21
s	s	48	0.72	(0.59, 0.82)	0.51

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	1	0.01	(0.0, 0.08)	0.08
e	e	1	0.01	(0.0, 0.08)	0.21
t	t	65	0.97	(0.9, 1.0)	0.71

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	3	0.04	(0.01, 0.13)	0.18
e	e	2	0.03	(0.0, 0.1)	0.09
y	y	62	0.93	(0.83, 0.98)	0.73

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	2	0.03	(0.0, 0.1)	0.24
e	e	1	0.01	(0.0, 0.08)	0.13
z	z	64	0.96	(0.87, 0.99)	0.63

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	14	0.21	(0.12, 0.33)	0.24
l	l	51	0.76	(0.64, 0.86)	0.61
p	p	2	0.03	(0.0, 0.1)	0.15

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	7	0.10	–	0.33
l	l	60	0.90	–	0.50
s	s	–	–	–	0.18

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	56	0.84	(0.73, 0.92)	0.81
l	l	8	0.12	(0.05, 0.22)	0.14
t	t	3	0.04	(0.01, 0.13)	0.05

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	21	0.31	(0.21, 0.44)	0.32
l	l	39	0.58	(0.46, 0.7)	0.35
y	y	7	0.10	(0.04, 0.2)	0.33

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	24	0.36	(0.24, 0.48)	0.30
l	l	39	0.58	(0.46, 0.7)	0.47
z	z	4	0.06	(0.02, 0.15)	0.23

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	2	0.03	(0.0, 0.1)	0.11
p	p	52	0.78	(0.66, 0.87)	0.39
s	s	13	0.19	(0.11, 0.31)	0.50

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	12	0.18	(0.1, 0.29)	0.27
p	p	5	0.07	(0.02, 0.17)	0.19
t	t	50	0.75	(0.63, 0.84)	0.54

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	34	0.51	(0.38, 0.63)	0.37
p	p	3	0.04	(0.01, 0.13)	0.09
y	y	30	0.45	(0.33, 0.57)	0.54

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	3	0.04	(0.01, 0.13)	0.05
p	p	22	0.33	(0.22, 0.45)	0.35
z	z	42	0.63	(0.5, 0.74)	0.60

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	4	0.06	(0.02, 0.15)	0.37
s	s	1	0.01	(0.0, 0.08)	0.24
t	t	62	0.93	(0.83, 0.98)	0.39

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	6	0.09	(0.03, 0.18)	0.33
s	s	2	0.03	(0.0, 0.1)	0.17
y	y	59	0.88	(0.78, 0.95)	0.50

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	26	0.39	(0.27, 0.51)	0.40
s	s	2	0.03	(0.0, 0.1)	0.22
z	z	39	0.58	(0.46, 0.7)	0.39

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	37	0.55	(0.43, 0.67)	0.34
t	t	19	0.28	(0.18, 0.41)	0.27
y	y	11	0.16	(0.08, 0.27)	0.39

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	14	0.21	(0.12, 0.33)	0.17
t	t	44	0.66	(0.53, 0.77)	0.47
z	z	9	0.13	(0.06, 0.24)	0.36

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	28	0.42	(0.3, 0.54)	0.49
y	y	31	0.46	(0.34, 0.59)	0.39
z	z	8	0.12	(0.05, 0.22)	0.12

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	11	0.16	(0.08, 0.27)	0.28
l	l	51	0.76	(0.64, 0.86)	0.66
p	p	5	0.07	(0.02, 0.17)	0.06

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	5	0.07	–	0.34
l	l	62	0.93	–	0.57
s	s	–	–	–	0.09

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	57	0.85	(0.74, 0.93)	0.88
l	l	3	0.04	(0.01, 0.13)	0.12
t	t	7	0.10	(0.04, 0.2)	0.00

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	21	0.31	(0.21, 0.44)	0.32
l	l	39	0.58	(0.46, 0.7)	0.47
y	y	7	0.10	(0.04, 0.2)	0.21

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	23	0.34	(0.23, 0.47)	0.31
l	l	40	0.60	(0.47, 0.72)	0.58
z	z	4	0.06	(0.02, 0.15)	0.11

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	2	0.03	(0.0, 0.1)	0.07
p	p	50	0.75	(0.63, 0.84)	0.42
s	s	15	0.22	(0.13, 0.34)	0.51

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	15	0.22	(0.13, 0.34)	0.30
p	p	6	0.09	(0.03, 0.18)	0.11
t	t	46	0.69	(0.56, 0.79)	0.60

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	33	0.49	(0.37, 0.62)	0.31
p	p	4	0.06	(0.02, 0.15)	0.14
y	y	30	0.45	(0.33, 0.57)	0.55

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	2	0.03	(0.0, 0.1)	0.00
p	p	25	0.37	(0.26, 0.5)	0.39
z	z	40	0.60	(0.47, 0.72)	0.61

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	3	0.04	(0.01, 0.13)	0.38
s	s	1	0.01	(0.0, 0.08)	0.13
t	t	63	0.94	(0.85, 0.98)	0.48

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	1	0.01	(0.0, 0.08)	0.26
s	s	3	0.04	(0.01, 0.13)	0.20
y	y	63	0.94	(0.85, 0.98)	0.54

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	25	0.37	(0.26, 0.5)	0.33
s	s	2	0.03	(0.0, 0.1)	0.26
z	z	40	0.60	(0.47, 0.72)	0.42

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	32	0.48	(0.35, 0.6)	0.35
t	t	22	0.33	(0.22, 0.45)	0.39
y	y	13	0.19	(0.11, 0.31)	0.26

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	9	0.13	(0.06, 0.24)	0.17
t	t	49	0.73	(0.61, 0.83)	0.61
z	z	9	0.13	(0.06, 0.24)	0.22

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	29	0.43	(0.31, 0.56)	0.42
y	y	32	0.48	(0.35, 0.6)	0.45
z	z	6	0.09	(0.03, 0.18)	0.14

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
l	l	36	0.54	(0.41, 0.66)	0.27
p	p	9	0.13	(0.06, 0.24)	0.38
s	s	22	0.33	(0.22, 0.45)	0.35

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
l	l	7	0.10	(0.04, 0.2)	0.18
p	p	55	0.82	(0.71, 0.9)	0.73
t	t	5	0.07	(0.02, 0.17)	0.09

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
l	l	46	0.69	(0.56, 0.79)	0.50
p	p	10	0.15	(0.07, 0.26)	0.16
y	y	11	0.16	(0.08, 0.27)	0.33

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
l	l	26	0.39	(0.27, 0.51)	0.21
p	p	19	0.28	(0.18, 0.41)	0.36
z	z	22	0.33	(0.22, 0.45)	0.43

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
l	l	6	0.09	(0.03, 0.18)	0.21
s	s	51	0.76	(0.64, 0.86)	0.69
t	t	10	0.15	(0.07, 0.26)	0.11

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
l	l	27	0.40	(0.28, 0.53)	0.40
s	s	28	0.42	(0.3, 0.54)	0.17
y	y	12	0.18	(0.1, 0.29)	0.43

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
l	l	58	0.87	(0.76, 0.94)	0.54
s	s	8	0.12	(0.05, 0.22)	0.17
z	z	1	0.01	(0.0, 0.08)	0.30

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
l	l	11	0.16	(0.08, 0.27)	0.27
t	t	4	0.06	(0.02, 0.15)	0.01
y	y	52	0.78	(0.66, 0.87)	0.72

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
l	l	13	0.19	(0.11, 0.31)	0.02
t	t	9	0.13	(0.06, 0.24)	0.15
z	z	45	0.67	(0.55, 0.78)	0.82

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
l	l	37	0.55	(0.43, 0.67)	0.57
y	y	17	0.25	(0.16, 0.37)	0.27
z	z	13	0.19	(0.11, 0.31)	0.16

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
p	p	8	0.12	(0.05, 0.22)	0.45
s	s	22	0.33	(0.22, 0.45)	0.38
t	t	37	0.55	(0.43, 0.67)	0.17

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
p	p	9	0.13	(0.06, 0.24)	0.37
s	s	43	0.64	(0.52, 0.76)	0.50
y	y	15	0.22	(0.13, 0.34)	0.13

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
p	p	48	0.72	(0.59, 0.82)	0.61
s	s	4	0.06	(0.02, 0.15)	0.13
z	z	15	0.22	(0.13, 0.34)	0.26

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
p	p	18	0.27	(0.17, 0.39)	0.22
t	t	43	0.64	(0.52, 0.76)	0.44
y	y	6	0.09	(0.03, 0.18)	0.34

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
p	p	20	0.30	(0.19, 0.42)	0.25
t	t	25	0.37	(0.26, 0.5)	0.21
z	z	22	0.33	(0.22, 0.45)	0.54

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
p	p	23	0.34	(0.23, 0.47)	0.51
y	y	7	0.10	(0.04, 0.2)	0.00
z	z	37	0.55	(0.43, 0.67)	0.49

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
s	s	38	0.57	(0.44, 0.69)	0.21
t	t	21	0.31	(0.21, 0.44)	0.32
y	y	8	0.12	(0.05, 0.22)	0.47

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
s	s	10	0.15	(0.07, 0.26)	0.02
t	t	53	0.79	(0.67, 0.88)	0.57
z	z	4	0.06	(0.02, 0.15)	0.42

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
s	s	7	0.10	(0.04, 0.2)	0.32
y	y	53	0.79	(0.67, 0.88)	0.50
z	z	7	0.10	(0.04, 0.2)	0.18

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
t	t	36	0.54	(0.41, 0.66)	0.49
y	y	8	0.12	(0.05, 0.22)	0.09
z	z	23	0.34	(0.23, 0.47)	0.42

Arial (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	74	0.96	(0.89, 0.99)	0.66
i	i	1	0.01	(0.0, 0.07)	0.20
j	j	2	0.03	(0.0, 0.09)	0.14

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	5	0.06	(0.02, 0.15)	0.03
i	i	66	0.86	(0.76, 0.93)	0.74
o	o	6	0.08	(0.03, 0.16)	0.24

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	2	0.03	–	0.04
i	i	75	0.97	–	0.78
p	p	–	–	–	0.18

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	43	0.56	(0.44, 0.67)	0.50
i	i	31	0.40	(0.29, 0.52)	0.34
r	r	3	0.04	(0.01, 0.11)	0.16

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	26	0.34	(0.23, 0.45)	0.34
i	i	26	0.34	(0.23, 0.45)	0.27
v	v	25	0.32	(0.22, 0.44)	0.39

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	22	0.29	(0.19, 0.4)	0.30
i	i	15	0.19	(0.11, 0.3)	0.24
x	x	40	0.52	(0.4, 0.63)	0.46

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	5	0.06	(0.02, 0.15)	0.03
j	j	66	0.86	(0.76, 0.93)	0.69
o	o	6	0.08	(0.03, 0.16)	0.28

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	4	0.05	(0.01, 0.13)	0.20
j	j	72	0.94	(0.85, 0.98)	0.65
p	p	1	0.01	(0.0, 0.07)	0.15

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	31	0.40	(0.29, 0.52)	0.50
j	j	35	0.45	(0.34, 0.57)	0.30
r	r	11	0.14	(0.07, 0.24)	0.20

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	17	0.22	(0.13, 0.33)	0.15
j	j	23	0.30	(0.2, 0.41)	0.33
v	v	37	0.48	(0.37, 0.6)	0.51

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	13	0.17	(0.09, 0.27)	0.14
j	j	17	0.22	(0.13, 0.33)	0.30
x	x	47	0.61	(0.49, 0.72)	0.57

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	6	0.08	(0.03, 0.16)	0.28
o	o	68	0.88	(0.79, 0.95)	0.44
p	p	3	0.04	(0.01, 0.11)	0.28

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	8	0.10	(0.05, 0.19)	0.11
o	o	8	0.10	(0.05, 0.19)	0.25
r	r	61	0.79	(0.68, 0.88)	0.64

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	11	0.14	–	0.23
o	o	–	–	–	0.00
v	v	66	0.86	–	0.77

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	4	0.05	–	0.20
o	o	–	–	–	0.00
x	x	73	0.95	–	0.80

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	6	0.08	(0.03, 0.16)	0.16
p	p	1	0.01	(0.0, 0.07)	0.17
r	r	70	0.91	(0.82, 0.96)	0.68

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	3	0.04	(0.01, 0.11)	0.06
p	p	1	0.01	(0.0, 0.07)	0.07
v	v	73	0.95	(0.87, 0.99)	0.87

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	1	0.01	(0.0, 0.07)	0.06
p	p	1	0.01	(0.0, 0.07)	0.06
x	x	75	0.97	(0.91, 1.0)	0.88

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	27	0.35	(0.25, 0.47)	0.29
r	r	8	0.10	(0.05, 0.19)	0.17
v	v	42	0.55	(0.43, 0.66)	0.54

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	24	0.31	(0.21, 0.43)	0.25
r	r	3	0.04	(0.01, 0.11)	0.15
x	x	50	0.65	(0.53, 0.75)	0.60

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	67	0.87	(0.77, 0.94)	0.85
v	v	1	0.01	(0.0, 0.07)	0.00
x	x	9	0.12	(0.05, 0.21)	0.15

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	1	0.01	(0.0, 0.07)	0.08
j	j	2	0.03	(0.0, 0.09)	0.08
o	o	74	0.96	(0.89, 0.99)	0.83

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	4	0.05	(0.01, 0.13)	0.27
j	j	1	0.01	(0.0, 0.07)	0.05
p	p	72	0.94	(0.85, 0.98)	0.68

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	3	0.04	(0.01, 0.11)	0.36
j	j	5	0.06	(0.02, 0.15)	0.19
r	r	69	0.90	(0.81, 0.95)	0.45

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	–	–	–	0.00
j	j	3	0.04	–	0.20
v	v	74	0.96	–	0.80

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	–	–	–	0.00
j	j	3	0.04	–	0.18
x	x	74	0.96	–	0.82

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	70	0.91	(0.82, 0.96)	0.76
o	o	2	0.03	(0.0, 0.09)	0.15
p	p	5	0.06	(0.02, 0.15)	0.09

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	17	0.22	(0.13, 0.33)	0.30
o	o	55	0.71	(0.6, 0.81)	0.67
r	r	5	0.06	(0.02, 0.15)	0.03

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	36	0.47	(0.35, 0.58)	0.51
o	o	33	0.43	(0.32, 0.55)	0.36
v	v	8	0.10	(0.05, 0.19)	0.14

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	34	0.44	(0.33, 0.56)	0.44
o	o	28	0.36	(0.26, 0.48)	0.31
x	x	15	0.19	(0.11, 0.3)	0.25

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	38	0.49	(0.38, 0.61)	0.39
p	p	35	0.45	(0.34, 0.57)	0.55
r	r	4	0.05	(0.01, 0.13)	0.06

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	25	0.32	(0.22, 0.44)	0.32
p	p	30	0.39	(0.28, 0.51)	0.40
v	v	22	0.29	(0.19, 0.4)	0.28

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	19	0.25	(0.16, 0.36)	0.29
p	p	24	0.31	(0.21, 0.43)	0.35
x	x	34	0.44	(0.33, 0.56)	0.36

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	22	0.29	(0.19, 0.4)	0.21
r	r	9	0.12	(0.05, 0.21)	0.18
v	v	46	0.60	(0.48, 0.71)	0.61

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	15	0.19	(0.11, 0.3)	0.18
r	r	3	0.04	(0.01, 0.11)	0.16
x	x	59	0.77	(0.66, 0.86)	0.65

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	68	0.88	(0.79, 0.95)	0.82
v	v	1	0.01	(0.0, 0.07)	0.01
x	x	8	0.10	(0.05, 0.19)	0.17

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
j	j	64	0.83	(0.73, 0.91)	0.66
o	o	9	0.12	(0.05, 0.21)	0.32
p	p	4	0.05	(0.01, 0.13)	0.03

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
j	j	27	0.35	(0.25, 0.47)	0.30
o	o	46	0.60	(0.48, 0.71)	0.66
r	r	4	0.05	(0.01, 0.13)	0.04

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
j	j	42	0.55	(0.43, 0.66)	0.60
o	o	20	0.26	(0.17, 0.37)	0.17
v	v	15	0.19	(0.11, 0.3)	0.23

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
j	j	32	0.42	(0.3, 0.53)	0.52
o	o	20	0.26	(0.17, 0.37)	0.15
x	x	25	0.32	(0.22, 0.44)	0.33

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
j	j	41	0.53	(0.42, 0.65)	0.28
p	p	25	0.32	(0.22, 0.44)	0.48
r	r	11	0.14	(0.07, 0.24)	0.24

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
j	j	19	0.25	(0.16, 0.36)	0.31
p	p	14	0.18	(0.1, 0.29)	0.15
v	v	44	0.57	(0.45, 0.68)	0.54

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
j	j	15	0.19	(0.11, 0.3)	0.28
p	p	8	0.10	(0.05, 0.19)	0.13
x	x	54	0.70	(0.59, 0.8)	0.59

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
j	j	29	0.38	(0.27, 0.49)	0.24
r	r	8	0.10	(0.05, 0.19)	0.00
v	v	40	0.52	(0.4, 0.63)	0.76

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
j	j	20	0.26	(0.17, 0.37)	0.22
r	r	4	0.05	(0.01, 0.13)	0.00
x	x	53	0.69	(0.57, 0.79)	0.78

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
j	j	67	0.87	(0.77, 0.94)	0.88
v	v	2	0.03	(0.0, 0.09)	0.00
x	x	8	0.10	(0.05, 0.19)	0.12

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
o	o	13	0.17	(0.09, 0.27)	0.25
p	p	4	0.05	(0.01, 0.13)	0.12
r	r	60	0.78	(0.67, 0.87)	0.64

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
o	o	1	0.01	(0.0, 0.07)	0.00
p	p	9	0.12	(0.05, 0.21)	0.24
v	v	67	0.87	(0.77, 0.94)	0.76

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
o	o	4	0.05	(0.01, 0.13)	0.00
p	p	6	0.08	(0.03, 0.16)	0.21
x	x	67	0.87	(0.77, 0.94)	0.79

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
o	o	33	0.43	(0.32, 0.55)	0.25
r	r	23	0.30	(0.2, 0.41)	0.43
v	v	21	0.27	(0.18, 0.39)	0.32

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
o	o	26	0.34	(0.23, 0.45)	0.21
r	r	17	0.22	(0.13, 0.33)	0.36
x	x	34	0.44	(0.33, 0.56)	0.43

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
o	o	65	0.84	(0.74, 0.92)	0.78
v	v	3	0.04	(0.01, 0.11)	0.01
x	x	9	0.12	(0.05, 0.21)	0.20

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
p	p	23	0.30	(0.2, 0.41)	0.30
r	r	5	0.06	(0.02, 0.15)	0.17
v	v	49	0.64	(0.52, 0.74)	0.53

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
p	p	20	0.26	(0.17, 0.37)	0.25
r	r	3	0.04	(0.01, 0.11)	0.15
x	x	54	0.70	(0.59, 0.8)	0.60

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
p	p	61	0.79	(0.68, 0.88)	0.85
v	v	4	0.05	(0.01, 0.13)	0.00
x	x	12	0.16	(0.08, 0.26)	0.15

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
r	r	63	0.82	(0.71, 0.9)	0.83
v	v	3	0.04	(0.01, 0.11)	0.00
x	x	11	0.14	(0.07, 0.24)	0.17

Calibri (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	32	0.43	(0.32, 0.55)	0.49
f	f	5	0.07	(0.02, 0.15)	0.24
i	i	37	0.50	(0.38, 0.62)	0.28

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	10	0.14	(0.07, 0.23)	0.26
f	f	26	0.35	(0.24, 0.47)	0.38
m	m	38	0.51	(0.39, 0.63)	0.36

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	11	0.15	(0.08, 0.25)	0.28
f	f	34	0.46	(0.34, 0.58)	0.39
n	n	29	0.39	(0.28, 0.51)	0.33

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	2	0.03	(0.0, 0.09)	0.05
f	f	60	0.81	(0.7, 0.89)	0.71
p	p	12	0.16	(0.09, 0.27)	0.24

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	1	0.01	(0.0, 0.07)	0.10
f	f	67	0.91	(0.81, 0.96)	0.65
q	q	6	0.08	(0.03, 0.17)	0.25

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	6	0.08	(0.03, 0.17)	0.31
f	f	8	0.11	(0.05, 0.2)	0.14
z	z	60	0.81	(0.7, 0.89)	0.56

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	24	0.32	(0.22, 0.44)	0.27
i	i	38	0.51	(0.39, 0.63)	0.42
m	m	12	0.16	(0.09, 0.27)	0.31

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	19	0.26	(0.16, 0.37)	0.30
i	i	51	0.69	(0.57, 0.79)	0.42
n	n	4	0.05	(0.01, 0.13)	0.28

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	4	0.05	(0.01, 0.13)	0.06
i	i	66	0.89	(0.8, 0.95)	0.75
p	p	4	0.05	(0.01, 0.13)	0.19

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	1	0.01	(0.0, 0.07)	0.03
i	i	71	0.96	(0.89, 0.99)	0.72
q	q	2	0.03	(0.0, 0.09)	0.25

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	25	0.34	(0.23, 0.46)	0.26
i	i	13	0.18	(0.1, 0.28)	0.20
z	z	36	0.49	(0.37, 0.61)	0.54

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	73	0.99	-	0.64
m	m	1	0.01	-	0.21
n	n	-	-	-	0.16

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	2	0.03	(0.0, 0.09)	0.20
m	m	70	0.95	(0.87, 0.99)	0.65
p	p	2	0.03	(0.0, 0.09)	0.15

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	1	0.01	(0.0, 0.07)	0.16
m	m	69	0.93	(0.85, 0.98)	0.62
q	q	4	0.05	(0.01, 0.13)	0.21

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	25	0.34	(0.23, 0.46)	0.28
m	m	4	0.05	(0.01, 0.13)	0.13
z	z	45	0.61	(0.49, 0.72)	0.58

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	2	0.03	(0.0, 0.09)	0.20
n	n	70	0.95	(0.87, 0.99)	0.65
p	p	2	0.03	(0.0, 0.09)	0.15

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	–	–	–	0.16
n	n	65	0.88	–	0.62
q	q	9	0.12	–	0.22

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	28	0.38	(0.27, 0.5)	0.31
n	n	2	0.03	(0.0, 0.09)	0.08
z	z	44	0.59	(0.47, 0.71)	0.61

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	54	0.73	(0.61, 0.83)	0.42
p	p	6	0.08	(0.03, 0.17)	0.24
q	q	14	0.19	(0.11, 0.3)	0.34

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	2	0.03	(0.0, 0.09)	0.07
p	p	2	0.03	(0.0, 0.09)	0.03
z	z	70	0.95	(0.87, 0.99)	0.89

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	–	–	–	0.12
q	q	2	0.03	–	0.04
z	z	72	0.97	–	0.85

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	20	0.27	(0.17, 0.39)	0.21
i	i	18	0.24	(0.15, 0.36)	0.24
m	m	36	0.49	(0.37, 0.61)	0.54

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	24	0.32	(0.22, 0.44)	0.22
i	i	19	0.26	(0.16, 0.37)	0.30
n	n	31	0.42	(0.31, 0.54)	0.48

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	16	0.22	(0.13, 0.33)	0.20
i	i	29	0.39	(0.28, 0.51)	0.21
p	p	29	0.39	(0.28, 0.51)	0.59

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	4	0.05	(0.01, 0.13)	0.12
i	i	26	0.35	(0.24, 0.47)	0.26
q	q	44	0.59	(0.47, 0.71)	0.62

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	13	0.18	(0.1, 0.28)	0.10
i	i	9	0.12	(0.06, 0.22)	0.22
z	z	52	0.70	(0.59, 0.8)	0.69

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	74	1.00	–	0.75
m	m	–	–	–	0.17
n	n	–	–	–	0.09

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	43	0.58	(0.46, 0.69)	0.49
m	m	23	0.31	(0.21, 0.43)	0.20
p	p	8	0.11	(0.05, 0.2)	0.31

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	37	0.50	(0.38, 0.62)	0.41
m	m	19	0.26	(0.16, 0.37)	0.26
q	q	18	0.24	(0.15, 0.36)	0.33

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	27	0.36	(0.26, 0.48)	0.26
m	m	3	0.04	(0.01, 0.11)	0.29
z	z	44	0.59	(0.47, 0.71)	0.45

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	43	0.58	(0.46, 0.69)	0.51
n	n	22	0.30	(0.2, 0.41)	0.15
p	p	9	0.12	(0.06, 0.22)	0.34

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	39	0.53	(0.41, 0.64)	0.42
n	n	17	0.23	(0.14, 0.34)	0.22
q	q	18	0.24	(0.15, 0.36)	0.36

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	39	0.53	(0.41, 0.64)	0.26
n	n	3	0.04	(0.01, 0.11)	0.22
z	z	32	0.43	(0.32, 0.55)	0.52

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	72	0.97	–	0.84
p	p	–	–	–	0.10
q	q	2	0.03	–	0.06

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	17	0.23	(0.14, 0.34)	0.23
p	p	4	0.05	(0.01, 0.13)	0.36
z	z	53	0.72	(0.6, 0.81)	0.41

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	8	0.11	(0.05, 0.2)	0.24
q	q	11	0.15	(0.08, 0.25)	0.33
z	z	55	0.74	(0.63, 0.84)	0.43

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	73	0.99	–	0.76
m	m	1	0.01	–	0.16
n	n	–	–	–	0.08

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	42	0.57	(0.45, 0.68)	0.50
m	m	6	0.08	(0.03, 0.17)	0.19
p	p	26	0.35	(0.24, 0.47)	0.30

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	33	0.45	(0.33, 0.57)	0.47
m	m	5	0.07	(0.02, 0.15)	0.16
q	q	36	0.49	(0.37, 0.61)	0.37

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	23	0.31	(0.21, 0.43)	0.32
m	m	13	0.18	(0.1, 0.28)	0.22
z	z	38	0.51	(0.39, 0.63)	0.46

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	49	0.66	(0.54, 0.77)	0.51
n	n	4	0.05	(0.01, 0.13)	0.15
p	p	21	0.28	(0.19, 0.4)	0.34

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	41	0.55	(0.43, 0.67)	0.47
n	n	2	0.03	(0.0, 0.09)	0.12
q	q	31	0.42	(0.31, 0.54)	0.41

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	26	0.35	(0.24, 0.47)	0.33
n	n	5	0.07	(0.02, 0.15)	0.16
z	z	43	0.58	(0.46, 0.69)	0.51

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	72	0.97	–	0.85
p	p	–	–	–	0.03
q	q	2	0.03	–	0.11

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	20	0.27	(0.17, 0.39)	0.28
p	p	19	0.26	(0.16, 0.37)	0.29
z	z	35	0.47	(0.36, 0.59)	0.43

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	14	0.19	(0.11, 0.3)	0.33
q	q	25	0.34	(0.23, 0.46)	0.30
z	z	35	0.47	(0.36, 0.59)	0.36

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	1	0.01	–	0.22
n	n	–	–	–	0.17
p	p	73	0.99	–	0.61

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	4	0.05	–	0.17
n	n	–	–	–	0.12
q	q	70	0.95	–	0.70

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	2	0.03	(0.0, 0.09)	0.08
n	n	1	0.01	(0.0, 0.07)	0.00
z	z	71	0.96	(0.89, 0.99)	0.92

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	69	0.93	(0.85, 0.98)	0.74
p	p	1	0.01	(0.0, 0.07)	0.08
q	q	4	0.05	(0.01, 0.13)	0.18

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	4	0.05	(0.01, 0.13)	0.14
p	p	21	0.28	(0.19, 0.4)	0.25
z	z	49	0.66	(0.54, 0.77)	0.61

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	2	0.03	(0.0, 0.09)	0.20
q	q	28	0.38	(0.27, 0.5)	0.26
z	z	44	0.59	(0.47, 0.71)	0.54

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
n	n	68	0.92	–	0.74
p	p	–	–	–	0.08
q	q	6	0.08	–	0.18

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
n	n	1	0.01	(0.0, 0.07)	0.08
p	p	22	0.30	(0.2, 0.41)	0.27
z	z	51	0.69	(0.57, 0.79)	0.65

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
n	n	2	0.03	(0.0, 0.09)	0.14
q	q	22	0.30	(0.2, 0.41)	0.29
z	z	50	0.68	(0.56, 0.78)	0.57

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
p	p	1	0.01	(0.0, 0.07)	0.04
q	q	2	0.03	(0.0, 0.09)	0.00
z	z	71	0.96	(0.89, 0.99)	0.96

Cambria (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	–	–	–	0.07
d	d	1	0.02	–	0.07
i	i	60	0.98	–	0.86

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	1	0.02	–	0.10
d	d	–	–	–	0.10
j	j	60	0.98	–	0.79

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	1	0.02	(0.0, 0.09)	0.10
d	d	3	0.05	(0.01, 0.14)	0.10
k	k	57	0.93	(0.84, 0.98)	0.81

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	–	–	–	0.10
d	d	3	0.05	–	0.10
l	l	58	0.95	–	0.79

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	1	0.02	–	0.00
d	d	–	–	–	0.00
v	v	60	0.98	–	1.00

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	1	0.02	(0.0, 0.09)	0.03
d	d	1	0.02	(0.0, 0.09)	0.03
y	y	59	0.97	(0.89, 1.0)	0.93

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	60	0.98	–	0.66
i	i	–	–	–	0.20
j	j	1	0.02	–	0.14

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	4	0.07	(0.02, 0.16)	0.35
i	i	49	0.80	(0.68, 0.89)	0.37
k	k	8	0.13	(0.06, 0.24)	0.29

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	49	0.80	–	0.66
i	i	12	0.20	–	0.26
l	l	–	–	–	0.08

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	18	0.30	(0.19, 0.43)	0.34
i	i	24	0.39	(0.27, 0.53)	0.27
v	v	19	0.31	(0.2, 0.44)	0.39

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	16	0.26	(0.16, 0.39)	0.15
i	i	27	0.44	(0.32, 0.58)	0.40
y	y	18	0.30	(0.19, 0.43)	0.45

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	9	0.15	(0.07, 0.26)	0.21
j	j	46	0.75	(0.63, 0.86)	0.41
k	k	6	0.10	(0.04, 0.2)	0.39

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	46	0.75	(0.63, 0.86)	0.47
j	j	14	0.23	(0.13, 0.35)	0.31
l	l	1	0.02	(0.0, 0.09)	0.21

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	22	0.36	(0.24, 0.49)	0.15
j	j	18	0.30	(0.19, 0.43)	0.33
v	v	21	0.34	(0.23, 0.48)	0.51

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	39	0.64	(0.51, 0.76)	0.32
j	j	14	0.23	(0.13, 0.35)	0.29
y	y	8	0.13	(0.06, 0.24)	0.39

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	8	0.13	(0.06, 0.24)	0.41
k	k	10	0.16	(0.08, 0.28)	0.34
l	l	43	0.70	(0.57, 0.81)	0.25

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	21	0.34	(0.23, 0.48)	0.50
k	k	3	0.05	(0.01, 0.14)	0.13
v	v	37	0.61	(0.47, 0.73)	0.37

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	22	0.36	(0.24, 0.49)	0.33
k	k	5	0.08	(0.03, 0.18)	0.25
y	y	34	0.56	(0.42, 0.68)	0.42

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	11	0.18	(0.09, 0.3)	0.36
l	l	19	0.31	(0.2, 0.44)	0.15
v	v	31	0.51	(0.38, 0.64)	0.49

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	6	0.10	(0.04, 0.2)	0.16
l	l	22	0.36	(0.24, 0.49)	0.29
y	y	33	0.54	(0.41, 0.67)	0.55

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	54	0.89	(0.78, 0.95)	0.72
v	v	1	0.02	(0.0, 0.09)	0.18
y	y	6	0.10	(0.04, 0.2)	0.10

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	59	0.97	(0.89, 1.0)	0.66
i	i	1	0.02	(0.0, 0.09)	0.20
j	j	1	0.02	(0.0, 0.09)	0.14

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	11	0.18	(0.09, 0.3)	0.35
i	i	44	0.72	(0.59, 0.83)	0.37
k	k	6	0.10	(0.04, 0.2)	0.29

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	48	0.79	(0.66, 0.88)	0.66
i	i	12	0.20	(0.11, 0.32)	0.26
l	l	1	0.02	(0.0, 0.09)	0.08

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	21	0.34	(0.23, 0.48)	0.34
i	i	22	0.36	(0.24, 0.49)	0.27
v	v	18	0.30	(0.19, 0.43)	0.39

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	10	0.16	(0.08, 0.28)	0.15
i	i	29	0.48	(0.35, 0.61)	0.40
y	y	22	0.36	(0.24, 0.49)	0.44

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	11	0.18	(0.09, 0.3)	0.21
j	j	45	0.74	(0.61, 0.84)	0.41
k	k	5	0.08	(0.03, 0.18)	0.39

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	50	0.82	(0.7, 0.91)	0.47
j	j	7	0.11	(0.05, 0.22)	0.31
l	l	4	0.07	(0.02, 0.16)	0.21

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	16	0.26	(0.16, 0.39)	0.15
j	j	22	0.36	(0.24, 0.49)	0.33
v	v	23	0.38	(0.26, 0.51)	0.51

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	30	0.49	(0.36, 0.62)	0.32
j	j	19	0.31	(0.2, 0.44)	0.29
y	y	12	0.20	(0.11, 0.32)	0.39

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	11	0.18	(0.09, 0.3)	0.41
k	k	12	0.20	(0.11, 0.32)	0.34
l	l	38	0.62	(0.49, 0.74)	0.25

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	27	0.44	(0.32, 0.58)	0.50
k	k	1	0.02	(0.0, 0.09)	0.13
v	v	33	0.54	(0.41, 0.67)	0.37

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	23	0.38	(0.26, 0.51)	0.33
k	k	4	0.07	(0.02, 0.16)	0.25
y	y	34	0.56	(0.42, 0.68)	0.42

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	17	0.28	(0.17, 0.41)	0.36
l	l	15	0.25	(0.14, 0.37)	0.15
v	v	29	0.48	(0.35, 0.61)	0.49

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	10	0.16	(0.08, 0.28)	0.16
l	l	19	0.31	(0.2, 0.44)	0.29
y	y	32	0.52	(0.39, 0.65)	0.55

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	55	0.90	(0.8, 0.96)	0.72
v	v	1	0.02	(0.0, 0.09)	0.18
y	y	5	0.08	(0.03, 0.18)	0.10

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	1	0.02	(0.0, 0.09)	0.09
j	j	4	0.07	(0.02, 0.16)	0.25
k	k	56	0.92	(0.82, 0.97)	0.67

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	–	–	–	0.17
j	j	14	0.23	–	0.41
l	l	47	0.77	–	0.42

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	–	–	–	0.00
j	j	3	0.05	–	0.20
v	v	58	0.95	–	0.80

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	6	0.10	–	0.31
j	j	–	–	–	0.00
y	y	55	0.90	–	0.69

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	12	0.20	(0.11, 0.32)	0.27
k	k	48	0.79	(0.66, 0.88)	0.64
l	l	1	0.02	(0.0, 0.09)	0.09

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	44	0.72	(0.59, 0.83)	0.52
k	k	8	0.13	(0.06, 0.24)	0.18
v	v	9	0.15	(0.07, 0.26)	0.31

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	39	0.64	(0.51, 0.76)	0.43
k	k	8	0.13	(0.06, 0.24)	0.12
y	y	14	0.23	(0.13, 0.35)	0.45

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	4	0.07	–	0.18
l	l	–	–	–	0.01
v	v	57	0.93	–	0.81

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	3	0.05	(0.01, 0.14)	0.13
l	l	1	0.02	(0.0, 0.09)	0.00
y	y	57	0.93	(0.84, 0.98)	0.87

Candara (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	53	0.87	(0.76, 0.94)	0.72
v	v	1	0.02	(0.0, 0.09)	0.00
y	y	7	0.11	(0.05, 0.22)	0.28

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	15	0.25	(0.14, 0.37)	0.41
k	k	42	0.69	(0.56, 0.8)	0.50
l	l	4	0.07	(0.02, 0.16)	0.08

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	44	0.72	(0.59, 0.83)	0.62
k	k	10	0.16	(0.08, 0.28)	0.12
v	v	7	0.11	(0.05, 0.22)	0.26

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	31	0.51	(0.38, 0.64)	0.35
k	k	27	0.44	(0.32, 0.58)	0.44
y	y	3	0.05	(0.01, 0.14)	0.21

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	11	0.18	–	0.37
l	l	–	–	–	0.00
v	v	50	0.82	–	0.63

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	3	0.05	(0.01, 0.14)	0.11
l	l	14	0.23	(0.13, 0.35)	0.37
y	y	44	0.72	(0.59, 0.83)	0.52

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	49	0.80	(0.68, 0.89)	0.59
v	v	9	0.15	(0.07, 0.26)	0.41
y	y	3	0.05	(0.01, 0.14)	0.00

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
k	k	10	0.16	(0.08, 0.28)	0.19
l	l	32	0.52	(0.39, 0.65)	0.39
v	v	19	0.31	(0.2, 0.44)	0.42

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
k	k	9	0.15	(0.07, 0.26)	0.12
l	l	33	0.54	(0.41, 0.67)	0.32
y	y	19	0.31	(0.2, 0.44)	0.55

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
k	k	53	0.87	–	0.52
v	v	–	–	–	0.10
y	y	8	0.13	–	0.38

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
l	l	55	0.90	–	0.67
v	v	–	–	–	0.00
y	y	6	0.10	–	0.33

Candara (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	53	0.90	–	0.66
i	i	–	–	–	0.18
j	j	6	0.10	–	0.16

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	3	0.05	(0.01, 0.14)	0.01
i	i	49	0.83	(0.71, 0.92)	0.67
o	o	7	0.12	(0.05, 0.23)	0.32

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	–	–	–	0.05
i	i	56	0.95	–	0.76
p	p	3	0.05	–	0.19

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	33	0.56	(0.42, 0.69)	0.54
i	i	21	0.36	(0.24, 0.49)	0.23
r	r	5	0.08	(0.03, 0.19)	0.23

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	20	0.34	(0.22, 0.47)	0.32
i	i	14	0.24	(0.14, 0.37)	0.23
v	v	25	0.42	(0.3, 0.56)	0.45

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	14	0.24	(0.14, 0.37)	0.32
i	i	9	0.15	(0.07, 0.27)	0.23
x	x	36	0.61	(0.47, 0.73)	0.45

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	1	0.02	(0.0, 0.09)	0.02
j	j	51	0.86	(0.75, 0.94)	0.64
o	o	7	0.12	(0.05, 0.23)	0.34

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	2	0.03	(0.0, 0.12)	0.14
j	j	56	0.95	(0.86, 0.99)	0.68
p	p	1	0.02	(0.0, 0.09)	0.17

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	31	0.53	(0.39, 0.66)	0.57
j	j	20	0.34	(0.22, 0.47)	0.22
r	r	8	0.14	(0.06, 0.25)	0.21

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	7	0.12	(0.05, 0.23)	0.21
j	j	11	0.19	(0.1, 0.31)	0.29
v	v	41	0.69	(0.56, 0.81)	0.51

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	7	0.12	(0.05, 0.23)	0.21
j	j	8	0.14	(0.06, 0.25)	0.28
x	x	44	0.75	(0.62, 0.85)	0.51

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	1	0.02	(0.0, 0.09)	0.30
o	o	54	0.92	(0.81, 0.97)	0.46
p	p	4	0.07	(0.02, 0.16)	0.24

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	6	0.10	(0.04, 0.21)	0.13
o	o	4	0.07	(0.02, 0.16)	0.25
r	r	49	0.83	(0.71, 0.92)	0.62

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	10	0.17	(0.08, 0.29)	0.30
o	o	2	0.03	(0.0, 0.12)	0.00
v	v	47	0.80	(0.67, 0.89)	0.70

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	3	0.05	(0.01, 0.14)	0.30
o	o	2	0.03	(0.0, 0.12)	0.00
x	x	54	0.92	(0.81, 0.97)	0.70

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	3	0.05	(0.01, 0.14)	0.17
p	p	2	0.03	(0.0, 0.12)	0.12
r	r	54	0.92	(0.81, 0.97)	0.72

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	3	0.05	(0.01, 0.14)	0.09
p	p	3	0.05	(0.01, 0.14)	0.04
v	v	53	0.90	(0.79, 0.96)	0.87

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	-	-	-	0.09
p	p	-	-	-	0.04
x	x	59	1.00	-	0.87

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	26	0.44	(0.31, 0.58)	0.37
r	r	3	0.05	(0.01, 0.14)	0.18
v	v	30	0.51	(0.37, 0.64)	0.44

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	19	0.32	(0.21, 0.46)	0.37
r	r	1	0.02	(0.0, 0.09)	0.18
x	x	39	0.66	(0.53, 0.78)	0.45

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	48	0.81	(0.69, 0.9)	0.99
v	v	2	0.03	(0.0, 0.12)	0.00
x	x	9	0.15	(0.07, 0.27)	0.01

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	-	-	-	0.05
j	j	7	0.12	-	0.07
o	o	52	0.88	-	0.88

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	6	0.10	(0.04, 0.21)	0.20
j	j	4	0.07	(0.02, 0.16)	0.08
p	p	49	0.83	(0.71, 0.92)	0.72

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	2	0.03	(0.0, 0.12)	0.31
j	j	10	0.17	(0.08, 0.29)	0.19
r	r	47	0.80	(0.67, 0.89)	0.50

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	1	0.02	(0.0, 0.09)	0.00
j	j	4	0.07	(0.02, 0.16)	0.14
v	v	54	0.92	(0.81, 0.97)	0.86

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	1	0.02	(0.0, 0.09)	0.00
j	j	3	0.05	(0.01, 0.14)	0.14
x	x	55	0.93	(0.84, 0.98)	0.86

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	46	0.78	(0.65, 0.88)	0.72
o	o	8	0.14	(0.06, 0.25)	0.22
p	p	5	0.08	(0.03, 0.19)	0.06

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	18	0.31	(0.19, 0.44)	0.23
o	o	40	0.68	(0.54, 0.79)	0.72
r	r	1	0.02	(0.0, 0.09)	0.04

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	22	0.37	(0.25, 0.51)	0.51
o	o	26	0.44	(0.31, 0.58)	0.34
v	v	11	0.19	(0.1, 0.31)	0.14

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	13	0.22	(0.12, 0.35)	0.51
o	o	27	0.46	(0.33, 0.59)	0.34
x	x	19	0.32	(0.21, 0.46)	0.15

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	27	0.46	(0.33, 0.59)	0.32
p	p	28	0.47	(0.34, 0.61)	0.57
r	r	4	0.07	(0.02, 0.16)	0.11

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	6	0.10	(0.04, 0.21)	0.31
p	p	24	0.41	(0.28, 0.54)	0.35
v	v	29	0.49	(0.36, 0.63)	0.34

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	10	0.17	(0.08, 0.29)	0.31
p	p	17	0.29	(0.18, 0.42)	0.34
x	x	32	0.54	(0.41, 0.67)	0.35

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	24	0.41	(0.28, 0.54)	0.21
r	r	4	0.07	(0.02, 0.16)	0.13
v	v	31	0.53	(0.39, 0.66)	0.66

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	13	0.22	(0.12, 0.35)	0.20
r	r	1	0.02	(0.0, 0.09)	0.13
x	x	45	0.76	(0.63, 0.86)	0.67

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	52	0.88	(0.77, 0.95)	0.98
v	v	1	0.02	(0.0, 0.09)	0.01
x	x	6	0.10	(0.04, 0.21)	0.02

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	46	0.78	(0.65, 0.88)	0.66
o	o	10	0.17	(0.08, 0.29)	0.32
p	p	3	0.05	(0.01, 0.14)	0.02

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	23	0.39	(0.27, 0.53)	0.23
o	o	32	0.54	(0.41, 0.67)	0.74
r	r	4	0.07	(0.02, 0.16)	0.03

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	20	0.34	(0.22, 0.47)	0.57
o	o	23	0.39	(0.27, 0.53)	0.23
v	v	16	0.27	(0.16, 0.4)	0.20

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	21	0.36	(0.24, 0.49)	0.57
o	o	15	0.25	(0.15, 0.38)	0.23
x	x	23	0.39	(0.27, 0.53)	0.21

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	34	0.58	(0.44, 0.7)	0.26
p	p	11	0.19	(0.1, 0.31)	0.55
r	r	14	0.24	(0.14, 0.37)	0.18

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	15	0.25	(0.15, 0.38)	0.32
p	p	7	0.12	(0.05, 0.23)	0.20
v	v	37	0.63	(0.49, 0.75)	0.48

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	11	0.19	(0.1, 0.31)	0.32
p	p	3	0.05	(0.01, 0.14)	0.20
x	x	45	0.76	(0.63, 0.86)	0.49

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	23	0.39	(0.27, 0.53)	0.23
r	r	3	0.05	(0.01, 0.14)	0.00
v	v	33	0.56	(0.42, 0.69)	0.77

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	22	0.37	(0.25, 0.51)	0.23
r	r	2	0.03	(0.0, 0.12)	0.00
x	x	35	0.59	(0.46, 0.72)	0.77

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	55	0.93	–	0.99
v	v	–	–	–	0.00
x	x	4	0.07	–	0.01

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
o	o	13	0.22	(0.12, 0.35)	0.31
p	p	4	0.07	(0.02, 0.16)	0.07
r	r	42	0.71	(0.58, 0.82)	0.63

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
o	o	1	0.02	(0.0, 0.09)	0.00
p	p	5	0.08	(0.03, 0.19)	0.26
v	v	53	0.90	(0.79, 0.96)	0.74

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
o	o	2	0.03	(0.0, 0.12)	0.00
p	p	5	0.08	(0.03, 0.19)	0.25
x	x	52	0.88	(0.77, 0.95)	0.75

Century Schoolbook (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
o	o	32	0.54	(0.41, 0.67)	0.32
r	r	9	0.15	(0.07, 0.27)	0.44
v	v	18	0.31	(0.19, 0.44)	0.24

Century Schoolbook (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
o	o	22	0.37	(0.25, 0.51)	0.32
r	r	6	0.10	(0.04, 0.21)	0.44
x	x	31	0.53	(0.39, 0.66)	0.24

Century Schoolbook (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
o	o	47	0.80	(0.67, 0.89)	0.97
v	v	2	0.03	(0.0, 0.12)	0.01
x	x	10	0.17	(0.08, 0.29)	0.02

Century Schoolbook (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
p	p	20	0.34	(0.22, 0.47)	0.32
r	r	5	0.08	(0.03, 0.19)	0.18
v	v	34	0.58	(0.44, 0.7)	0.50

Century Schoolbook (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
p	p	18	0.31	(0.19, 0.44)	0.31
r	r	2	0.03	(0.0, 0.12)	0.18
x	x	39	0.66	(0.53, 0.78)	0.50

Century Schoolbook (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
p	p	50	0.85	(0.73, 0.93)	0.99
v	v	2	0.03	(0.0, 0.12)	0.00
x	x	7	0.12	(0.05, 0.23)	0.01

Century Schoolbook (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
r	r	47	0.80	(0.67, 0.89)	0.99
v	v	1	0.02	(0.0, 0.09)	0.00
x	x	11	0.19	(0.1, 0.31)	0.01

Century Schoolbook (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	1	0.01	(0.0, 0.06)	0.18
e	e	3	0.04	(0.01, 0.1)	0.17
f	f	81	0.95	(0.88, 0.99)	0.66

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	1	0.01	(0.0, 0.06)	0.19
e	e	7	0.08	(0.03, 0.16)	0.15
h	h	77	0.91	(0.82, 0.96)	0.66

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	-	-	-	0.16
e	e	5	0.06	-	0.12
k	k	80	0.94	-	0.73

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	23	0.27	(0.18, 0.38)	0.47
e	e	42	0.49	(0.38, 0.6)	0.30
o	o	20	0.24	(0.15, 0.34)	0.23

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	2	0.02	(0.0, 0.08)	0.19
e	e	6	0.07	(0.03, 0.15)	0.14
w	w	77	0.91	(0.82, 0.96)	0.67

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	3	0.04	(0.01, 0.1)	0.16
e	e	4	0.05	(0.01, 0.12)	0.14
y	y	78	0.92	(0.84, 0.97)	0.70

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	65	0.76	(0.66, 0.85)	0.76
f	f	18	0.21	(0.13, 0.31)	0.08
h	h	2	0.02	(0.0, 0.08)	0.16

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	70	0.82	(0.73, 0.9)	0.72
f	f	10	0.12	(0.06, 0.21)	0.04
k	k	5	0.06	(0.02, 0.13)	0.24

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	2	0.02	(0.0, 0.08)	0.14
f	f	80	0.94	(0.87, 0.98)	0.66
o	o	3	0.04	(0.01, 0.1)	0.19

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	27	0.32	(0.22, 0.43)	0.47
f	f	43	0.51	(0.4, 0.62)	0.25
w	w	15	0.18	(0.1, 0.27)	0.29

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	62	0.73	(0.62, 0.82)	0.47
f	f	10	0.12	(0.06, 0.21)	0.23
y	y	13	0.15	(0.08, 0.25)	0.30

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	71	0.84	–	0.67
h	h	–	–	–	0.11
k	k	14	0.16	–	0.23

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	1	0.01	(0.0, 0.06)	0.13
h	h	74	0.87	(0.78, 0.93)	0.69
o	o	10	0.12	(0.06, 0.21)	0.18

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	34	0.40	(0.3, 0.51)	0.44
h	h	28	0.33	(0.23, 0.44)	0.30
w	w	23	0.27	(0.18, 0.38)	0.26

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	56	0.66	(0.55, 0.76)	0.43
h	h	3	0.04	(0.01, 0.1)	0.29
y	y	26	0.31	(0.21, 0.42)	0.29

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	1	0.01	(0.0, 0.06)	0.10
k	k	78	0.92	(0.84, 0.97)	0.76
o	o	6	0.07	(0.03, 0.15)	0.14

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	56	0.66	(0.55, 0.76)	0.70
k	k	26	0.31	(0.21, 0.42)	0.24
w	w	3	0.04	(0.01, 0.1)	0.05

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	66	0.78	(0.67, 0.86)	0.69
k	k	5	0.06	(0.02, 0.13)	0.22
y	y	14	0.16	(0.09, 0.26)	0.09

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	–	–	–	0.13
o	o	7	0.08	–	0.16
w	w	78	0.92	–	0.70

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	1	0.01	(0.0, 0.06)	0.11
o	o	6	0.07	(0.03, 0.15)	0.16
y	y	78	0.92	(0.84, 0.97)	0.72

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	56	0.66	(0.55, 0.76)	0.84
w	w	4	0.05	(0.01, 0.12)	0.06
y	y	25	0.29	(0.2, 0.4)	0.10

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	75	0.88	(0.79, 0.94)	0.79
f	f	8	0.09	(0.04, 0.18)	0.06
h	h	2	0.02	(0.0, 0.08)	0.15

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	63	0.74	(0.63, 0.83)	0.73
f	f	8	0.09	(0.04, 0.18)	0.03
k	k	14	0.16	(0.09, 0.26)	0.23

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	–	–	–	0.04
f	f	81	0.95	–	0.84
o	o	4	0.05	–	0.12

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	30	0.35	(0.25, 0.46)	0.47
f	f	37	0.44	(0.33, 0.55)	0.24
w	w	18	0.21	(0.13, 0.31)	0.29

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	59	0.69	(0.58, 0.79)	0.46
f	f	9	0.11	(0.05, 0.19)	0.23
y	y	17	0.20	(0.12, 0.3)	0.31

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	77	0.91	(0.82, 0.96)	0.67
h	h	1	0.01	(0.0, 0.06)	0.11
k	k	7	0.08	(0.03, 0.16)	0.22

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	3	0.04	(0.01, 0.1)	0.04
h	h	79	0.93	(0.85, 0.97)	0.83
o	o	3	0.04	(0.01, 0.1)	0.14

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	35	0.41	(0.31, 0.52)	0.42
h	h	21	0.25	(0.16, 0.35)	0.31
w	w	29	0.34	(0.24, 0.45)	0.27

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	63	0.74	(0.63, 0.83)	0.41
h	h	4	0.05	(0.01, 0.12)	0.29
y	y	18	0.21	(0.13, 0.31)	0.30

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	3	0.04	(0.01, 0.1)	0.00
k	k	80	0.94	(0.87, 0.98)	0.90
o	o	2	0.02	(0.0, 0.08)	0.10

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	64	0.75	(0.65, 0.84)	0.71
k	k	17	0.20	(0.12, 0.3)	0.24
w	w	4	0.05	(0.01, 0.12)	0.06

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	69	0.81	(0.71, 0.89)	0.70
k	k	4	0.05	(0.01, 0.12)	0.21
y	y	12	0.14	(0.08, 0.23)	0.09

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	2	0.02	(0.0, 0.08)	0.03
o	o	2	0.02	(0.0, 0.08)	0.13
w	w	81	0.95	(0.88, 0.99)	0.84

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	2	0.02	(0.0, 0.08)	0.01
o	o	3	0.04	(0.01, 0.1)	0.10
y	y	80	0.94	(0.87, 0.98)	0.90

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
e	e	60	0.71	(0.6, 0.8)	0.87
w	w	2	0.02	(0.0, 0.08)	0.04
y	y	23	0.27	(0.18, 0.38)	0.09

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	44	0.52	(0.41, 0.63)	0.21
h	h	11	0.13	(0.07, 0.22)	0.29
k	k	30	0.35	(0.25, 0.46)	0.51

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	12	0.14	(0.08, 0.23)	0.04
h	h	2	0.02	(0.0, 0.08)	0.12
o	o	71	0.84	(0.74, 0.91)	0.84

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	10	0.12	(0.06, 0.21)	0.07
h	h	3	0.04	(0.01, 0.1)	0.17
w	w	72	0.85	(0.75, 0.92)	0.76

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	18	0.21	(0.13, 0.31)	0.07
h	h	10	0.12	(0.06, 0.21)	0.17
y	y	57	0.67	(0.56, 0.77)	0.75

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	5	0.06	(0.02, 0.13)	0.00
k	k	8	0.09	(0.04, 0.18)	0.21
o	o	72	0.85	(0.75, 0.92)	0.79

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	38	0.45	(0.34, 0.56)	0.37
k	k	2	0.02	(0.0, 0.08)	0.09
w	w	45	0.53	(0.42, 0.64)	0.53

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	34	0.40	(0.3, 0.51)	0.39
k	k	13	0.15	(0.08, 0.25)	0.10
y	y	38	0.45	(0.34, 0.56)	0.52

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	34	0.40	(0.3, 0.51)	0.24
o	o	45	0.53	(0.42, 0.64)	0.51
w	w	6	0.07	(0.03, 0.15)	0.25

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	9	0.11	(0.05, 0.19)	0.20
o	o	68	0.80	(0.7, 0.88)	0.52
y	y	8	0.09	(0.04, 0.18)	0.28

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	56	0.66	(0.55, 0.76)	0.73
w	w	26	0.31	(0.21, 0.42)	0.12
y	y	3	0.04	(0.01, 0.1)	0.15

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	2	0.02	(0.0, 0.08)	0.08
k	k	8	0.09	(0.04, 0.18)	0.20
o	o	75	0.88	(0.79, 0.94)	0.73

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	25	0.29	(0.2, 0.4)	0.43
k	k	6	0.07	(0.03, 0.15)	0.09
w	w	54	0.64	(0.52, 0.74)	0.48

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	23	0.27	(0.18, 0.38)	0.44
k	k	11	0.13	(0.07, 0.22)	0.09
y	y	51	0.60	(0.49, 0.7)	0.47

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	19	0.22	(0.14, 0.33)	0.30
o	o	39	0.46	(0.35, 0.57)	0.48
w	w	27	0.32	(0.22, 0.43)	0.23

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	1	0.01	(0.0, 0.06)	0.27
o	o	63	0.74	(0.63, 0.83)	0.47
y	y	21	0.25	(0.16, 0.35)	0.26

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	38	0.45	(0.34, 0.56)	0.76
w	w	31	0.36	(0.26, 0.48)	0.10
y	y	16	0.19	(0.11, 0.29)	0.14

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
k	k	16	0.19	(0.11, 0.29)	0.23
o	o	65	0.76	(0.66, 0.85)	0.77
w	w	4	0.05	(0.01, 0.12)	0.00

Courier New (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
k	k	3	0.04	(0.01, 0.1)	0.20
o	o	71	0.84	(0.74, 0.91)	0.76
y	y	11	0.13	(0.07, 0.22)	0.05

Courier New (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
k	k	29	0.34	(0.24, 0.45)	0.53
w	w	34	0.40	(0.3, 0.51)	0.22
y	y	22	0.26	(0.17, 0.37)	0.25

Courier New (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
o	o	59	0.69	(0.58, 0.79)	0.92
w	w	4	0.05	(0.01, 0.12)	0.00
y	y	22	0.26	(0.17, 0.37)	0.08

Courier New (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
g	g	53	0.79	(0.67, 0.88)	0.51
h	h	3	0.04	(0.01, 0.13)	0.23
m	m	11	0.16	(0.08, 0.27)	0.27

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
g	g	58	0.87	(0.76, 0.94)	0.64
h	h	2	0.03	(0.0, 0.1)	0.23
n	n	7	0.10	(0.04, 0.2)	0.13

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
g	g	4	0.06	(0.02, 0.15)	0.11
h	h	61	0.91	(0.82, 0.97)	0.79
q	q	2	0.03	(0.0, 0.1)	0.10

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
g	g	44	0.66	(0.53, 0.77)	0.42
h	h	5	0.07	(0.02, 0.17)	0.27
r	r	18	0.27	(0.17, 0.39)	0.32

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
g	g	9	0.13	(0.06, 0.24)	0.24
h	h	30	0.45	(0.33, 0.57)	0.38
y	y	28	0.42	(0.3, 0.54)	0.38

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
g	g	9	0.13	(0.06, 0.24)	0.27
h	h	4	0.06	(0.02, 0.15)	0.14
z	z	54	0.81	(0.69, 0.89)	0.59

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
g	g	65	0.97	–	0.54
m	m	2	0.03	–	0.30
n	n	–	–	–	0.16

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
g	g	1	0.01	(0.0, 0.08)	0.13
m	m	64	0.96	(0.87, 0.99)	0.76
q	q	2	0.03	(0.0, 0.1)	0.11

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
g	g	61	0.91	(0.82, 0.97)	0.36
m	m	2	0.03	(0.0, 0.1)	0.32
r	r	4	0.06	(0.02, 0.15)	0.32

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
g	g	2	0.03	(0.0, 0.1)	0.21
m	m	38	0.57	(0.44, 0.69)	0.42
y	y	27	0.40	(0.28, 0.53)	0.37

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
g	g	24	0.36	–	0.23
m	m	–	–	–	0.20
z	z	43	0.64	–	0.57

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
g	g	–	–	–	0.13
n	n	67	1.00	–	0.77
q	q	–	–	–	0.11

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
g	g	63	0.94	–	0.45
n	n	–	–	–	0.20
r	r	4	0.06	–	0.35

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	2	0.03	(0.0, 0.1)	0.26
n	n	38	0.57	(0.44, 0.69)	0.32
y	y	27	0.40	(0.28, 0.53)	0.43

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	30	0.45	(0.33, 0.57)	0.29
n	n	2	0.03	(0.0, 0.1)	0.06
z	z	35	0.52	(0.4, 0.65)	0.65

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	3	0.04	(0.01, 0.13)	0.09
q	q	1	0.01	(0.0, 0.08)	0.09
r	r	63	0.94	(0.85, 0.98)	0.81

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	6	0.09	(0.03, 0.18)	0.08
q	q	1	0.01	(0.0, 0.08)	0.09
y	y	60	0.90	(0.8, 0.96)	0.83

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	1	0.01	(0.0, 0.08)	0.00
q	q	1	0.01	(0.0, 0.08)	0.00
z	z	65	0.97	(0.9, 1.0)	1.00

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	3	0.04	(0.01, 0.13)	0.22
r	r	34	0.51	(0.38, 0.63)	0.37
y	y	30	0.45	(0.33, 0.57)	0.41

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	27	0.40	(0.28, 0.53)	0.26
r	r	3	0.04	(0.01, 0.13)	0.15
z	z	37	0.55	(0.43, 0.67)	0.59

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	31	0.46	(0.34, 0.59)	0.53
y	y	2	0.03	(0.0, 0.1)	0.00
z	z	34	0.51	(0.38, 0.63)	0.47

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	46	0.69	(0.56, 0.79)	0.32
m	m	20	0.30	(0.19, 0.42)	0.43
n	n	1	0.01	(0.0, 0.08)	0.25

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	2	0.03	(0.0, 0.1)	0.25
m	m	25	0.37	(0.26, 0.5)	0.27
q	q	40	0.60	(0.47, 0.72)	0.48

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	27	0.40	(0.28, 0.53)	0.21
m	m	11	0.16	(0.08, 0.27)	0.31
r	r	29	0.43	(0.31, 0.56)	0.49

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	2	0.03	(0.0, 0.1)	0.14
m	m	7	0.10	(0.04, 0.2)	0.23
y	y	58	0.87	(0.76, 0.94)	0.62

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	14	0.21	(0.12, 0.33)	0.07
m	m	1	0.01	(0.0, 0.08)	0.15
z	z	52	0.78	(0.66, 0.87)	0.77

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	1	0.01	(0.0, 0.08)	0.24
n	n	14	0.21	(0.12, 0.33)	0.16
q	q	52	0.78	(0.66, 0.87)	0.60

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	30	0.45	-	0.25
n	n	-	-	-	0.16
r	r	37	0.55	-	0.60

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	–	–	–	0.18
n	n	11	0.16	–	0.09
y	y	56	0.84	–	0.73

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	11	0.16	–	0.07
n	n	–	–	–	0.00
z	z	56	0.84	–	0.93

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	3	0.04	(0.01, 0.13)	0.26
q	q	39	0.58	(0.46, 0.7)	0.40
r	r	25	0.37	(0.26, 0.5)	0.34

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	27	0.40	(0.28, 0.53)	0.36
q	q	7	0.10	(0.04, 0.2)	0.24
y	y	33	0.49	(0.37, 0.62)	0.40

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	2	0.03	(0.0, 0.1)	0.13
q	q	9	0.13	(0.06, 0.24)	0.25
z	z	56	0.84	(0.73, 0.92)	0.62

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	4	0.06	(0.02, 0.15)	0.24
r	r	11	0.16	(0.08, 0.27)	0.27
y	y	52	0.78	(0.66, 0.87)	0.49

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	11	0.16	(0.08, 0.27)	0.14
r	r	1	0.01	(0.0, 0.08)	0.16
z	z	55	0.82	(0.71, 0.9)	0.70

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	34	0.51	(0.38, 0.63)	0.46
y	y	12	0.18	(0.1, 0.29)	0.09
z	z	21	0.31	(0.21, 0.44)	0.45

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	–	–	–	0.30
n	n	–	–	–	0.19
q	q	67	1.00	–	0.51

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	14	0.21	–	0.34
n	n	–	–	–	0.13
r	r	53	0.79	–	0.53

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	–	–	–	0.26
n	n	1	0.01	–	0.07
y	y	66	0.99	–	0.67

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	–	–	–	0.17
n	n	–	–	–	0.00
z	z	67	1.00	–	0.83

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	4	0.06	(0.02, 0.15)	0.31
q	q	60	0.90	(0.8, 0.96)	0.35
r	r	3	0.04	(0.01, 0.13)	0.35

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	37	0.55	(0.43, 0.67)	0.40
q	q	2	0.03	(0.0, 0.1)	0.21
y	y	28	0.42	(0.3, 0.54)	0.40

Futura (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	1	0.01	(0.0, 0.08)	0.18
q	q	28	0.42	(0.3, 0.54)	0.21
z	z	38	0.57	(0.44, 0.69)	0.61

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
m	m	3	0.04	(0.01, 0.13)	0.31
r	r	2	0.03	(0.0, 0.1)	0.24
y	y	62	0.93	(0.83, 0.98)	0.45

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
m	m	3	0.04	(0.01, 0.13)	0.23
r	r	1	0.01	(0.0, 0.08)	0.16
z	z	63	0.94	(0.85, 0.98)	0.61

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
m	m	35	0.52	(0.4, 0.65)	0.56
y	y	26	0.39	(0.27, 0.51)	0.09
z	z	6	0.09	(0.03, 0.18)	0.35

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
n	n	1	0.01	(0.0, 0.08)	0.20
q	q	63	0.94	(0.85, 0.98)	0.43
r	r	3	0.04	(0.01, 0.13)	0.37

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
n	n	38	0.57	(0.44, 0.69)	0.30
q	q	1	0.01	(0.0, 0.08)	0.25
y	y	28	0.42	(0.3, 0.54)	0.45

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
n	n	3	0.04	(0.01, 0.13)	0.06
q	q	24	0.36	(0.24, 0.48)	0.26
z	z	40	0.60	(0.47, 0.72)	0.68

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
n	n	1	0.01	–	0.17
r	r	–	–	–	0.29
y	y	66	0.99	–	0.54

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
n	n	–	–	–	0.06
r	r	5	0.07	–	0.17
z	z	62	0.93	–	0.77

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
n	n	31	0.46	(0.34, 0.59)	0.41
y	y	29	0.43	(0.31, 0.56)	0.10
z	z	7	0.10	(0.04, 0.2)	0.49

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
q	q	6	0.09	(0.03, 0.18)	0.23
r	r	37	0.55	(0.43, 0.67)	0.38
y	y	24	0.36	(0.24, 0.48)	0.38

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
q	q	30	0.45	(0.33, 0.57)	0.26
r	r	1	0.01	(0.0, 0.08)	0.15
z	z	36	0.54	(0.41, 0.66)	0.59

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
q	q	25	0.37	(0.26, 0.5)	0.51
y	y	3	0.04	(0.01, 0.13)	0.00
z	z	39	0.58	(0.46, 0.7)	0.49

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
r	r	34	0.51	(0.38, 0.63)	0.45
y	y	24	0.36	(0.24, 0.48)	0.08
z	z	9	0.13	(0.06, 0.24)	0.47

Futura (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	–	–	–	0.08
d	d	–	–	–	0.08
i	i	64	1.00	–	0.83

Georgia (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
b	b	–	–	–	0.10
d	d	–	–	–	0.10
j	j	64	1.00	–	0.80

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	2	0.03	(0.0, 0.11)	0.14
d	d	2	0.03	(0.0, 0.11)	0.12
k	k	60	0.94	(0.85, 0.98)	0.75

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	3	0.05	-	0.12
d	d	-	-	-	0.12
l	l	61	0.95	-	0.75

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	-	-	-	0.00
d	d	1	0.02	-	0.00
v	v	63	0.98	-	1.00

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	-	-	-	0.02
d	d	-	-	-	0.02
y	y	64	1.00	-	0.96

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	59	0.92	(0.83, 0.97)	0.66
i	i	2	0.03	(0.0, 0.11)	0.18
j	j	3	0.05	(0.01, 0.13)	0.16

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	7	0.11	(0.05, 0.21)	0.39
i	i	51	0.80	(0.68, 0.89)	0.39
k	k	6	0.09	(0.04, 0.19)	0.23

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	31	0.48	-	0.64
i	i	33	0.52	-	0.26
l	l	-	-	-	0.10

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	27	0.42	(0.3, 0.55)	0.31
i	i	20	0.31	(0.2, 0.44)	0.22
v	v	17	0.27	(0.16, 0.39)	0.47

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	18	0.28	(0.18, 0.41)	0.20
i	i	22	0.34	(0.23, 0.47)	0.29
y	y	24	0.38	(0.26, 0.5)	0.51

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	4	0.06	(0.02, 0.15)	0.29
j	j	52	0.81	(0.7, 0.9)	0.42
k	k	8	0.12	(0.06, 0.23)	0.29

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	22	0.34	(0.23, 0.47)	0.51
j	j	36	0.56	(0.43, 0.69)	0.31
l	l	6	0.09	(0.04, 0.19)	0.18

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	11	0.17	(0.09, 0.29)	0.20
j	j	18	0.28	(0.18, 0.41)	0.28
v	v	35	0.55	(0.42, 0.67)	0.52

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	51	0.80	(0.68, 0.89)	0.34
j	j	5	0.08	(0.03, 0.17)	0.22
y	y	8	0.12	(0.06, 0.23)	0.44

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	17	0.27	(0.16, 0.39)	0.45
k	k	9	0.14	(0.07, 0.25)	0.28
l	l	38	0.59	(0.46, 0.71)	0.28

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	22	0.34	-	0.54
k	k	-	-	-	0.01
v	v	42	0.66	-	0.45

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	15	0.23	(0.14, 0.36)	0.43
k	k	3	0.05	(0.01, 0.13)	0.08
y	y	46	0.72	(0.59, 0.82)	0.49

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	7	0.11	(0.05, 0.21)	0.31
l	l	6	0.09	(0.04, 0.19)	0.12
v	v	51	0.80	(0.68, 0.89)	0.57

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	7	0.11	(0.05, 0.21)	0.20
l	l	4	0.06	(0.02, 0.15)	0.19
y	y	53	0.83	(0.71, 0.91)	0.60

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	54	0.84	(0.73, 0.92)	0.81
v	v	4	0.06	(0.02, 0.15)	0.10
y	y	6	0.09	(0.04, 0.19)	0.09

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	54	0.84	(0.73, 0.92)	0.66
i	i	3	0.05	(0.01, 0.13)	0.18
j	j	7	0.11	(0.05, 0.21)	0.16

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	8	0.12	(0.06, 0.23)	0.38
i	i	48	0.75	(0.63, 0.85)	0.40
k	k	8	0.12	(0.06, 0.23)	0.22

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	35	0.55	(0.42, 0.67)	0.64
i	i	27	0.42	(0.3, 0.55)	0.26
l	l	2	0.03	(0.0, 0.11)	0.10

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	24	0.38	(0.26, 0.5)	0.32
i	i	18	0.28	(0.18, 0.41)	0.23
v	v	22	0.34	(0.23, 0.47)	0.45

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	8	0.12	(0.06, 0.23)	0.21
i	i	18	0.28	(0.18, 0.41)	0.30
y	y	38	0.59	(0.46, 0.71)	0.49

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	6	0.09	(0.04, 0.19)	0.28
j	j	53	0.83	(0.71, 0.91)	0.43
k	k	5	0.08	(0.03, 0.17)	0.28

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	20	0.31	(0.2, 0.44)	0.52
j	j	42	0.66	(0.53, 0.77)	0.31
l	l	2	0.03	(0.0, 0.11)	0.18

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	12	0.19	(0.1, 0.3)	0.21
j	j	16	0.25	(0.15, 0.37)	0.29
v	v	36	0.56	(0.43, 0.69)	0.51

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	52	0.81	(0.7, 0.9)	0.35
j	j	8	0.12	(0.06, 0.23)	0.23
y	y	4	0.06	(0.02, 0.15)	0.42

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	16	0.25	(0.15, 0.37)	0.44
k	k	10	0.16	(0.08, 0.27)	0.27
l	l	38	0.59	(0.46, 0.71)	0.29

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	18	0.28	(0.18, 0.41)	0.52
k	k	3	0.05	(0.01, 0.13)	0.01
v	v	43	0.67	(0.54, 0.78)	0.47

Georgia (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	22	0.34	(0.23, 0.47)	0.41
k	k	2	0.03	(0.0, 0.11)	0.08
y	y	40	0.62	(0.5, 0.74)	0.51

Georgia (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	7	0.11	(0.05, 0.21)	0.32
l	l	6	0.09	(0.04, 0.19)	0.13
v	v	51	0.80	(0.68, 0.89)	0.55

Georgia (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	6	0.09	(0.04, 0.19)	0.21
l	l	8	0.12	(0.06, 0.23)	0.20
y	y	50	0.78	(0.66, 0.87)	0.59

Georgia (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	55	0.86	(0.75, 0.93)	0.81
v	v	2	0.03	(0.0, 0.11)	0.10
y	y	7	0.11	(0.05, 0.21)	0.09

Georgia (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	2	0.03	(0.0, 0.11)	0.11
j	j	6	0.09	(0.04, 0.19)	0.25
k	k	56	0.88	(0.77, 0.94)	0.64

Georgia (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	5	0.08	(0.03, 0.17)	0.21
j	j	19	0.30	(0.19, 0.42)	0.38
l	l	40	0.62	(0.5, 0.74)	0.41

Georgia (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	2	0.03	(0.0, 0.11)	0.00
j	j	7	0.11	(0.05, 0.21)	0.14
v	v	55	0.86	(0.75, 0.93)	0.86

Georgia (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	10	0.16	(0.08, 0.27)	0.23
j	j	1	0.02	(0.0, 0.08)	0.00
y	y	53	0.83	(0.71, 0.91)	0.77

Georgia (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	24	0.38	(0.26, 0.5)	0.31
k	k	39	0.61	(0.48, 0.73)	0.57
l	l	1	0.02	(0.0, 0.08)	0.12

Georgia (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	42	0.66	(0.53, 0.77)	0.57
k	k	9	0.14	(0.07, 0.25)	0.03
v	v	13	0.20	(0.11, 0.32)	0.40

Georgia (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	45	0.70	(0.58, 0.81)	0.48
k	k	5	0.08	(0.03, 0.17)	0.01
y	y	14	0.22	(0.13, 0.34)	0.51

Georgia (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	10	0.16	(0.08, 0.27)	0.13
l	l	4	0.06	(0.02, 0.15)	0.01
v	v	50	0.78	(0.66, 0.87)	0.87

Georgia (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	4	0.06	(0.02, 0.15)	0.10
l	l	1	0.02	(0.0, 0.08)	0.00
y	y	59	0.92	(0.83, 0.97)	0.90

Georgia (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
i	i	57	0.89	-	0.81
v	v	-	-	-	0.00
y	y	7	0.11	-	0.19

Georgia (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
j	j	40	0.62	(0.5, 0.74)	0.42
k	k	23	0.36	(0.24, 0.49)	0.48
l	l	1	0.02	(0.0, 0.08)	0.11

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
j	j	44	0.69	(0.56, 0.8)	0.64
k	k	8	0.12	(0.06, 0.23)	0.01
v	v	12	0.19	(0.1, 0.3)	0.35

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
j	j	20	0.31	(0.2, 0.44)	0.41
k	k	40	0.62	(0.5, 0.74)	0.29
y	y	4	0.06	(0.02, 0.15)	0.30

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
j	j	14	0.22	(0.13, 0.34)	0.27
l	l	2	0.03	(0.0, 0.11)	0.00
v	v	48	0.75	(0.63, 0.85)	0.73

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
j	j	7	0.11	(0.05, 0.21)	0.09
l	l	38	0.59	(0.46, 0.71)	0.27
y	y	19	0.30	(0.19, 0.42)	0.64

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
j	j	49	0.77	(0.64, 0.86)	0.70
v	v	14	0.22	(0.13, 0.34)	0.30
y	y	1	0.02	(0.0, 0.08)	0.00

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
k	k	4	0.06	(0.02, 0.15)	0.03
l	l	22	0.34	(0.23, 0.47)	0.43
v	v	38	0.59	(0.46, 0.71)	0.54

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
k	k	5	0.08	(0.03, 0.17)	0.01
l	l	16	0.25	(0.15, 0.37)	0.37
y	y	43	0.67	(0.54, 0.78)	0.62

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
k	k	53	0.83	(0.71, 0.91)	0.52
v	v	5	0.08	(0.03, 0.17)	0.12
y	y	6	0.09	(0.04, 0.19)	0.36

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
l	l	59	0.92	(0.83, 0.97)	0.78
v	v	1	0.02	(0.0, 0.08)	0.00
y	y	4	0.06	(0.02, 0.15)	0.22

Georgia (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	21	0.32	(0.21, 0.45)	0.51
f	f	11	0.17	(0.09, 0.28)	0.10
i	i	33	0.51	(0.38, 0.63)	0.39

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	5	0.08	(0.03, 0.17)	0.31
f	f	41	0.63	(0.5, 0.75)	0.35
n	n	19	0.29	(0.19, 0.42)	0.34

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	–	–	–	0.08
f	f	59	0.91	–	0.63
p	p	6	0.09	–	0.28

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	–	–	–	0.09
f	f	62	0.95	–	0.63
q	q	3	0.05	–	0.28

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	6	0.09	(0.03, 0.19)	0.16
f	f	8	0.12	(0.05, 0.23)	0.14
w	w	51	0.78	(0.67, 0.88)	0.70

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	6	0.09	(0.03, 0.19)	0.17
f	f	10	0.15	(0.08, 0.26)	0.14
z	z	49	0.75	(0.63, 0.85)	0.69

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	6	0.09	(0.03, 0.19)	0.22
i	i	53	0.82	(0.7, 0.9)	0.57
n	n	6	0.09	(0.03, 0.19)	0.21

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	2	0.03	(0.0, 0.11)	0.04
i	i	60	0.92	(0.83, 0.97)	0.78
p	p	3	0.05	(0.01, 0.13)	0.18

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	2	0.03	(0.0, 0.11)	0.04
i	i	60	0.92	(0.83, 0.97)	0.78
q	q	3	0.05	(0.01, 0.13)	0.18

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	15	0.23	(0.14, 0.35)	0.30
i	i	10	0.15	(0.08, 0.26)	0.24
w	w	40	0.62	(0.49, 0.73)	0.46

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	18	0.28	(0.17, 0.4)	0.28
i	i	18	0.28	(0.17, 0.4)	0.26
z	z	29	0.45	(0.32, 0.57)	0.47

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	6	0.09	(0.03, 0.19)	0.20
n	n	57	0.88	(0.77, 0.95)	0.60
p	p	2	0.03	(0.0, 0.11)	0.20

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	2	0.03	(0.0, 0.11)	0.20
n	n	56	0.86	(0.75, 0.93)	0.60
q	q	7	0.11	(0.04, 0.21)	0.20

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	11	0.17	-	0.25
n	n	-	-	-	0.06
w	w	54	0.83	-	0.69

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	12	0.18	(0.1, 0.3)	0.25
n	n	5	0.08	(0.03, 0.17)	0.06
z	z	48	0.74	(0.61, 0.84)	0.69

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	33	0.51	(0.38, 0.63)	0.47
p	p	21	0.32	(0.21, 0.45)	0.26
q	q	11	0.17	(0.09, 0.28)	0.26

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	2	0.03	(0.0, 0.11)	0.05
p	p	3	0.05	(0.01, 0.13)	0.06
w	w	60	0.92	(0.83, 0.97)	0.88

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	2	0.03	(0.0, 0.11)	0.06
p	p	2	0.03	(0.0, 0.11)	0.06
z	z	61	0.94	(0.85, 0.98)	0.88

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	1	0.02	(0.0, 0.08)	0.05
q	q	3	0.05	(0.01, 0.13)	0.06
w	w	61	0.94	(0.85, 0.98)	0.88

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	2	0.03	(0.0, 0.11)	0.06
q	q	1	0.02	(0.0, 0.08)	0.06
z	z	62	0.95	(0.87, 0.99)	0.88

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
b	b	43	0.66	(0.53, 0.77)	0.67
w	w	16	0.25	(0.15, 0.37)	0.18
z	z	6	0.09	(0.03, 0.19)	0.15

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	13	0.20	(0.11, 0.32)	0.07
i	i	24	0.37	(0.25, 0.5)	0.47
n	n	28	0.43	(0.31, 0.56)	0.47

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	10	0.15	(0.08, 0.26)	0.07
i	i	28	0.43	(0.31, 0.56)	0.32
p	p	27	0.42	(0.29, 0.54)	0.61

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	8	0.12	(0.05, 0.23)	0.07
i	i	26	0.40	(0.28, 0.53)	0.32
q	q	31	0.48	(0.35, 0.6)	0.61

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	12	0.18	(0.1, 0.3)	0.17
i	i	10	0.15	(0.08, 0.26)	0.12
w	w	43	0.66	(0.53, 0.77)	0.71

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	13	0.20	(0.11, 0.32)	0.14
i	i	12	0.18	(0.1, 0.3)	0.14
z	z	40	0.62	(0.49, 0.73)	0.72

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	50	0.77	(0.65, 0.86)	0.42
n	n	8	0.12	(0.05, 0.23)	0.18
p	p	7	0.11	(0.04, 0.21)	0.40

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	41	0.63	(0.5, 0.75)	0.42
n	n	9	0.14	(0.07, 0.25)	0.18
q	q	15	0.23	(0.14, 0.35)	0.40

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	19	0.29	-	0.23
n	n	-	-	-	0.06
w	w	46	0.71	-	0.72

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	19	0.29	(0.19, 0.42)	0.22
n	n	6	0.09	(0.03, 0.19)	0.07
z	z	40	0.62	(0.49, 0.73)	0.71

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	59	0.91	(0.81, 0.97)	0.83
p	p	2	0.03	(0.0, 0.11)	0.08
q	q	4	0.06	(0.02, 0.15)	0.08

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	8	0.12	(0.05, 0.23)	0.21
p	p	7	0.11	(0.04, 0.21)	0.23
w	w	50	0.77	(0.65, 0.86)	0.56

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	10	0.15	(0.08, 0.26)	0.21
p	p	14	0.22	(0.12, 0.33)	0.24
z	z	41	0.63	(0.5, 0.75)	0.55

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	7	0.11	(0.04, 0.21)	0.21
q	q	7	0.11	(0.04, 0.21)	0.23
w	w	51	0.78	(0.67, 0.88)	0.56

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	14	0.22	(0.12, 0.33)	0.21
q	q	12	0.18	(0.1, 0.3)	0.24
z	z	39	0.60	(0.47, 0.72)	0.55

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	53	0.82	(0.7, 0.9)	0.66
w	w	11	0.17	(0.09, 0.28)	0.19
z	z	1	0.02	(0.0, 0.08)	0.15

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	54	0.83	(0.72, 0.91)	0.60
n	n	5	0.08	(0.03, 0.17)	0.11
p	p	6	0.09	(0.03, 0.19)	0.29

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	51	0.78	(0.67, 0.88)	0.61
n	n	3	0.05	(0.01, 0.13)	0.11
q	q	11	0.17	(0.09, 0.28)	0.29

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	21	0.32	(0.21, 0.45)	0.34
n	n	10	0.15	(0.08, 0.26)	0.24
w	w	34	0.52	(0.4, 0.65)	0.42

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	30	0.46	(0.34, 0.59)	0.37
n	n	11	0.17	(0.09, 0.28)	0.21
z	z	24	0.37	(0.25, 0.5)	0.42

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	60	0.92	(0.83, 0.97)	0.91
p	p	3	0.05	(0.01, 0.13)	0.04
q	q	2	0.03	(0.0, 0.11)	0.04

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	17	0.26	(0.16, 0.39)	0.29
p	p	18	0.28	(0.17, 0.4)	0.35
w	w	30	0.46	(0.34, 0.59)	0.37

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	19	0.29	(0.19, 0.42)	0.30
p	p	17	0.26	(0.16, 0.39)	0.33
z	z	29	0.45	(0.32, 0.57)	0.37

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	16	0.25	(0.15, 0.37)	0.29
q	q	21	0.32	(0.21, 0.45)	0.35
w	w	28	0.43	(0.31, 0.56)	0.37

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	21	0.32	(0.21, 0.45)	0.30
q	q	23	0.35	(0.24, 0.48)	0.33
z	z	21	0.32	(0.21, 0.45)	0.37

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
i	i	55	0.85	(0.74, 0.92)	0.64
w	w	7	0.11	(0.04, 0.21)	0.17
z	z	3	0.05	(0.01, 0.13)	0.19

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
n	n	58	0.89	–	0.75
p	p	–	–	–	0.12
q	q	7	0.11	–	0.12

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
n	n	3	0.05	(0.01, 0.13)	0.06
p	p	12	0.18	(0.1, 0.3)	0.25
w	w	50	0.77	(0.65, 0.86)	0.69

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
n	n	1	0.02	(0.0, 0.08)	0.06
p	p	12	0.18	(0.1, 0.3)	0.26
z	z	52	0.80	(0.68, 0.89)	0.68

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
n	n	1	0.02	(0.0, 0.08)	0.06
q	q	16	0.25	(0.15, 0.37)	0.25
w	w	48	0.74	(0.61, 0.84)	0.69

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
n	n	1	0.02	(0.0, 0.08)	0.06
q	q	20	0.31	(0.2, 0.43)	0.26
z	z	44	0.68	(0.55, 0.79)	0.68

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
n	n	40	0.62	(0.49, 0.73)	0.58
w	w	16	0.25	(0.15, 0.37)	0.22
z	z	9	0.14	(0.07, 0.25)	0.19

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
p	p	2	0.03	(0.0, 0.11)	0.00
q	q	3	0.05	(0.01, 0.13)	0.00
w	w	60	0.92	(0.83, 0.97)	1.00

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
p	p	3	0.05	(0.01, 0.13)	0.00
q	q	1	0.02	(0.0, 0.08)	0.00
z	z	61	0.94	(0.85, 0.98)	1.00

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
p	p	49	0.75	(0.63, 0.85)	0.67
w	w	14	0.22	(0.12, 0.33)	0.18
z	z	2	0.03	(0.0, 0.11)	0.15

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
q	q	52	0.80	(0.68, 0.89)	0.67
w	w	11	0.17	(0.09, 0.28)	0.18
z	z	2	0.03	(0.0, 0.11)	0.15

PT Sans (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	56	0.89	(0.78, 0.95)	0.81
h	h	3	0.05	(0.01, 0.13)	0.12
m	m	4	0.06	(0.02, 0.15)	0.08

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	56	0.89	(0.78, 0.95)	0.87
h	h	1	0.02	(0.0, 0.09)	0.11
n	n	6	0.10	(0.04, 0.2)	0.02

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	37	0.59	(0.46, 0.71)	0.39
h	h	22	0.35	(0.23, 0.48)	0.59
q	q	4	0.06	(0.02, 0.15)	0.02

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	49	0.78	(0.66, 0.87)	0.67
h	h	8	0.13	(0.06, 0.23)	0.17
r	r	6	0.10	(0.04, 0.2)	0.16

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	29	0.46	(0.33, 0.59)	0.42
h	h	20	0.32	(0.21, 0.45)	0.52
y	y	14	0.22	(0.13, 0.34)	0.06

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	31	0.49	(0.36, 0.62)	0.36
h	h	4	0.06	(0.02, 0.15)	0.47
z	z	28	0.44	(0.32, 0.58)	0.17

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	62	0.98	-	0.88
m	m	1	0.02	-	0.08
n	n	-	-	-	0.04

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	22	0.35	(0.23, 0.48)	0.41
m	m	36	0.57	(0.44, 0.7)	0.56
q	q	5	0.08	(0.03, 0.18)	0.03

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	50	0.79	(0.67, 0.89)	0.67
m	m	4	0.06	(0.02, 0.15)	0.15
r	r	9	0.14	(0.07, 0.25)	0.18

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	28	0.44	(0.32, 0.58)	0.43
m	m	23	0.37	(0.25, 0.5)	0.50
y	y	12	0.19	(0.1, 0.31)	0.08

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	32	0.51	(0.38, 0.64)	0.36
m	m	8	0.13	(0.06, 0.23)	0.44
z	z	23	0.37	(0.25, 0.5)	0.20

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	25	0.40	(0.28, 0.53)	0.45
n	n	33	0.52	(0.39, 0.65)	0.53
q	q	5	0.08	(0.03, 0.18)	0.02

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	55	0.87	(0.77, 0.94)	0.74
n	n	2	0.03	(0.0, 0.11)	0.09
r	r	6	0.10	(0.04, 0.2)	0.17

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	30	0.48	(0.35, 0.61)	0.47
n	n	25	0.40	(0.28, 0.53)	0.46
y	y	8	0.13	(0.06, 0.23)	0.07

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	39	0.62	(0.49, 0.74)	0.41
n	n	2	0.03	(0.0, 0.11)	0.41
z	z	22	0.35	(0.23, 0.48)	0.19

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	20	0.32	(0.21, 0.45)	0.40
q	q	9	0.14	(0.07, 0.25)	0.01
r	r	34	0.54	(0.41, 0.67)	0.59

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	23	0.37	(0.25, 0.5)	0.26
q	q	8	0.13	(0.06, 0.23)	0.25
y	y	32	0.51	(0.38, 0.64)	0.50

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	13	0.21	(0.11, 0.33)	0.08
q	q	3	0.05	(0.01, 0.13)	0.26
z	z	47	0.75	(0.62, 0.85)	0.66

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	28	0.44	(0.32, 0.58)	0.50
r	r	27	0.43	(0.3, 0.56)	0.48
y	y	8	0.13	(0.06, 0.23)	0.01

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	44	0.70	(0.57, 0.81)	0.38
r	r	3	0.05	(0.01, 0.13)	0.47
z	z	16	0.25	(0.15, 0.38)	0.15

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	37	0.59	(0.46, 0.71)	0.52
y	y	2	0.03	(0.0, 0.11)	0.14
z	z	24	0.38	(0.26, 0.51)	0.34

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	46	0.73	–	0.39
m	m	17	0.27	–	0.34
n	n	–	–	–	0.27

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	–	–	–	0.21
m	m	9	0.14	–	0.15
q	q	54	0.86	–	0.64

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	14	0.22	(0.13, 0.34)	0.24
m	m	5	0.08	(0.03, 0.18)	0.20
r	r	44	0.70	(0.57, 0.81)	0.55

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	1	0.02	(0.0, 0.09)	0.16
m	m	6	0.10	(0.04, 0.2)	0.12
y	y	56	0.89	(0.78, 0.95)	0.72

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	8	0.13	(0.06, 0.23)	0.10
m	m	2	0.03	(0.0, 0.11)	0.07
z	z	53	0.84	(0.73, 0.92)	0.83

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	1	0.02	(0.0, 0.09)	0.20
n	n	8	0.13	(0.06, 0.23)	0.09
q	q	54	0.86	(0.75, 0.93)	0.71

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	17	0.27	-	0.25
n	n	-	-	-	0.13
r	r	46	0.73	-	0.62

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	1	0.02	(0.0, 0.09)	0.17
n	n	5	0.08	(0.03, 0.18)	0.06
y	y	57	0.90	(0.8, 0.96)	0.77

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	6	0.10	(0.04, 0.2)	0.10
n	n	1	0.02	(0.0, 0.09)	0.00
z	z	56	0.89	(0.78, 0.95)	0.90

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	3	0.05	(0.01, 0.13)	0.26
q	q	41	0.65	(0.52, 0.77)	0.47
r	r	19	0.30	(0.19, 0.43)	0.27

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	22	0.35	(0.23, 0.48)	0.38
q	q	12	0.19	(0.1, 0.31)	0.28
y	y	29	0.46	(0.33, 0.59)	0.34

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	-	-	-	0.23
q	q	13	0.21	-	0.27
z	z	50	0.79	-	0.50

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	14	0.22	(0.13, 0.34)	0.31
r	r	8	0.13	(0.06, 0.23)	0.22
y	y	41	0.65	(0.52, 0.77)	0.47

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	23	0.37	(0.25, 0.5)	0.19
r	r	2	0.03	(0.0, 0.11)	0.17
z	z	38	0.60	(0.47, 0.72)	0.64

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	27	0.43	(0.3, 0.56)	0.56
y	y	9	0.14	(0.07, 0.25)	0.08
z	z	27	0.43	(0.3, 0.56)	0.36

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	1	0.02	-	0.17
n	n	-	-	-	0.12
q	q	62	0.98	-	0.70

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	5	0.08	-	0.24
n	n	-	-	-	0.14
r	r	58	0.92	-	0.62

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	1	0.02	(0.0, 0.09)	0.14
n	n	1	0.02	(0.0, 0.09)	0.05
y	y	61	0.97	(0.89, 1.0)	0.81

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	2	0.03	-	0.08
n	n	-	-	-	0.00
z	z	61	0.97	-	0.92

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	9	0.14	(0.07, 0.25)	0.22
q	q	48	0.76	(0.64, 0.86)	0.47
r	r	6	0.10	(0.04, 0.2)	0.30

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	24	0.38	(0.26, 0.51)	0.35
q	q	5	0.08	(0.03, 0.18)	0.28
y	y	34	0.54	(0.41, 0.67)	0.37

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	2	0.03	(0.0, 0.11)	0.20
q	q	21	0.33	(0.22, 0.46)	0.26
z	z	40	0.63	(0.5, 0.75)	0.54

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	5	0.08	(0.03, 0.18)	0.29
r	r	2	0.03	(0.0, 0.11)	0.21
y	y	56	0.89	(0.78, 0.95)	0.49

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	12	0.19	(0.1, 0.31)	0.17
r	r	1	0.02	(0.0, 0.09)	0.18
z	z	50	0.79	(0.67, 0.89)	0.64

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
m	m	34	0.54	(0.41, 0.67)	0.58
y	y	17	0.27	(0.17, 0.4)	0.09
z	z	12	0.19	(0.1, 0.31)	0.33

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
n	n	4	0.06	(0.02, 0.15)	0.17
q	q	47	0.75	(0.62, 0.85)	0.52
r	r	12	0.19	(0.1, 0.31)	0.31

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
n	n	24	0.38	(0.26, 0.51)	0.30
q	q	9	0.14	(0.07, 0.25)	0.30
y	y	30	0.48	(0.35, 0.61)	0.40

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
n	n	2	0.03	(0.0, 0.11)	0.14
q	q	21	0.33	(0.22, 0.46)	0.29
z	z	40	0.63	(0.5, 0.75)	0.57

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
n	n	6	0.10	(0.04, 0.2)	0.22
r	r	8	0.13	(0.06, 0.23)	0.24
y	y	49	0.78	(0.66, 0.87)	0.54

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
n	n	2	0.03	(0.0, 0.11)	0.10
r	r	7	0.11	(0.05, 0.22)	0.18
z	z	54	0.86	(0.75, 0.93)	0.72

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
n	n	26	0.41	(0.29, 0.54)	0.51
y	y	23	0.37	(0.25, 0.5)	0.09
z	z	14	0.22	(0.13, 0.34)	0.40

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
q	q	19	0.30	(0.19, 0.43)	0.34
r	r	27	0.43	(0.3, 0.56)	0.34
y	y	17	0.27	(0.17, 0.4)	0.32

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
q	q	25	0.40	(0.28, 0.53)	0.28
r	r	3	0.05	(0.01, 0.13)	0.22
z	z	35	0.56	(0.42, 0.68)	0.50

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
q	q	32	0.51	(0.38, 0.64)	0.57
y	y	3	0.05	(0.01, 0.13)	0.03
z	z	28	0.44	(0.32, 0.58)	0.39

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
r	r	24	0.38	(0.26, 0.51)	0.54
y	y	20	0.32	(0.21, 0.45)	0.05
z	z	19	0.30	(0.19, 0.43)	0.41

PT Serif (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	4	0.07	(0.02, 0.16)	0.31
c	c	5	0.08	(0.03, 0.18)	0.35
f	f	52	0.85	(0.74, 0.93)	0.34

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	6	0.10	(0.04, 0.2)	0.27
c	c	1	0.02	(0.0, 0.09)	0.27
l	l	54	0.89	(0.78, 0.95)	0.46

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	7	0.11	(0.05, 0.22)	0.27
c	c	16	0.26	(0.16, 0.39)	0.47
m	m	38	0.62	(0.49, 0.74)	0.26

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	22	0.36	(0.24, 0.49)	0.26
c	c	17	0.28	(0.17, 0.41)	0.54
n	n	22	0.36	(0.24, 0.49)	0.20

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	10	0.16	(0.08, 0.28)	0.29
c	c	7	0.11	(0.05, 0.22)	0.33
t	t	44	0.72	(0.59, 0.83)	0.38

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	18	0.30	(0.19, 0.43)	0.26
c	c	23	0.38	(0.26, 0.51)	0.54
u	u	20	0.33	(0.21, 0.46)	0.20

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	47	0.77	–	0.72
f	f	–	–	–	0.07
l	l	14	0.23	–	0.21

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	19	0.31	(0.2, 0.44)	0.28
f	f	28	0.46	(0.33, 0.59)	0.48
m	m	14	0.23	(0.13, 0.35)	0.24

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	15	0.25	(0.14, 0.37)	0.30
f	f	39	0.64	(0.51, 0.76)	0.52
n	n	7	0.11	(0.05, 0.22)	0.18

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	43	0.70	(0.57, 0.81)	0.82
f	f	11	0.18	(0.09, 0.3)	0.06
t	t	7	0.11	(0.05, 0.22)	0.12

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	6	0.10	(0.04, 0.2)	0.30
f	f	49	0.80	(0.68, 0.89)	0.52
u	u	6	0.10	(0.04, 0.2)	0.18

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	20	0.33	(0.21, 0.46)	0.41
l	l	35	0.57	(0.44, 0.7)	0.50
m	m	6	0.10	(0.04, 0.2)	0.09

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	26	0.43	–	0.44
l	l	35	0.57	–	0.54
n	n	–	–	–	0.02

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	47	0.77	–	0.76
l	l	14	0.23	–	0.16
t	t	–	–	–	0.08

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	17	0.28	–	0.41
l	l	44	0.72	–	0.57
u	u	–	–	–	0.03

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	59	0.97	–	0.64
m	m	2	0.03	–	0.21
n	n	–	–	–	0.15

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	15	0.25	(0.14, 0.37)	0.33
m	m	14	0.23	(0.13, 0.35)	0.18
t	t	32	0.52	(0.39, 0.65)	0.50

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	46	0.75	(0.63, 0.86)	0.61
m	m	14	0.23	(0.13, 0.35)	0.22
u	u	1	0.02	(0.0, 0.09)	0.17

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	16	0.26	(0.16, 0.39)	0.34
n	n	9	0.15	(0.07, 0.26)	0.11
t	t	36	0.59	(0.46, 0.71)	0.54

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	57	0.93	(0.84, 0.98)	0.67
n	n	3	0.05	(0.01, 0.14)	0.16
u	u	1	0.02	(0.0, 0.09)	0.16

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
a	a	18	0.30	(0.19, 0.43)	0.34
t	t	39	0.64	(0.51, 0.76)	0.54
u	u	4	0.07	(0.02, 0.16)	0.11

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	45	0.74	(0.61, 0.84)	0.75
f	f	7	0.11	(0.05, 0.22)	0.05
l	l	9	0.15	(0.07, 0.26)	0.20

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	20	0.33	(0.21, 0.46)	0.35
f	f	23	0.38	(0.26, 0.51)	0.35
m	m	18	0.30	(0.19, 0.43)	0.30

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	16	0.26	(0.16, 0.39)	0.40
f	f	38	0.62	(0.49, 0.74)	0.34
n	n	7	0.11	(0.05, 0.22)	0.25

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	40	0.66	(0.52, 0.77)	0.88
f	f	14	0.23	(0.13, 0.35)	0.03
t	t	7	0.11	(0.05, 0.22)	0.09

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	10	0.16	(0.08, 0.28)	0.40
f	f	43	0.70	(0.57, 0.81)	0.34
u	u	8	0.13	(0.06, 0.24)	0.25

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	25	0.41	(0.29, 0.54)	0.48
l	l	28	0.46	(0.33, 0.59)	0.37
m	m	8	0.13	(0.06, 0.24)	0.15

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	26	0.43	(0.3, 0.56)	0.56
l	l	33	0.54	(0.41, 0.67)	0.35
n	n	2	0.03	(0.0, 0.11)	0.09

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	47	0.77	(0.65, 0.87)	0.81
l	l	10	0.16	(0.08, 0.28)	0.14
t	t	4	0.07	(0.02, 0.16)	0.06

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	21	0.34	(0.23, 0.48)	0.53
l	l	39	0.64	(0.51, 0.76)	0.37
u	u	1	0.02	(0.0, 0.09)	0.10

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	58	0.95	-	0.81
m	m	3	0.05	-	0.12
n	n	-	-	-	0.07

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	17	0.28	(0.17, 0.41)	0.40
m	m	24	0.39	(0.27, 0.53)	0.24
t	t	20	0.33	(0.21, 0.46)	0.36

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	52	0.85	(0.74, 0.93)	0.79
m	m	8	0.13	(0.06, 0.24)	0.13
u	u	1	0.02	(0.0, 0.09)	0.08

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	24	0.39	(0.27, 0.53)	0.46
n	n	13	0.21	(0.12, 0.34)	0.19
t	t	24	0.39	(0.27, 0.53)	0.36

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	56	0.92	(0.82, 0.97)	0.91
n	n	4	0.07	(0.02, 0.16)	0.05
u	u	1	0.02	(0.0, 0.09)	0.05

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
c	c	20	0.33	(0.21, 0.46)	0.46
t	t	34	0.56	(0.42, 0.68)	0.36
u	u	7	0.11	(0.05, 0.22)	0.19

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	14	0.23	(0.13, 0.35)	0.24
l	l	8	0.13	(0.06, 0.24)	0.14
m	m	39	0.64	(0.51, 0.76)	0.62

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	18	0.30	(0.19, 0.43)	0.24
l	l	5	0.08	(0.03, 0.18)	0.20
n	n	38	0.62	(0.49, 0.74)	0.55

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	16	0.26	(0.16, 0.39)	0.23
l	l	29	0.48	(0.35, 0.61)	0.54
t	t	16	0.26	(0.16, 0.39)	0.23

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	16	0.26	(0.16, 0.39)	0.20
l	l	5	0.08	(0.03, 0.18)	0.21
u	u	40	0.66	(0.52, 0.77)	0.58

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	61	1.00	-	0.75
m	m	-	-	-	0.17
n	n	-	-	-	0.09

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	7	0.11	(0.05, 0.22)	0.11
m	m	44	0.72	(0.59, 0.83)	0.81
t	t	10	0.16	(0.08, 0.28)	0.08

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	56	0.92	(0.82, 0.97)	0.73
m	m	2	0.03	(0.0, 0.11)	0.18
u	u	3	0.05	(0.01, 0.14)	0.09

Times New Roman (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
f	f	11	0.18	(0.09, 0.3)	0.13
n	n	38	0.62	(0.49, 0.74)	0.76
t	t	12	0.20	(0.11, 0.32)	0.11

Times New Roman (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
f	f	58	0.95	–	0.78
n	n	–	–	–	0.11
u	u	3	0.05	–	0.11

Times New Roman (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
f	f	19	0.31	(0.2, 0.44)	0.13
t	t	2	0.03	(0.0, 0.11)	0.11
u	u	40	0.66	(0.52, 0.77)	0.76

Times New Roman (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
l	l	59	0.97	–	0.70
m	m	2	0.03	–	0.19
n	n	–	–	–	0.11

Times New Roman (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
l	l	3	0.05	(0.01, 0.14)	0.14
m	m	42	0.69	(0.56, 0.8)	0.63
t	t	16	0.26	(0.16, 0.39)	0.23

Times New Roman (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
l	l	50	0.82	(0.7, 0.91)	0.72
m	m	3	0.05	(0.01, 0.14)	0.19
u	u	8	0.13	(0.06, 0.24)	0.09

Times New Roman (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
l	l	7	0.11	(0.05, 0.22)	0.21
n	n	39	0.64	(0.51, 0.76)	0.56
t	t	15	0.25	(0.14, 0.37)	0.24

Times New Roman (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
l	l	55	0.90	–	0.81
n	n	–	–	–	0.09
u	u	6	0.10	–	0.09

Times New Roman (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
l	l	12	0.20	(0.11, 0.32)	0.22
t	t	10	0.16	(0.08, 0.28)	0.19
u	u	39	0.64	(0.51, 0.76)	0.59

Times New Roman (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
m	m	2	0.03	–	0.17
n	n	–	–	–	0.08
t	t	59	0.97	–	0.75

Times New Roman (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
m	m	23	0.38	–	0.42
n	n	–	–	–	0.29
u	u	38	0.62	–	0.29

Times New Roman (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
m	m	2	0.03	(0.0, 0.11)	0.18
t	t	58	0.95	(0.86, 0.99)	0.73
u	u	1	0.02	(0.0, 0.09)	0.09

Times New Roman (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
n	n	2	0.03	(0.0, 0.11)	0.11
t	t	58	0.95	(0.86, 0.99)	0.79
u	u	1	0.02	(0.0, 0.09)	0.11

Times New Roman (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	2	0.03	(0.0, 0.11)	0.10
g	g	30	0.48	(0.35, 0.61)	0.34
h	h	31	0.49	(0.36, 0.62)	0.56

Verdana (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	6	0.10	(0.04, 0.2)	0.31
g	g	10	0.16	(0.08, 0.27)	0.28
o	o	47	0.75	(0.62, 0.85)	0.41

Verdana (Latin)

character		studies			model
		count	freq.	confidence interval	freq.
d	d	4	0.06	(0.02, 0.15)	0.13
g	g	5	0.08	(0.03, 0.18)	0.12
s	s	54	0.86	(0.75, 0.93)	0.75

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	2	0.03	(0.0, 0.11)	0.18
g	g	13	0.21	(0.11, 0.33)	0.22
u	u	48	0.76	(0.64, 0.86)	0.60

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	–	–	–	0.06
g	g	3	0.05	–	0.07
w	w	60	0.95	–	0.88

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	–	–	–	0.06
g	g	3	0.05	–	0.07
x	x	60	0.95	–	0.87

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	3	0.05	(0.01, 0.13)	0.05
h	h	18	0.29	(0.18, 0.41)	0.54
o	o	42	0.67	(0.54, 0.78)	0.41

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	5	0.08	(0.03, 0.18)	0.18
h	h	2	0.03	(0.0, 0.11)	0.15
s	s	56	0.89	(0.78, 0.95)	0.67

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	35	0.56	(0.42, 0.68)	0.51
h	h	7	0.11	(0.05, 0.22)	0.16
u	u	21	0.33	(0.22, 0.46)	0.33

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	2	0.03	(0.0, 0.11)	0.17
h	h	1	0.02	(0.0, 0.09)	0.06
w	w	60	0.95	(0.87, 0.99)	0.78

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	2	0.03	–	0.17
h	h	–	–	–	0.06
x	x	61	0.97	–	0.78

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	16	0.25	(0.15, 0.38)	0.24
o	o	5	0.08	(0.03, 0.18)	0.08
s	s	42	0.67	(0.54, 0.78)	0.68

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	22	0.35	(0.23, 0.48)	0.13
o	o	24	0.38	(0.26, 0.51)	0.30
u	u	17	0.27	(0.17, 0.4)	0.57

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	5	0.08	(0.03, 0.18)	0.20
o	o	5	0.08	(0.03, 0.18)	0.00
w	w	53	0.84	(0.73, 0.92)	0.80

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	7	0.11	(0.05, 0.22)	0.20
o	o	2	0.03	(0.0, 0.11)	0.00
x	x	54	0.86	(0.75, 0.93)	0.80

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	15	0.24	(0.14, 0.36)	0.26
s	s	44	0.70	(0.57, 0.81)	0.59
u	u	4	0.06	(0.02, 0.15)	0.16

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	19	0.30	(0.19, 0.43)	0.34
s	s	3	0.05	(0.01, 0.13)	0.26
w	w	41	0.65	(0.52, 0.77)	0.41

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	11	0.17	(0.09, 0.29)	0.59
s	s	3	0.05	(0.01, 0.13)	0.14
x	x	49	0.78	(0.66, 0.87)	0.27

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	22	0.35	–	0.25
u	u	–	–	–	0.06
w	w	41	0.65	–	0.69

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	10	0.16	(0.08, 0.27)	0.25
u	u	1	0.02	(0.0, 0.09)	0.06
x	x	52	0.83	(0.71, 0.91)	0.69

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
d	d	50	0.79	(0.67, 0.89)	0.71
w	w	3	0.05	(0.01, 0.13)	0.15
x	x	10	0.16	(0.08, 0.27)	0.13

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	10	0.16	(0.08, 0.27)	0.13
h	h	23	0.37	(0.25, 0.5)	0.65
o	o	30	0.48	(0.35, 0.61)	0.22

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	7	0.11	(0.05, 0.22)	0.25
h	h	7	0.11	(0.05, 0.22)	0.22
s	s	49	0.78	(0.66, 0.87)	0.52

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	48	0.76	(0.64, 0.86)	0.64
h	h	6	0.10	(0.04, 0.2)	0.23
u	u	9	0.14	(0.07, 0.25)	0.13

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	5	0.08	(0.03, 0.18)	0.26
h	h	2	0.03	(0.0, 0.11)	0.14
w	w	56	0.89	(0.78, 0.95)	0.60

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	4	0.06	(0.02, 0.15)	0.27
h	h	3	0.05	(0.01, 0.13)	0.14
x	x	56	0.89	(0.78, 0.95)	0.59

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	19	0.30	(0.19, 0.43)	0.23
o	o	8	0.13	(0.06, 0.23)	0.09
s	s	36	0.57	(0.44, 0.7)	0.68

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	24	0.38	(0.26, 0.51)	0.14
o	o	23	0.37	(0.25, 0.5)	0.25
u	u	16	0.25	(0.15, 0.38)	0.61

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	8	0.13	(0.06, 0.23)	0.20
o	o	3	0.05	(0.01, 0.13)	0.00
w	w	52	0.83	(0.71, 0.91)	0.80

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	8	0.13	(0.06, 0.23)	0.20
o	o	2	0.03	(0.0, 0.11)	0.00
x	x	53	0.84	(0.73, 0.92)	0.80

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	20	0.32	(0.21, 0.45)	0.27
s	s	37	0.59	(0.46, 0.71)	0.57
u	u	6	0.10	(0.04, 0.2)	0.17

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	17	0.27	(0.17, 0.4)	0.34
s	s	4	0.06	(0.02, 0.15)	0.26
w	w	42	0.67	(0.54, 0.78)	0.41

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	11	0.17	(0.09, 0.29)	0.59
s	s	3	0.05	(0.01, 0.13)	0.14
x	x	49	0.78	(0.66, 0.87)	0.27

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	17	0.27	–	0.28
u	u	–	–	–	0.06
w	w	46	0.73	–	0.66

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	9	0.14	(0.07, 0.25)	0.28
u	u	2	0.03	(0.0, 0.11)	0.06
x	x	52	0.83	(0.71, 0.91)	0.65

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
g	g	49	0.78	(0.66, 0.87)	0.72
w	w	3	0.05	(0.01, 0.13)	0.15
x	x	11	0.17	(0.09, 0.29)	0.13

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	30	0.48	(0.35, 0.61)	0.36
o	o	6	0.10	(0.04, 0.2)	0.22
s	s	27	0.43	(0.3, 0.56)	0.43

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	18	0.29	(0.18, 0.41)	0.17
o	o	44	0.70	(0.57, 0.81)	0.75
u	u	1	0.02	(0.0, 0.09)	0.08

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	13	0.21	(0.11, 0.33)	0.32
o	o	8	0.13	(0.06, 0.23)	0.20
w	w	42	0.67	(0.54, 0.78)	0.48

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	8	0.13	(0.06, 0.23)	0.32
o	o	7	0.11	(0.05, 0.22)	0.20
x	x	48	0.76	(0.64, 0.86)	0.48

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	8	0.13	(0.06, 0.23)	0.14
s	s	54	0.86	(0.75, 0.93)	0.80
u	u	1	0.02	(0.0, 0.09)	0.06

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	15	0.24	(0.14, 0.36)	0.30
s	s	8	0.13	(0.06, 0.23)	0.33
w	w	40	0.63	(0.5, 0.75)	0.37

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	19	0.30	(0.19, 0.43)	0.63
s	s	4	0.06	(0.02, 0.15)	0.18
x	x	40	0.63	(0.5, 0.75)	0.20

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	9	0.14	(0.07, 0.25)	0.07
u	u	2	0.03	(0.0, 0.11)	0.00
w	w	52	0.83	(0.71, 0.91)	0.93

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	8	0.13	–	0.07
u	u	–	–	–	0.00
x	x	55	0.87	–	0.93

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
h	h	45	0.71	(0.59, 0.82)	0.68
w	w	7	0.11	(0.05, 0.22)	0.17
x	x	11	0.17	(0.09, 0.29)	0.15

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
o	o	11	0.17	(0.09, 0.29)	0.23
s	s	39	0.62	(0.49, 0.74)	0.46
u	u	13	0.21	(0.11, 0.33)	0.31

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
o	o	10	0.16	(0.08, 0.27)	0.20
s	s	4	0.06	(0.02, 0.15)	0.33
w	w	49	0.78	(0.66, 0.87)	0.47

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
o	O	6	0.10	(0.04, 0.2)	0.49
s	S	3	0.05	(0.01, 0.13)	0.19
x	X	54	0.86	(0.75, 0.93)	0.32

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
o	O	14	0.22	(0.13, 0.34)	0.21
u	U	1	0.02	(0.0, 0.09)	0.26
w	W	48	0.76	(0.64, 0.86)	0.53

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
o	O	4	0.06	(0.02, 0.15)	0.22
u	U	3	0.05	(0.01, 0.13)	0.26
x	X	56	0.89	(0.78, 0.95)	0.52

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
o	O	46	0.73	(0.6, 0.83)	0.63
w	W	5	0.08	(0.03, 0.18)	0.20
x	X	12	0.19	(0.1, 0.31)	0.18

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
s	S	13	0.21	(0.11, 0.33)	0.36
u	U	5	0.08	(0.03, 0.18)	0.24
w	W	45	0.71	(0.59, 0.82)	0.40

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
s	S	7	0.11	(0.05, 0.22)	0.19
u	U	4	0.06	(0.02, 0.15)	0.60
x	X	52	0.83	(0.71, 0.91)	0.21

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
s	S	45	0.71	(0.59, 0.82)	0.50
w	W	6	0.10	(0.04, 0.2)	0.50
x	X	12	0.19	(0.1, 0.31)	0.01

Verdana (Latin)

		studies			model
character		count	freq.	confidence interval	freq.
u	U	40	0.63	(0.5, 0.75)	0.65
w	W	6	0.10	(0.04, 0.2)	0.18
x	X	17	0.27	(0.17, 0.4)	0.16

Verdana (Latin)

Appendix 3: similarity matrices

Similarity matrices for all the typefaces used in the studies described in chapter 5.

zhe	Ж	–							
short i	Й	0.36	–						
em	М	0.70	0.65	–					
er	Р	0.09	0.07	0.13	–				
es	С	0.16	0.17	0.24	0.62	–			
u	У	0.31	0.46	0.44	0.40	0.19	–		
ef	Ф	0.17	0.05	0.09	0.80	0.43	0.22	–	
yu	Ю	0.18	0.13	0.22	0.64	0.63	0.14	0.64	–
characters		ж	й	м	р	с	у	ф	ю
		zhe	short i	em	er	es	u	ef	yu

Arial (Cyrillic)

a	а	–							
ghe	г	0.16	–						
de	д	0.20	0.66	–					
io	ё	0.56	0.08	0.08	–				
zhe	ж	0.25	0.18	0.21	0.15	–			
ka	к	0.27	0.34	0.29	0.10	0.80	–		
pe	п	0.12	0.83	0.79	0.08	0.20	0.32	–	
ya	я	0.65	0.21	0.26	0.25	0.49	0.60	0.20	–
characters		а	г	д	ё	ж	к	п	я
		a	ghe	de	io	zhe	ka	pe	ya

Century Schoolbook (Cyrillic)

de	Д	–							
ka	К	0.39	–						
el	Л	0.80	0.50	–					
o	О	0.10	0.03	0.05	–				
es	С	0.10	0.14	0.10	0.80	–			
tse	Ц	0.80	0.44	0.67	0.07	0.14	–		
yu	Ю	0.16	0.37	0.15	0.57	0.44	0.15	–	
ya	Я	0.36	0.54	0.47	0.18	0.14	0.26	0.40	–
characters		Д	К	Л	О	С	Ц	Ю	Я
		de	ka	el	o	es	tse	yu	ya

Courier New (Cyrillic)

be	б	–							
ie	е	0.51	–						
ze	з	0.44	0.66	–					
i	и	0.04	0.26	0.17	–				
em	м	0.06	0.14	0.15	0.75	–			
en	н	0.07	0.17	0.19	0.87	0.76	–		
ef	ф	0.66	0.36	0.36	0.04	0.10	0.07	–	
che	ч	0.15	0.32	0.33	0.47	0.44	0.60	0.18	–
characters		б	е	з	и	м	н	ф	ч
		be	ie	ze	i	em	en	ef	che

Georgia (Cyrillic)

ghe	г	–							
ie	е	0.13	–						
o	о	0.13	0.80	–					
pe	п	0.77	0.11	0.14	–				
te	т	0.89	0.10	0.12	0.77	–			
ha	х	0.44	0.14	0.12	0.44	0.46	–		
shcha	щ	0.36	0.05	0.03	0.45	0.36	0.35	–	
e	э	0.11	0.88	0.69	0.10	0.09	0.17	0.12	–
characters		г	е	о	п	т	х	щ	э
		ghe	ie	o	pe	te	ha	shcha	e

PT Sans (Cyrillic)

ve	в	–							
ze	з	0.78	–						
el	л	0.24	0.20	–					
er	р	0.59	0.39	0.23	–				
te	т	0.24	0.17	0.66	0.19	–			
u	у	0.04	0.07	0.50	0.32	0.37	–		
ha	х	0.20	0.16	0.45	0.09	0.55	0.62	–	
e	э	0.57	0.81	0.15	0.42	0.12	0.11	0.09	–
characters		в	з	л	р	т	у	х	э
		ve	ze	el	er	te	u	ha	e

PT Serif (Cyrillic)

a	а	-							
be	б	0.69	-						
io	ё	0.61	0.60	-					
short i	й	0.10	0.12	0.32	-				
tse	ц	0.13	0.10	0.07	0.51	-			
sha	ш	0.12	0.11	0.06	0.46	0.73	-		
shcha	щ	0.08	0.06	0.09	0.47	0.86	0.87	-	
hard sign	ъ	0.42	0.26	0.22	0.26	0.39	0.28	0.34	-
characters		а	б	ё	й	ц	ш	щ	ъ
		a	be	io	short i	tse	sha	shcha	hard sign

Times New Roman (Cyrillic)

a	а	-							
ve	в	0.54	-						
i	и	0.24	0.20	-					
ha	х	0.10	0.16	0.63	-				
che	ч	0.43	0.33	0.35	0.27	-			
hard sign	ъ	0.40	0.47	0.11	0.08	0.24	-		
yeru	ы	0.37	0.43	0.22	0.04	0.21	0.57	-	
soft sign	ь	0.43	0.56	0.15	0.07	0.26	0.83	0.65	-
characters		а	в	и	х	ч	ъ	ы	ь
		a	ve	i	ha	che	hard sign	yeru	soft sign

Verdana (Cyrillic)

A	अ	-							
Ka	ख	0.24	-						
Gha	घ	0.35	0.28	-					
Ta	त	0.18	0.24	0.32	-				
Ba	ब	0.15	0.29	0.40	0.34	-			
Ra	र	0.21	0.51	0.22	0.50	0.16	-		
Ssa	ष	0.14	0.18	0.48	0.33	0.83	0.15	-	
Sa	स	0.36	0.66	0.32	0.42	0.20	0.55	0.31	-
characters		अ	ख	घ	त	ब	र	ष	स
		A	Ka	Gha	Ta	Ba	Ra	Ssa	Sa

Adobe Devanagari (Devanagari)

U	उ	-							
Gha	घ	0.35	-						
Ca	च	0.12	0.47	-					
Cha	छ	0.30	0.57	0.17	-				
Tta	ट	0.34	0.14	0.22	0.15	-			
Dda	ड	0.53	0.20	0.23	0.19	0.65	-		
Dha	ध	0.31	0.77	0.29	0.63	0.17	0.18	-	
Ha	ह	0.31	0.25	0.20	0.30	0.47	0.66	0.19	-
characters		उ	घ	च	छ	ट	ड	ध	ह
		U	Gha	Ca	Cha	Tta	Dda	Dha	Ha

Devanagari MT (Devanagari)

E	ए	-							
Ka	क	0.15	-						
Ja	ज	0.14	0.55	-					
Ddha	ढ	0.18	0.49	0.33	-				
Na	न	0.15	0.32	0.62	0.22	-			
Bha	भ	0.17	0.25	0.41	0.21	0.55	-		
Ma	म	0.27	0.29	0.43	0.16	0.76	0.78	-	
Sha	श	0.21	0.32	0.26	0.23	0.20	0.46	0.22	-
characters		ए	क	ज	ढ	न	भ	म	श
		E	Ka	Ja	Ddha	Na	Bha	Ma	Sha

Ek Mukta (Devanagari)

Ca	च	-							
Jha	झ	0.31	-						
Nya	ञ	0.69	0.39	-					
Ttha	ठ	0.43	0.15	0.31	-				
Bha	भ	0.33	0.39	0.23	0.17	-			
Ma	म	0.30	0.39	0.21	0.18	0.84	-		
Ra	र	0.21	0.12	0.13	0.25	0.28	0.26	-	
Va	व	0.73	0.16	0.54	0.63	0.20	0.24	0.25	-
characters		च	झ	ञ	ठ	भ	म	र	व
		Ca	Jha	Nya	Ttha	Bha	Ma	Ra	Va

ITF Devanagari (Devanagari)

I	इ	-							
Ga	ग	0.08	-						
Dda	ड	0.86	0.09	-					
Na	न	0.10	0.80	0.11	-				
Pa	प	0.11	0.63	0.18	0.55	-			
Ba	ब	0.20	0.21	0.23	0.28	0.52	-		
Sa	स	0.17	0.51	0.16	0.43	0.53	0.31	-	
Ha	ह	0.75	0.08	0.75	0.11	0.13	0.21	0.23	-
characters		इ	ग	ड	न	प	ब	स	ह
		I	Ga	Dda	Na	Pa	Ba	Sa	Ha

Kohinoor Devanagari (Devanagari)

Ka	क	-							
Ga	ग	0.12	-						
Nna	ण	0.19	0.47	-					
Tha	थ	0.25	0.16	0.19	-				
Pa	प	0.20	0.33	0.61	0.33	-			
Pha	फ	0.70	0.14	0.38	0.26	0.61	-		
Ya	य	0.23	0.28	0.34	0.69	0.65	0.37	-	
La	ल	0.55	0.11	0.17	0.24	0.21	0.33	0.23	-
characters		क	ग	ण	थ	प	फ	य	ल
		Ka	Ga	Nna	Tha	Pa	Pha	Ya	La

Lohit Devanagari (Devanagari)

Ddha	ढ	-							
Nna	ण	0.10	-						
Ta	त	0.40	0.22	-					
Tha	थ	0.27	0.21	0.35	-				
Da	द	0.83	0.07	0.53	0.29	-			
Ya	य	0.18	0.30	0.41	0.79	0.24	-		
Sha	श	0.17	0.44	0.14	0.47	0.14	0.35	-	
Ssa	ष	0.11	0.69	0.30	0.34	0.10	0.54	0.33	-
characters		ढ	ण	त	थ	द	य	श	ष
		Ddha	Nna	Ta	Tha	Da	Ya	Sha	Ssa

Murty Hindi (Devanagari)

U	उ	-							
Nga	ङ	0.34	-						
Cha	छ	0.47	0.24	-					
Tta	ट	0.19	0.33	0.12	-				
Ttha	ठ	0.17	0.22	0.17	0.78	-			
Da	द	0.19	0.33	0.21	0.78	0.58	-		
Dha	ध	0.41	0.17	0.65	0.19	0.12	0.24	-	
La	ल	0.62	0.22	0.47	0.24	0.15	0.27	0.46	-
characters		उ	ङ	छ	ट	ठ	द	ध	ल
		U	Nga	Cha	Tta	Ttha	Da	Dha	La

Nirmala UI (Devanagari)

a	a	-							
e	e	0.89	-						
l	l	0.04	0.06	-					
p	p	0.47	0.48	0.21	-				
s	s	0.71	0.71	0.11	0.27	-			
t	t	0.09	0.08	0.79	0.21	0.08	-		
y	y	0.18	0.19	0.24	0.59	0.18	0.40	-	
z	z	0.35	0.37	0.25	0.17	0.65	0.18	0.40	-
characters		a	e	l	p	s	t	y	z
		a	e	l	p	s	t	y	z

Arial (Latin)

d	d	-							
i	i	0.16	-						
j	j	0.22	0.95	-					
o	o	0.73	0.09	0.12	-				
p	p	0.94	0.14	0.27	0.72	-			
r	r	0.36	0.53	0.43	0.24	0.44	-		
v	v	0.15	0.24	0.15	0.28	0.14	0.31	-	
x	x	0.08	0.18	0.10	0.21	0.10	0.23	0.85	-
characters		d	i	j	o	p	r	v	x
		d	i	j	o	p	r	v	x

Calibri (Latin)

b	b	-							
f	f	0.41	-						
i	i	0.14	0.50	-					
m	m	0.26	0.22	0.32	-				
n	n	0.32	0.20	0.31	0.98	-			
p	p	0.79	0.30	0.15	0.32	0.34	-		
q	q	0.79	0.27	0.11	0.26	0.30	0.91	-	
z	z	0.07	0.08	0.23	0.28	0.31	0.10	0.06	-
characters		b	f	i	m	n	p	q	z
		b	f	i	m	n	p	q	z

Cambria (Latin)

b	b	-							
d	d	0.97	-						
i	i	0.13	0.13	-					
j	j	0.10	0.13	0.92	-				
k	k	0.58	0.53	0.12	0.10	-			
l	l	0.28	0.25	0.75	0.63	0.23	-		
v	v	0.19	0.18	0.16	0.14	0.48	0.16	-	
y	y	0.19	0.20	0.10	0.34	0.40	0.08	0.87	-
characters		b	d	i	j	k	l	v	y
		b	d	i	j	k	l	v	y

Candara (Latin)

d	d	-							
i	i	0.23	-						
j	j	0.29	0.88	-					
o	o	0.72	0.13	0.13	-				
p	p	0.94	0.22	0.32	0.68	-			
r	r	0.33	0.53	0.41	0.28	0.42	-		
v	v	0.12	0.24	0.13	0.21	0.10	0.38	-	
x	x	0.06	0.18	0.08	0.14	0.08	0.27	0.84	-
characters		d	i	j	o	p	r	v	x
		d	i	j	o	p	r	v	x

Century Schoolbook (Latin)

c	c	-							
e	e	0.81	-						
f	f	0.08	0.11	-					
h	h	0.19	0.13	0.73	-				
k	k	0.07	0.05	0.59	0.73	-			
o	o	0.84	0.83	0.05	0.14	0.05	-		
w	w	0.26	0.20	0.22	0.27	0.53	0.18	-	
y	y	0.06	0.05	0.47	0.47	0.58	0.04	0.58	-
characters		c	e	f	h	k	o	w	y
		c	e	f	h	k	o	w	y

Courier New (Latin)

g	g	-							
h	h	0.30	-						
m	m	0.20	0.58	-					
n	n	0.17	0.69	0.91	-				
q	q	0.95	0.39	0.18	0.18	-			
r	r	0.19	0.50	0.68	0.74	0.18	-		
y	y	0.44	0.16	0.04	0.02	0.46	0.06	-	
z	z	0.03	0.08	0.23	0.25	0.03	0.24	0.47	-
characters		g	h	m	n	q	r	y	z
		g	h	m	n	q	r	y	z

Futura (Latin)

b	b	-							
d	d	0.98	-						
i	i	0.13	0.20	-					
j	j	0.15	0.13	0.83	-				
k	k	0.60	0.58	0.13	0.08	-			
l	l	0.47	0.48	0.61	0.36	0.46	-		
v	v	0.13	0.13	0.20	0.09	0.40	0.12	-	
y	y	0.10	0.10	0.09	0.53	0.32	0.08	0.85	-
characters		b	d	i	j	k	l	v	y
		b	d	i	j	k	l	v	y

Georgia (Latin)

b	b	-							
f	f	0.41	-						
i	i	0.24	0.49	-					
n	n	0.53	0.35	0.24	-				
p	p	0.79	0.34	0.20	0.56	-			
q	q	0.83	0.33	0.17	0.48	0.85	-		
w	w	0.08	0.08	0.20	0.23	0.09	0.07	-	
z	z	0.13	0.16	0.23	0.28	0.12	0.13	0.75	-
characters		b	f	i	n	p	q	w	z
		b	f	i	n	p	q	w	z

PT Sans (Latin)

g	g	-							
h	h	0.16	-						
m	m	0.14	0.70	-					
n	n	0.11	0.76	0.93	-				
q	q	0.54	0.40	0.27	0.29	-			
r	r	0.12	0.46	0.58	0.60	0.23	-		
y	y	0.34	0.23	0.13	0.17	0.39	0.24	-	
z	z	0.06	0.15	0.21	0.26	0.06	0.33	0.48	-
characters		g	h	m	n	q	r	y	z
		g	h	m	n	q	r	y	z

PT Serif (Latin)

a	a	-							
c	c	0.63	-						
f	f	0.14	0.14	-					
l	l	0.02	0.08	0.61	-				
m	m	0.31	0.22	0.16	0.23	-			
n	n	0.36	0.33	0.14	0.25	0.91	-		
t	t	0.16	0.21	0.64	0.63	0.12	0.17	-	
u	u	0.47	0.40	0.07	0.18	0.72	0.84	0.20	-
characters		a	c	f	l	m	n	t	u
		a	c	f	l	m	n	t	u

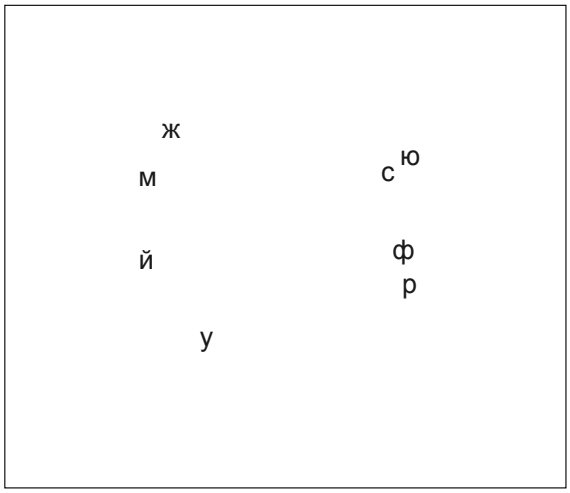
Times New Roman (Latin)

d	d	-							
g	g	0.79	-						
h	h	0.71	0.53	-					
o	o	0.51	0.49	0.35	-				
s	s	0.28	0.31	0.26	0.48	-			
u	u	0.48	0.44	0.76	0.55	0.40	-		
w	w	0.06	0.05	0.10	0.11	0.21	0.24	-	
x	x	0.03	0.03	0.06	0.09	0.15	0.12	0.73	-
characters		d	g	h	o	s	u	w	x
		d	g	h	o	s	u	w	x

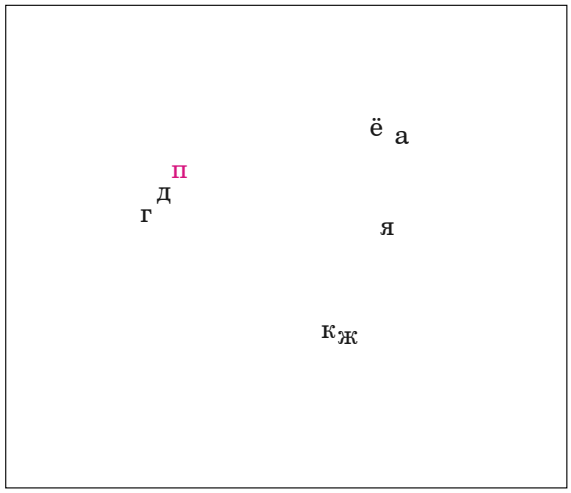
Verdana (Latin)

Appendix 4: cluster plots

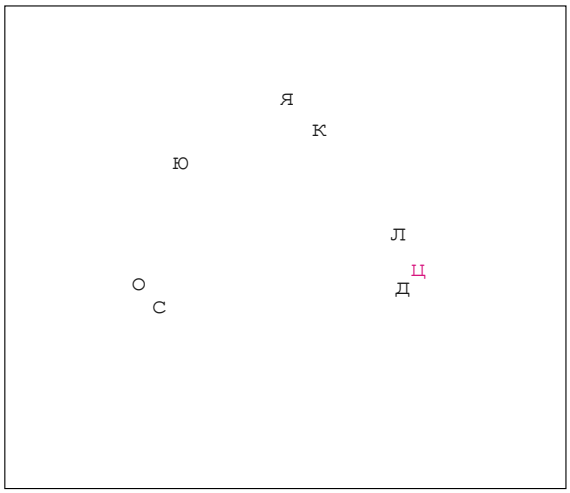
Plots with character clusters based on the generalized pair similarity measures derived from the similarity matrices in appendix 3. For transliteration, refer to the corresponding similarity matrix. Characters in magenta have been shifted slightly from their default position to avoid their collision with other characters in the plot.



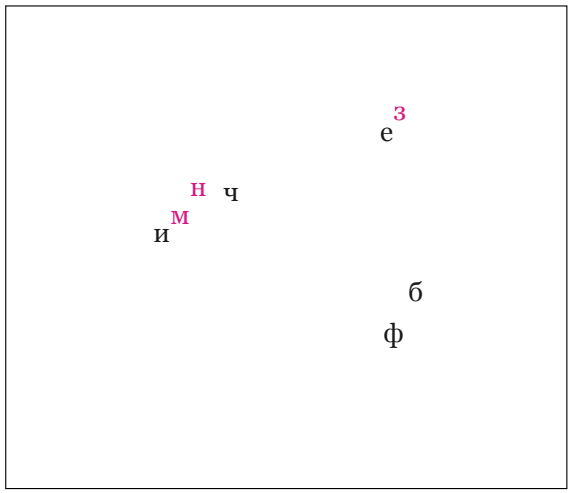
Arial (Cyrillic)



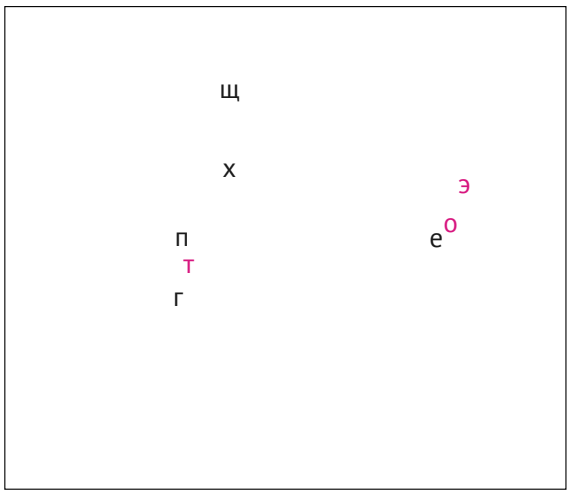
Century Schoolbook (Cyrillic)



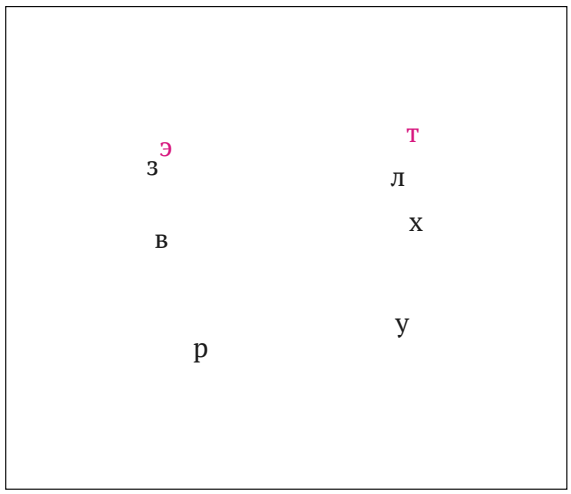
Courier New (Cyrillic)



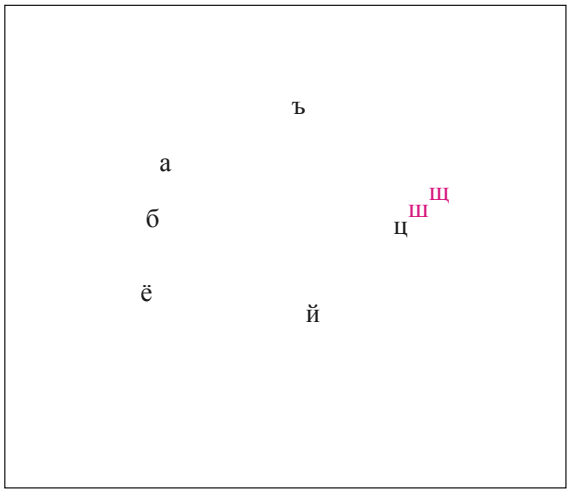
Georgia (Cyrillic)



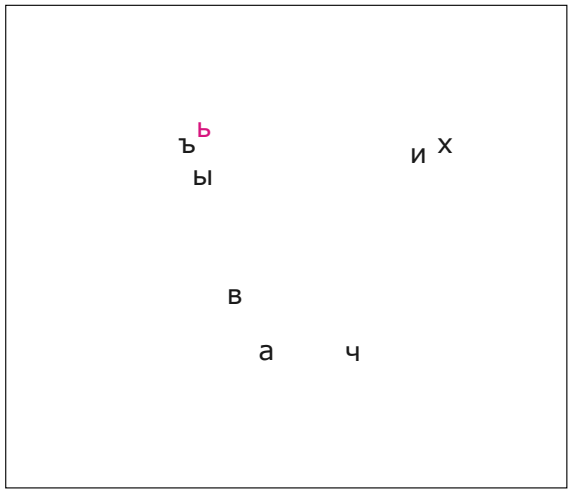
PT Sans (Cyrillic)



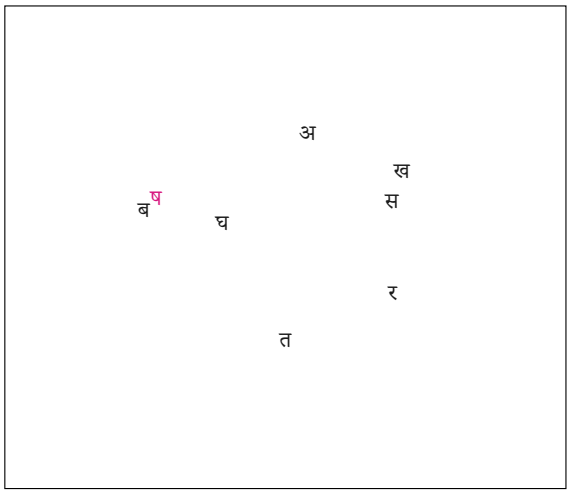
PT Serif (Cyrillic)



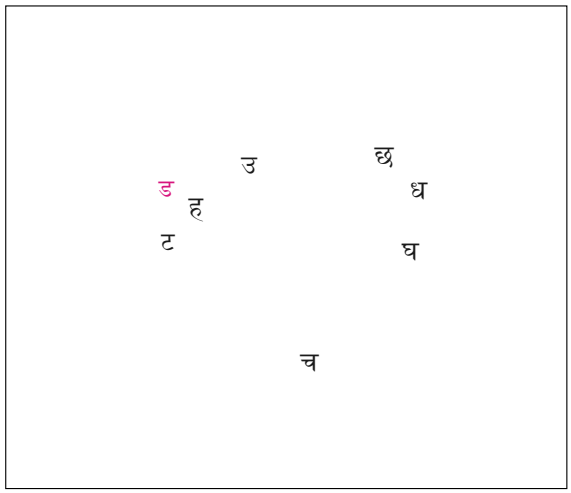
Times New Roman (Cyrillic)



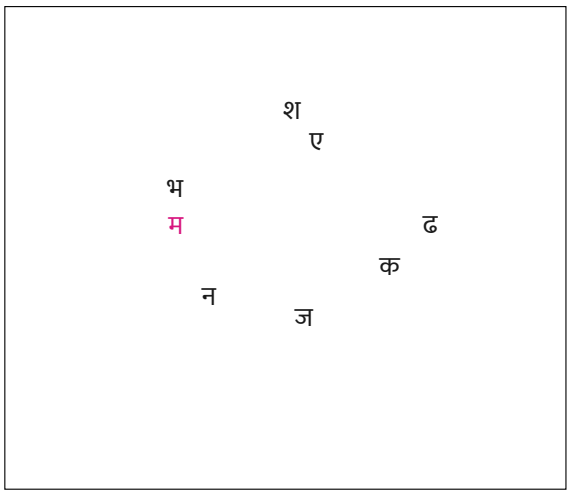
Verdana (Cyrillic)



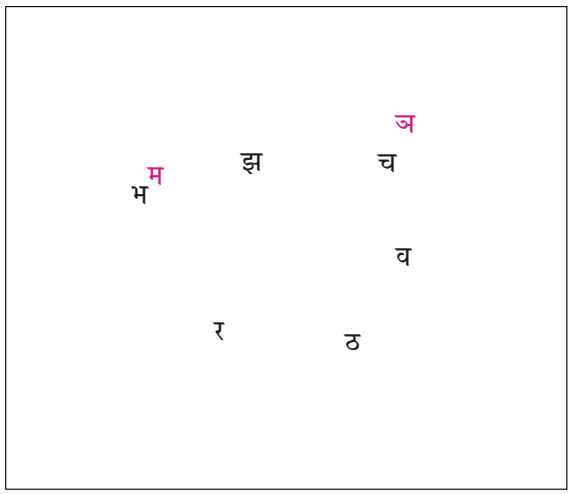
Adobe Devanagari (Devanagari)



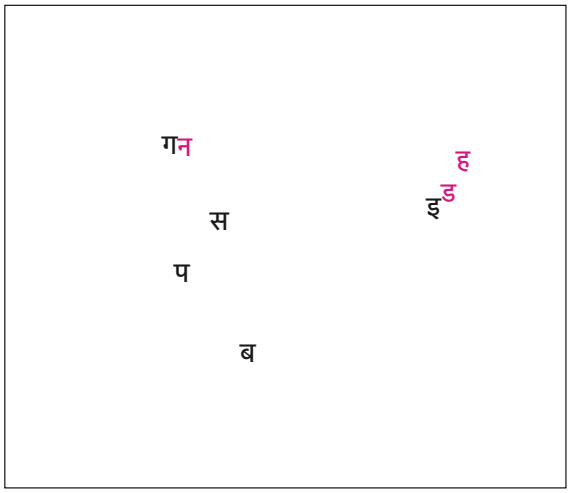
Devanagari MT (Devanagari)



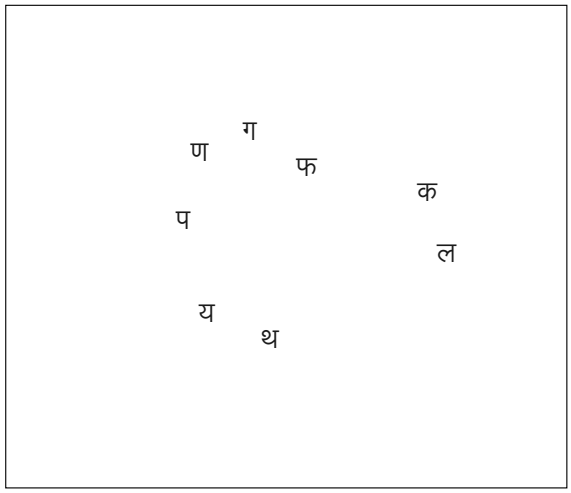
Ek Mukta (Devanagari)



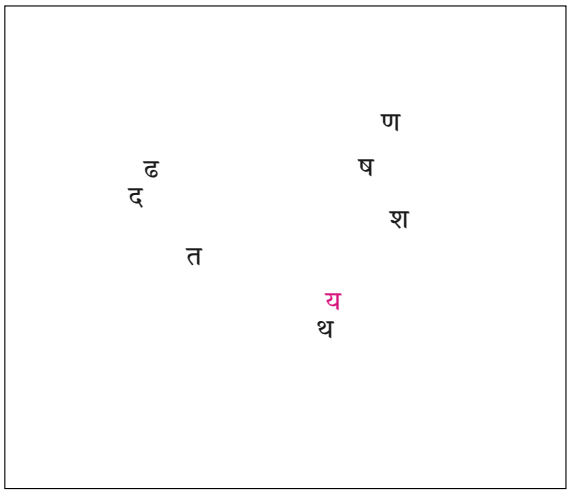
ITF Devanagari (Devanagari)



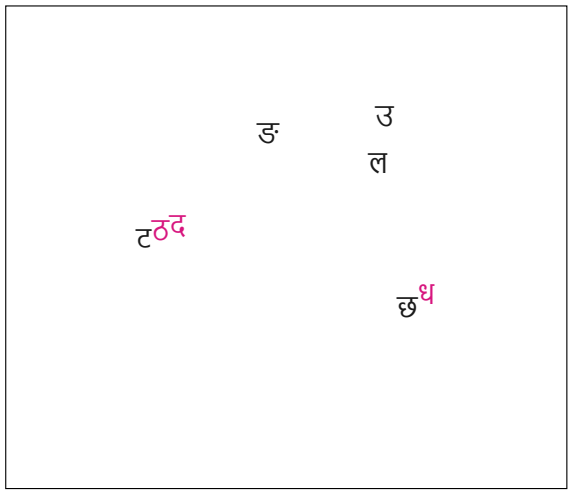
Kohinoor Devanagari (Devanagari)



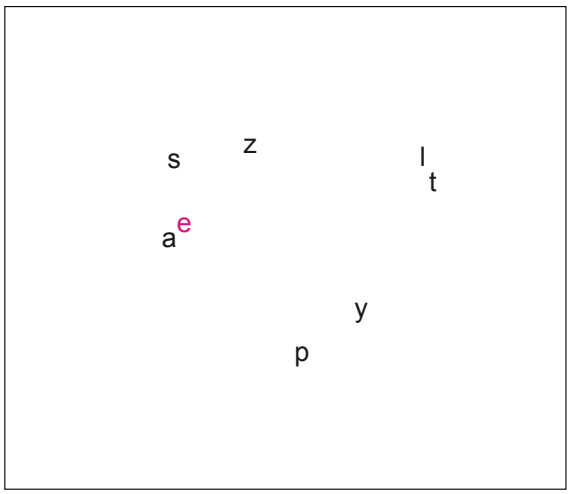
Lohit Devanagari (Devanagari)



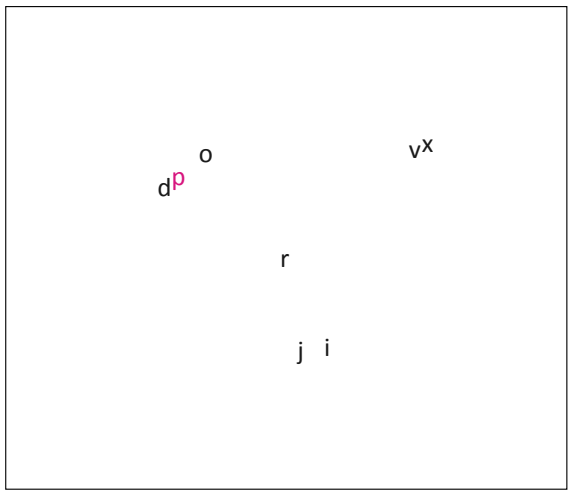
Murty Hindi (Devanagari)



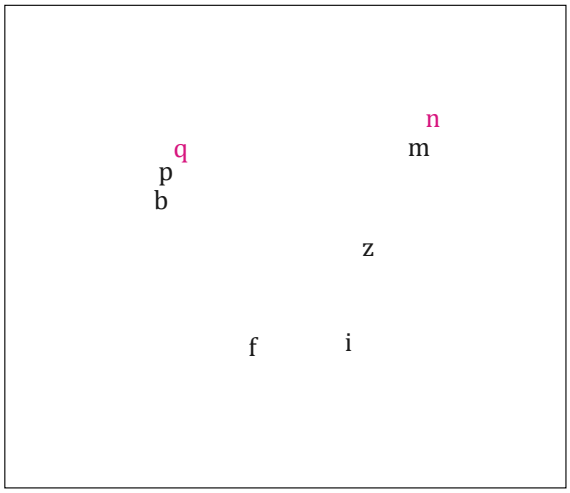
Nirmala UI (Devanagari)



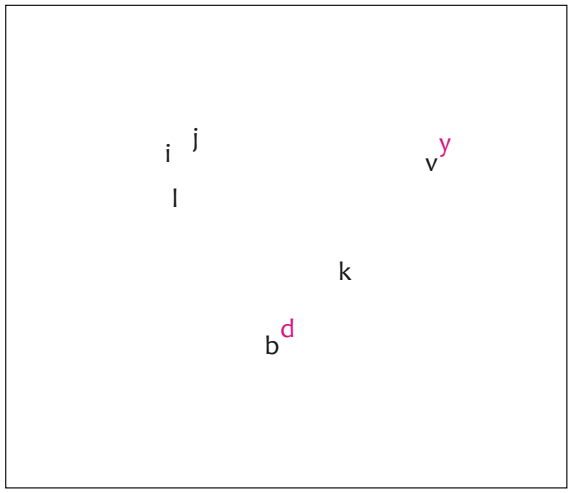
Arial (Latin)



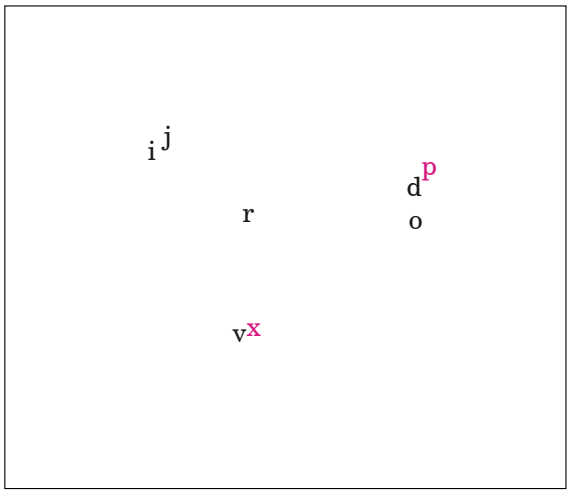
Calibri (Latin)



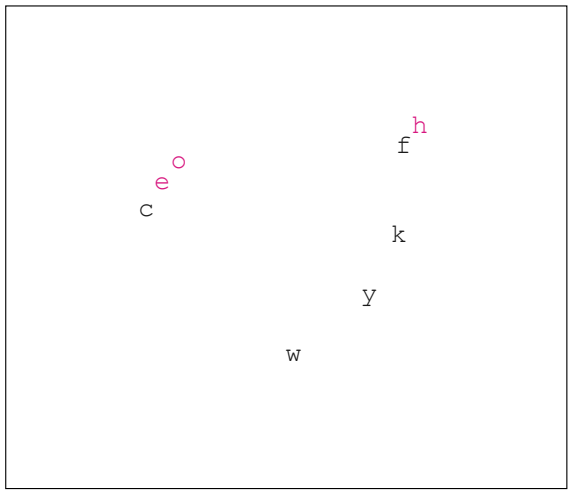
Cambria (Latin)



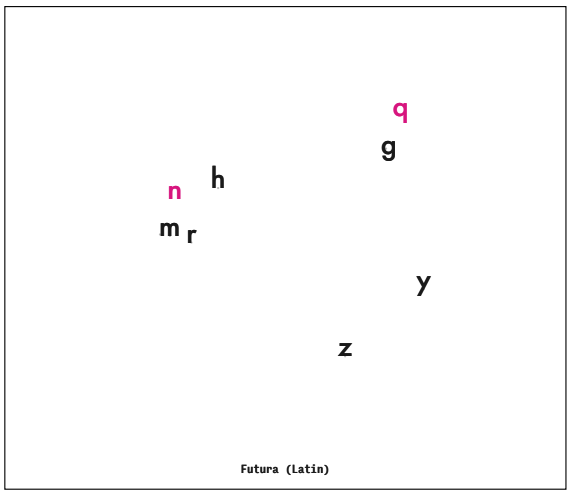
Candara (Latin)



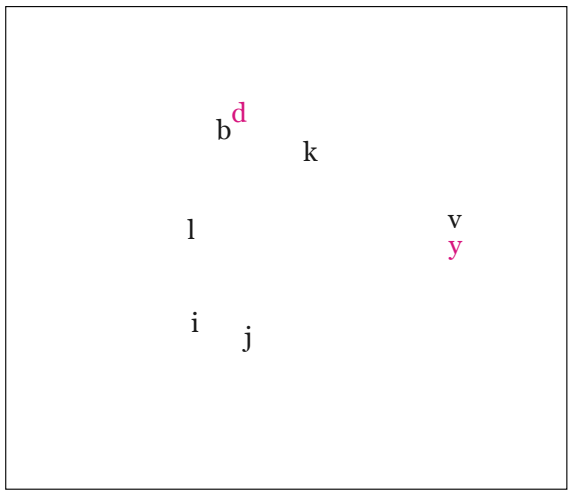
Century Schoolbook (Latin)



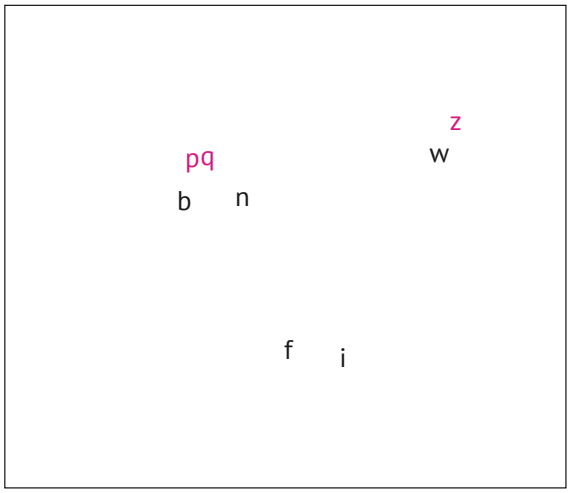
Courier New (Latin)



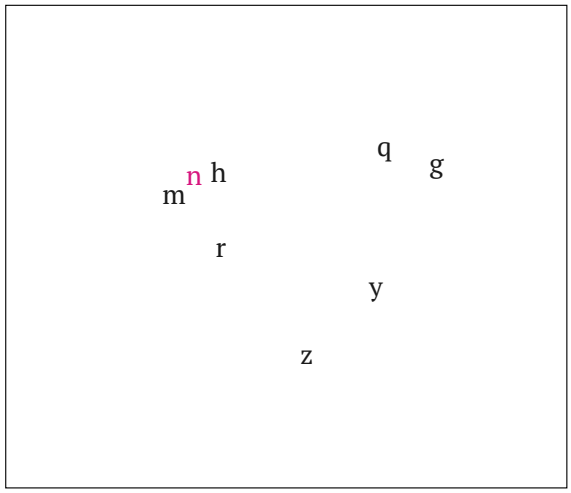
Futura (Latin)



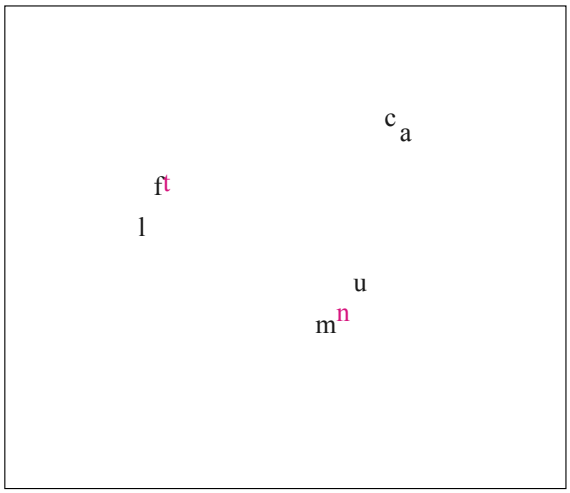
Georgia (Latin)



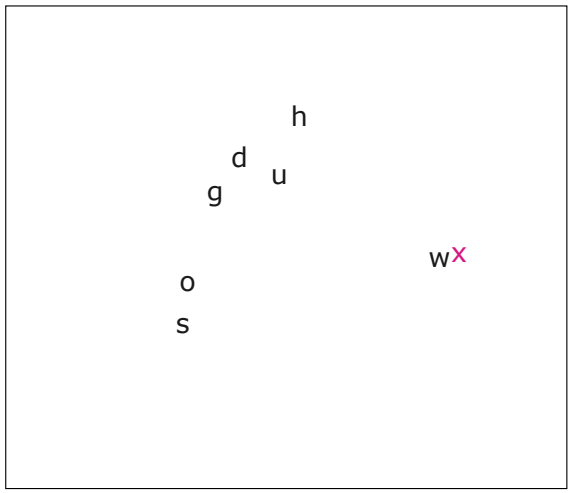
PT Sans (Latin)



PT Serif (Latin)



Times New Roman (Latin)



Verdana (Latin)

Appendix 5: distinctiveness measures

Character distinctiveness measures for all the typefaces used in the studies described in chapter 5.

character		score
zhe	ж	0.44
short i	й	0.46
em	м	0.31
er	р	0.21
es	с	0.30
u	у	0.38
ef	ф	0.31
yu	ю	0.26

Arial
(Cyrillic)

character		score
a	а	0.37
ghe	г	0.30
de	д	0.30
io	ё	0.63
zhe	ж	0.35
ka	к	0.23
pe	п	0.27
ya	я	0.24

Century Schoolbook
(Cyrillic)

character		score
de	Д	0.23
ka	К	0.31
el	Л	0.22
o	О	0.48
es	С	0.49
tse	Ц	0.28
yu	Ю	0.36
ya	Я	0.33

Courier New
(Cyrillic)

character		score
be	б	0.44
ie	е	0.31
ze	з	0.34
i	и	0.25
em	м	0.33
en	н	0.29
ef	ф	0.49
che	ч	0.30

Georgia
(Cyrillic)

character		score
ghe	г	0.19
ie	е	0.38
o	о	0.42
pe	п	0.22
te	т	0.20
ha	х	0.39
shcha	щ	0.51
e	э	0.38

PT Sans
(Cyrillic)

character		score
ve	в	0.24
ze	з	0.28
el	л	0.31
er	р	0.36
te	т	0.34
u	у	0.42
ha	х	0.38
e	э	0.35

PT Serif
(Cyrillic)

character		score
a	а	0.38
be	б	0.44
io	ё	0.44
short i	й	0.36
tse	ц	0.21
sha	ш	0.27
shcha	щ	0.22
hard sign	ъ	0.38

Times New Roman
(Cyrillic)

character		score
a	а	0.28
ve	в	0.23
i	и	0.48
ha	х	0.61
che	ч	0.40
hard sign	ъ	0.23
yeru	ы	0.29
soft sign	ь	0.17

Verdana
(Cyrillic)

character		score
A	अ	0.54
Ka	ख	0.31
Gha	घ	0.32
Ta	त	0.33
Ba	ब	0.32
Ra	र	0.34
Ssa	ष	0.31
Sa	स	0.19

Adobe Devanagari
(Devanagari)

character		score
U	उ	0.36
Gha	घ	0.22
Ca	च	0.51
Cha	छ	0.34
Tta	ट	0.39
Dda	ड	0.25
Dha	ध	0.28
Ha	ह	0.32

Devanagari MT
(Devanagari)

character		score
E	ए	0.64
Ka	क	0.32
Ja	ज	0.23
Ddha	ढ	0.48
Na	न	0.19
Bha	भ	0.19
Ma	म	0.17
Sha	श	0.46

Ek Mukta
(Devanagari)

character		score
Ca	च	0.15
Jha	झ	0.45
Nya	ञ	0.28
Ttha	ठ	0.39
Bha	भ	0.30
Ma	म	0.31
Ra	र	0.57
Va	व	0.21

ITF Devanagari
(Devanagari)

character		score
I	इ	0.35
Ga	ग	0.32
Dda	ड	0.32
Na	न	0.32
Pa	प	0.24
Ba	ब	0.44
Sa	स	0.33
Ha	ह	0.35

Kohinoor Devanagari
(Devanagari)

character		score
Ka	क	0.36
Ga	ग	0.54
Nna	ण	0.33
Tha	थ	0.40
Pa	प	0.16
Pha	फ	0.20
Ya	य	0.20
La	ल	0.47

Lohit Devanagari
(Devanagari)

character		score
Ddha	ढ	0.41
Nna	ण	0.42
Ta	त	0.33
Tha	थ	0.22
Da	द	0.37
Ya	य	0.20
Sha	श	0.42
Ssa	ष	0.31

Murty Hindi
(Devanagari)

character		score
U	उ	0.32
Nga	ङ	0.47
Cha	छ	0.34
Tta	ट	0.26
Ttha	ठ	0.37
Da	द	0.26
Dha	ध	0.36
La	ल	0.30

Nirmala UI
(Devanagari)

character		score
a	a	0.22
e	e	0.20
l	l	0.51
p	p	0.32
s	s	0.25
t	t	0.48
y	y	0.38
z	z	0.32

Arial
(Latin)

character		score
d	d	0.25
i	i	0.39
j	j	0.36
o	o	0.35
p	p	0.23
r	r	0.27
v	v	0.40
x	x	0.50

Calibri
(Latin)

character		score
b	b	0.23
f	f	0.44
i	i	0.50
m	m	0.26
n	n	0.28
p	p	0.20
q	q	0.23
z	z	0.68

Cambria
(Latin)

character		score
b	b	0.34
d	d	0.35
i	i	0.40
j	j	0.35
k	k	0.30
l	l	0.37
v	v	0.42
y	y	0.37

Candara
(Latin)

character		score
d	d	0.26
i	i	0.35
j	j	0.36
o	o	0.34
p	p	0.22
r	r	0.25
v	v	0.44
x	x	0.53

Century Schoolbook
(Latin)

character		score
c	c	0.37
e	e	0.40
f	f	0.36
h	h	0.25
k	k	0.26
o	o	0.39
w	w	0.36
y	y	0.35

Courier New
(Latin)

character		score
g	g	0.37
h	h	0.24
m	m	0.24
n	n	0.25
q	q	0.34
r	r	0.27
y	y	0.53
z	z	0.62

Futura
(Latin)

character		score
b	b	0.33
d	d	0.32
i	i	0.38
j	j	0.38
k	k	0.28
l	l	0.28
v	v	0.47
y	y	0.41

Georgia
(Latin)

character		score
b	b	0.16
f	f	0.39
i	i	0.50
n	n	0.26
p	p	0.17
q	q	0.19
w	w	0.57
z	z	0.49

PT Sans
(Latin)

character		score
g	g	0.58
h	h	0.20
m	m	0.16
n	n	0.16
q	q	0.38
r	r	0.27
y	y	0.44
z	z	0.56

PT Serif
(Latin)

character		score
a	a	0.40
c	c	0.43
f	f	0.48
l	l	0.43
m	m	0.25
n	n	0.25
t	t	0.41
u	u	0.19

Times New Roman
(Latin)

character		score
d	d	0.20
g	g	0.24
h	h	0.22
o	o	0.26
s	s	0.40
u	u	0.17
w	w	0.57
x	x	0.65

Verdana
(Latin)

Appendix 6: typeface comparisons

Juxtaposed response frequencies (column “freq.”) that occured in two different typefaces from the studies described in chapter 5. The p-value provided is from Fisher’s exact test on corresponding response counts which are shown in appendix 2.

Century Schoolbook			Courier New		
character		freq.	freq.	character	
de	д	0.65	0.43	д	de
ka	к	0.23	0.25	к	ka
ya	я	0.13	0.31	я	ya

p-value = 0.024 (typeface has an effect)

PT Sans			PT Serif		
character		freq.	freq.	character	
e	э	0.73	0.73	э	e
ha	х	0.13	0.15	х	ha
te	т	0.15	0.12	т	te

p-value = 1.000 (typeface has no effect)

Devanagari MT			Nirmala UI		
character		freq.	freq.	character	
Cha	छ	0.13	0.25	छ	Cha
Dha	ध	0.11	0.08	ध	Dha
Tta	ट	0.76	0.67	ट	Tta

p-value = 0.215 (typeface has no effect)

Devanagari MT			Nirmala UI		
character		freq.	freq.	character	
Cha	छ	0.18	0.18	छ	Cha
Dha	ध	0.13	0.20	ध	Dha
U	उ	0.69	0.63	उ	U

p-value = 0.595 (typeface has no effect)

Devanagari MT			Nirmala UI		
character		freq.	freq.	character	
Cha	छ	0.44	0.29	छ	Cha
Tta	ट	0.38	0.59	ट	Tta
U	उ	0.18	0.12	उ	U

p-value = 0.123 (typeface has no effect)

Devanagari MT			Nirmala UI		
character		freq.	freq.	character	
Dha	ध	0.29	0.18	ध	Dha
Tta	ट	0.53	0.55	ट	Tta
U	उ	0.18	0.27	उ	U

p-value = 0.318 (typeface has no effect)

Lohit Devanagari			Murty Hindi		
character		freq.	freq.	character	
Nna	ण	0.70	0.84	ण	Nna
Tha	थ	0.25	0.08	थ	Tha
Ya	य	0.05	0.08	य	Ya

p-value = 0.025 (typeface has an effect)

Arial			Times New Roman		
character		freq.	freq.	character	
a	a	0.84	0.77	a	a
l	l	0.12	0.23	l	l
t	t	0.04	0.00	t	t

p-value = 0.057 (typeface has no effect)

Calibri			Candara		
character		freq.	freq.	character	
d	d	0.96	0.97	d	d
i	i	0.01	0.02	i	i
j	j	0.03	0.02	j	j

p-value = 1.000 (typeface has no effect)

Calibri			Candara		
character		freq.	freq.	character	
d	d	0.34	0.34	d	d
i	i	0.34	0.36	i	i
v	v	0.32	0.30	v	v

p-value = 0.881 (typeface has no effect)

Calibri			Candara		
character		freq.	freq.	character	
d	d	0.22	0.26	d	d
j	j	0.30	0.36	j	j
v	v	0.48	0.38	v	v

p-value = 0.452 (typeface has no effect)

Calibri			Candara		
character		freq.	freq.	character	
i	i	0.00	0.00	i	i
j	j	0.04	0.05	j	j
v	v	0.96	0.95	v	v

p-value = 1.000 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.96	0.90	d	d
i	i	0.01	0.00	i	i
j	j	0.03	0.10	j	j

p-value = 0.098 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.06	0.05	d	d
i	i	0.86	0.83	i	i
o	o	0.08	0.12	o	o

p-value = 0.791 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.03	0.00	d	d
i	i	0.97	0.95	i	i
p	p	0.00	0.05	p	p

p-value = 0.060 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.56	0.56	d	d
i	i	0.40	0.36	i	i
r	r	0.04	0.08	r	r

p-value = 0.426 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.34	0.34	d	d
i	i	0.34	0.24	i	i
v	v	0.32	0.42	v	v

p-value = 0.292 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.29	0.24	d	d
i	i	0.19	0.15	i	i
x	x	0.52	0.61	x	x

p-value = 0.546 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.06	0.02	d	d
j	j	0.86	0.86	j	j
o	o	0.08	0.12	o	o

p-value = 0.214 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.05	0.03	d	d
j	j	0.94	0.95	j	j
p	p	0.01	0.02	p	p

p-value = 1.000 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.40	0.53	d	d
j	j	0.45	0.34	j	j
r	r	0.14	0.14	r	r

p-value = 0.267 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.22	0.12	d	d
j	j	0.30	0.19	j	j
v	v	0.48	0.69	v	v

p-value = 0.022 (typeface has an effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.17	0.12	d	d
j	j	0.22	0.14	j	j
x	x	0.61	0.75	x	x

p-value = 0.175 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.08	0.02	d	d
o	o	0.88	0.92	o	o
p	p	0.04	0.07	p	p

p-value = 0.094 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.10	0.10	d	d
o	o	0.10	0.07	o	o
r	r	0.79	0.83	r	r

p-value = 0.750 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.14	0.17	d	d
o	o	0.00	0.03	o	o
v	v	0.86	0.80	v	v

p-value = 0.251 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.05	0.05	d	d
o	o	0.00	0.03	o	o
x	x	0.95	0.92	x	x

p-value = 0.293 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.08	0.05	d	d
p	p	0.01	0.03	p	p
r	r	0.91	0.92	r	r

p-value = 0.617 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.04	0.05	d	d
p	p	0.01	0.05	p	p
v	v	0.95	0.90	v	v

p-value = 0.353 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.01	0.00	d	d
p	p	0.01	0.00	p	p
x	x	0.97	1.00	x	x

p-value = 0.497 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.35	0.44	d	d
r	r	0.10	0.05	r	r
v	v	0.55	0.51	v	v

p-value = 0.338 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.31	0.32	d	d
r	r	0.04	0.02	r	r
x	x	0.65	0.66	x	x

p-value = 0.733 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.87	0.81	d	d
v	v	0.01	0.03	v	v
x	x	0.12	0.15	x	x

p-value = 0.526 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
i	i	0.01	0.00	i	i
j	j	0.03	0.12	j	j
o	o	0.96	0.88	o	o

p-value = 0.056 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
i	i	0.05	0.10	i	i
j	j	0.01	0.07	j	j
p	p	0.94	0.83	p	p

p-value = 0.096 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
i	i	0.04	0.03	i	i
j	j	0.06	0.17	j	j
r	r	0.90	0.80	r	r

p-value = 0.127 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
i	i	0.00	0.02	i	i
j	j	0.04	0.07	j	j
v	v	0.96	0.92	v	v

p-value = 0.495 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
i	i	0.00	0.02	i	i
j	j	0.04	0.05	j	j
x	x	0.96	0.93	x	x

p-value = 0.719 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
i	i	0.91	0.78	i	i
o	o	0.03	0.14	o	o
p	p	0.06	0.08	p	p

p-value = 0.041 (typeface has an effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
i	i	0.22	0.31	i	i
o	o	0.71	0.68	o	o
r	r	0.06	0.02	r	r

p-value = 0.342 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
i	i	0.47	0.37	i	i
o	o	0.43	0.44	o	o
v	v	0.10	0.19	v	v

p-value = 0.284 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
i	i	0.44	0.22	i	i
o	o	0.36	0.46	o	o
x	x	0.19	0.32	x	x

p-value = 0.011 (typeface has an effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
i	i	0.49	0.46	i	i
p	p	0.45	0.47	p	p
r	r	0.05	0.07	r	r

p-value = 0.869 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
i	i	0.32	0.10	i	i
p	p	0.39	0.41	p	p
v	v	0.29	0.49	v	v

p-value = 0.001 (typeface has an effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
i	i	0.25	0.17	i	i
p	p	0.31	0.29	p	p
x	x	0.44	0.54	x	x

p-value = 0.366 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
i	i	0.29	0.41	i	i
r	r	0.12	0.07	r	r
v	v	0.60	0.53	v	v

p-value = 0.245 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
i	i	0.19	0.22	i	i
r	r	0.04	0.02	r	r
x	x	0.77	0.76	x	x

p-value = 0.663 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
i	i	0.88	0.88	i	i
v	v	0.01	0.02	v	v
x	x	0.10	0.10	x	x

p-value = 1.000 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
j	j	0.83	0.78	j	j
o	o	0.12	0.17	o	o
p	p	0.05	0.05	p	p

p-value = 0.630 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
j	j	0.35	0.39	j	j
o	o	0.60	0.54	o	o
r	r	0.05	0.07	r	r

p-value = 0.804 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
j	j	0.55	0.34	j	j
o	o	0.26	0.39	o	o
v	v	0.19	0.27	v	v

p-value = 0.039 (typeface has an effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
j	j	0.42	0.36	j	j
o	o	0.26	0.25	o	o
x	x	0.32	0.39	x	x

p-value = 0.621 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
j	j	0.53	0.58	j	j
p	p	0.32	0.19	p	p
r	r	0.14	0.24	r	r

p-value = 0.074 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
j	j	0.25	0.25	j	j
p	p	0.18	0.12	p	p
v	v	0.57	0.63	v	v

p-value = 0.551 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
j	j	0.19	0.19	j	j
p	p	0.10	0.05	p	p
x	x	0.70	0.76	x	x

p-value = 0.498 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
j	j	0.38	0.39	j	j
r	r	0.10	0.05	r	r
v	v	0.52	0.56	v	v

p-value = 0.564 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
j	j	0.26	0.37	j	j
r	r	0.05	0.03	r	r
x	x	0.69	0.59	x	x

p-value = 0.279 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
j	j	0.87	0.93	j	j
v	v	0.03	0.00	v	v
x	x	0.10	0.07	x	x

p-value = 0.237 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
o	o	0.17	0.22	o	o
p	p	0.05	0.07	p	p
r	r	0.78	0.71	r	r

p-value = 0.706 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
o	o	0.01	0.02	o	o
p	p	0.12	0.08	p	p
v	v	0.87	0.90	v	v

p-value = 0.896 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
o	o	0.05	0.03	o	o
p	p	0.08	0.08	p	p
x	x	0.87	0.88	x	x

p-value = 1.000 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
o	o	0.43	0.54	o	o
r	r	0.30	0.15	r	r
v	v	0.27	0.31	v	v

p-value = 0.100 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
o	o	0.34	0.37	o	o
r	r	0.22	0.10	r	r
x	x	0.44	0.53	x	x

p-value = 0.148 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
o	o	0.84	0.80	o	o
v	v	0.04	0.03	v	v
x	x	0.12	0.17	x	x

p-value = 0.671 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
p	p	0.30	0.34	p	p
r	r	0.06	0.08	r	r
v	v	0.64	0.58	v	v

p-value = 0.718 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
p	p	0.26	0.31	p	p
r	r	0.04	0.03	r	r
x	x	0.70	0.66	x	x

p-value = 0.910 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
p	p	0.79	0.85	p	p
v	v	0.05	0.03	v	v
x	x	0.16	0.12	x	x

p-value = 0.715 (typeface has no effect)

Calibri			Century Schoolbook		
character		freq.	freq.	character	
r	r	0.82	0.80	r	r
v	v	0.04	0.02	v	v
x	x	0.14	0.19	x	x

p-value = 0.486 (typeface has no effect)

Calibri			Georgia		
character		freq.	freq.	character	
d	d	0.96	0.84	d	d
i	i	0.01	0.05	i	i
j	j	0.03	0.11	j	j

p-value = 0.047 (typeface has an effect)

Calibri			Georgia		
character		freq.	freq.	character	
d	d	0.34	0.38	d	d
i	i	0.34	0.28	i	i
v	v	0.32	0.34	v	v

p-value = 0.808 (typeface has no effect)

Calibri			Georgia		
character		freq.	freq.	character	
d	d	0.22	0.19	d	d
j	j	0.30	0.25	j	j
v	v	0.48	0.56	v	v

p-value = 0.512 (typeface has no effect)

Calibri			Georgia		
character		freq.	freq.	character	
i	i	0.00	0.03	i	i
j	j	0.04	0.11	j	j
v	v	0.96	0.86	v	v

p-value = 0.092 (typeface has no effect)

Calibri			Verdana		
character		freq.	freq.	character	
d	d	0.05	0.11	d	d
o	o	0.00	0.03	o	o
x	x	0.95	0.86	x	x

p-value = 0.117 (typeface has no effect)

Cambria			Futura		
character		freq.	freq.	character	
m	m	0.05	0.00	m	m
n	n	0.00	0.00	n	n
q	q	0.95	1.00	q	q

p-value = 0.120 (typeface has no effect)

Cambria			Futura		
character		freq.	freq.	character	
m	m	0.03	0.00	m	m
n	n	0.01	0.00	n	n
z	z	0.96	1.00	z	z

p-value = 0.245 (typeface has no effect)

Cambria			Futura		
character		freq.	freq.	character	
m	m	0.03	0.01	m	m
q	q	0.38	0.42	q	q
z	z	0.59	0.57	z	z

p-value = 0.801 (typeface has no effect)

Cambria			Futura		
character		freq.	freq.	character	
n	n	0.03	0.04	n	n
q	q	0.30	0.36	q	q
z	z	0.68	0.60	z	z

p-value = 0.608 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
b	b	0.43	0.32	b	b
f	f	0.07	0.17	f	f
i	i	0.50	0.51	i	i

p-value = 0.091 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
b	b	0.15	0.08	b	b
f	f	0.46	0.63	f	f
n	n	0.39	0.29	n	n

p-value = 0.095 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
b	b	0.03	0.00	b	b
f	f	0.81	0.91	f	f
p	p	0.16	0.09	p	p

p-value = 0.136 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
b	b	0.01	0.00	b	b
f	f	0.91	0.95	f	f
q	q	0.08	0.05	q	q

p-value = 0.327 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
b	b	0.08	0.09	b	b
f	f	0.11	0.15	f	f
z	z	0.81	0.75	z	z

p-value = 0.768 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
b	b	0.26	0.09	b	b
i	i	0.69	0.82	i	i
n	n	0.05	0.09	n	n

p-value = 0.029 (typeface has an effect)

Cambria			PT Sans		
character		freq.	freq.	character	
b	b	0.05	0.03	b	b
i	i	0.89	0.92	i	i
p	p	0.05	0.05	p	p

p-value = 0.769 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
b	b	0.01	0.03	b	b
i	i	0.96	0.92	i	i
q	q	0.03	0.05	q	q

p-value = 0.578 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
b	b	0.34	0.28	b	b
i	i	0.18	0.28	i	i
z	z	0.49	0.45	z	z

p-value = 0.335 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
b	b	0.03	0.09	b	b
n	n	0.95	0.88	n	n
p	p	0.03	0.03	p	p

p-value = 0.266 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
b	b	0.00	0.03	b	b
n	n	0.88	0.86	n	n
q	q	0.12	0.11	q	q

p-value = 0.649 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
b	b	0.38	0.18	b	b
n	n	0.03	0.08	n	n
z	z	0.59	0.74	z	z

p-value = 0.023 (typeface has an effect)

Cambria			PT Sans		
character		freq.	freq.	character	
b	b	0.73	0.51	b	b
p	p	0.08	0.32	p	p
q	q	0.19	0.17	q	q

p-value = 0.001 (typeface has an effect)

Cambria			PT Sans		
character		freq.	freq.	character	
b	b	0.03	0.03	b	b
p	p	0.03	0.03	p	p
z	z	0.95	0.94	z	z

p-value = 1.000 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
b	b	0.00	0.03	b	b
q	q	0.03	0.02	q	q
z	z	0.97	0.95	z	z

p-value = 0.620 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
f	f	0.32	0.20	f	f
i	i	0.26	0.37	i	i
n	n	0.42	0.43	n	n

p-value = 0.185 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
f	f	0.22	0.15	f	f
i	i	0.39	0.43	i	i
p	p	0.39	0.42	p	p

p-value = 0.587 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
f	f	0.05	0.12	f	f
i	i	0.35	0.40	i	i
q	q	0.59	0.48	q	q

p-value = 0.202 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
f	f	0.18	0.20	f	f
i	i	0.12	0.18	i	i
z	z	0.70	0.62	z	z

p-value = 0.427 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
f	f	0.58	0.77	f	f
n	n	0.30	0.12	n	n
p	p	0.12	0.11	p	p

p-value = 0.024 (typeface has an effect)

Cambria			PT Sans		
character		freq.	freq.	character	
f	f	0.53	0.63	f	f
n	n	0.23	0.14	n	n
q	q	0.24	0.23	q	q

p-value = 0.275 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
f	f	0.53	0.29	f	f
n	n	0.04	0.09	n	n
z	z	0.43	0.62	z	z

p-value = 0.013 (typeface has an effect)

Cambria			PT Sans		
character		freq.	freq.	character	
f	f	0.97	0.91	f	f
p	p	0.00	0.03	p	p
q	q	0.03	0.06	q	q

p-value = 0.193 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
f	f	0.23	0.15	f	f
p	p	0.05	0.22	p	p
z	z	0.72	0.63	z	z

p-value = 0.010 (typeface has an effect)

Cambria			PT Sans		
character		freq.	freq.	character	
f	f	0.11	0.22	f	f
q	q	0.15	0.18	q	q
z	z	0.74	0.60	z	z

p-value = 0.116 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
i	i	0.66	0.83	i	i
n	n	0.05	0.08	n	n
p	p	0.28	0.09	p	p

p-value = 0.012 (typeface has an effect)

Cambria			PT Sans		
character		freq.	freq.	character	
i	i	0.55	0.78	i	i
n	n	0.03	0.05	n	n
q	q	0.42	0.17	q	q

p-value = 0.003 (typeface has an effect)

Cambria			PT Sans		
character		freq.	freq.	character	
i	i	0.35	0.46	i	i
n	n	0.07	0.17	n	n
z	z	0.58	0.37	z	z

p-value = 0.016 (typeface has an effect)

Cambria			PT Sans		
character		freq.	freq.	character	
i	i	0.97	0.92	i	i
p	p	0.00	0.05	p	p
q	q	0.03	0.03	q	q

p-value = 0.210 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
i	i	0.27	0.29	i	i
p	p	0.26	0.26	p	p
z	z	0.47	0.45	z	z

p-value = 0.977 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
i	i	0.19	0.32	i	i
q	q	0.34	0.35	q	q
z	z	0.47	0.32	z	z

p-value = 0.106 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
n	n	0.92	0.89	n	n
p	p	0.00	0.00	p	p
q	q	0.08	0.11	q	q

p-value = 0.780 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
n	n	0.01	0.02	n	n
p	p	0.30	0.18	p	p
z	z	0.69	0.80	z	z

p-value = 0.258 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
n	n	0.03	0.02	n	n
q	q	0.30	0.31	q	q
z	z	0.68	0.68	z	z

p-value = 1.000 (typeface has no effect)

Cambria			PT Sans		
character		freq.	freq.	character	
p	p	0.01	0.05	p	p
q	q	0.03	0.02	q	q
z	z	0.96	0.94	z	z

p-value = 0.713 (typeface has no effect)

Cambria			PT Serif		
character		freq.	freq.	character	
m	m	0.05	0.02	m	m
n	n	0.00	0.00	n	n
q	q	0.95	0.98	q	q

p-value = 0.366 (typeface has no effect)

Cambria			PT Serif		
character		freq.	freq.	character	
m	m	0.03	0.03	m	m
n	n	0.01	0.00	n	n
z	z	0.96	0.97	z	z

p-value = 1.000 (typeface has no effect)

Cambria			PT Serif		
character		freq.	freq.	character	
m	m	0.03	0.03	m	m
q	q	0.38	0.33	q	q
z	z	0.59	0.63	z	z

p-value = 0.897 (typeface has no effect)

Cambria			PT Serif		
character		freq.	freq.	character	
n	n	0.03	0.03	n	n
q	q	0.30	0.33	q	q
z	z	0.68	0.63	z	z

p-value = 0.894 (typeface has no effect)

Cambria			Times New Roman		
character		freq.	freq.	character	
f	f	1.00	1.00	f	f
m	m	0.00	0.00	m	m
n	n	0.00	0.00	n	n

p-value = 1.000 (typeface has no effect)

Candara			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.97	0.90	d	d
i	i	0.02	0.00	i	i
j	j	0.02	0.10	j	j

p-value = 0.114 (typeface has no effect)

Candara			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.34	0.34	d	d
i	i	0.36	0.24	i	i
v	v	0.30	0.42	v	v

p-value = 0.213 (typeface has no effect)

Candara			Century Schoolbook		
character		freq.	freq.	character	
d	d	0.26	0.12	d	d
j	j	0.36	0.19	j	j
v	v	0.38	0.69	v	v

p-value = 0.001 (typeface has an effect)

Candara			Century Schoolbook		
character		freq.	freq.	character	
i	i	0.00	0.02	i	i
j	j	0.05	0.07	j	j
v	v	0.95	0.92	v	v

p-value = 0.717 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.00	0.00	b	b
d	d	0.02	0.00	d	d
i	i	0.98	1.00	i	i

p-value = 1.000 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.02	0.00	b	b
d	d	0.00	0.00	d	d
j	j	0.98	1.00	j	j

p-value = 1.000 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.02	0.03	b	b
d	d	0.05	0.03	d	d
k	k	0.93	0.94	k	k

p-value = 1.000 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.00	0.05	b	b
d	d	0.05	0.00	d	d
l	l	0.95	0.95	l	l

p-value = 0.060 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.02	0.00	b	b
d	d	0.00	0.02	d	d
v	v	0.98	0.98	v	v

p-value = 1.000 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.02	0.00	b	b
d	d	0.02	0.00	d	d
y	y	0.97	1.00	y	y

p-value = 0.496 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.98	0.92	b	b
i	i	0.00	0.03	i	i
j	j	0.02	0.05	j	j

p-value = 0.239 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.07	0.11	b	b
i	i	0.80	0.80	i	i
k	k	0.13	0.09	k	k

p-value = 0.521 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.80	0.48	b	b
i	i	0.20	0.52	i	i
l	l	0.00	0.00	l	l

p-value < 0.001 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.30	0.42	b	b
i	i	0.39	0.31	i	i
v	v	0.31	0.27	v	v

p-value = 0.340 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.26	0.28	b	b
i	i	0.44	0.34	i	i
y	y	0.30	0.38	y	y

p-value = 0.510 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.15	0.06	b	b
j	j	0.75	0.81	j	j
k	k	0.10	0.12	k	k

p-value = 0.369 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.75	0.34	b	b
j	j	0.23	0.56	j	j
l	l	0.02	0.09	l	l

p-value < 0.001 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.36	0.17	b	b
j	j	0.30	0.28	j	j
v	v	0.34	0.55	v	v

p-value = 0.027 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.64	0.80	b	b
j	j	0.23	0.08	j	j
y	y	0.13	0.12	y	y

p-value = 0.046 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.13	0.27	b	b
k	k	0.16	0.14	k	k
l	l	0.70	0.59	l	l

p-value = 0.144 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.34	0.34	b	b
k	k	0.05	0.00	k	k
v	v	0.61	0.66	v	v

p-value = 0.312 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.36	0.23	b	b
k	k	0.08	0.05	k	k
y	y	0.56	0.72	y	y

p-value = 0.186 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.18	0.11	b	b
l	l	0.31	0.09	l	l
v	v	0.51	0.80	v	v

p-value = 0.001 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.10	0.11	b	b
l	l	0.36	0.06	l	l
y	y	0.54	0.83	y	y

p-value < 0.001 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
b	b	0.89	0.84	b	b
v	v	0.02	0.06	v	v
y	y	0.10	0.09	y	y

p-value = 0.518 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
d	d	0.97	0.84	d	d
i	i	0.02	0.05	i	i
j	j	0.02	0.11	j	j

p-value = 0.036 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
d	d	0.18	0.12	d	d
i	i	0.72	0.75	i	i
k	k	0.10	0.12	k	k

p-value = 0.621 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
d	d	0.79	0.55	d	d
i	i	0.20	0.42	i	i
l	l	0.02	0.03	l	l

p-value = 0.011 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
d	d	0.34	0.38	d	d
i	i	0.36	0.28	i	i
v	v	0.30	0.34	v	v

p-value = 0.698 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
d	d	0.16	0.12	d	d
i	i	0.48	0.28	i	i
y	y	0.36	0.59	y	y

p-value = 0.049 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
d	d	0.18	0.09	d	d
j	j	0.74	0.83	j	j
k	k	0.08	0.08	k	k

p-value = 0.334 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
d	d	0.82	0.31	d	d
j	j	0.11	0.66	j	j
l	l	0.07	0.03	l	l

p-value < 0.001 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
d	d	0.26	0.19	d	d
j	j	0.36	0.25	j	j
v	v	0.38	0.56	v	v

p-value = 0.112 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
d	d	0.49	0.81	d	d
j	j	0.31	0.12	j	j
y	y	0.20	0.06	y	y

p-value = 0.001 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
d	d	0.18	0.25	d	d
k	k	0.20	0.16	k	k
l	l	0.62	0.59	l	l

p-value = 0.602 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
d	d	0.44	0.28	d	d
k	k	0.02	0.05	k	k
v	v	0.54	0.67	v	v

p-value = 0.165 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
d	d	0.38	0.34	d	d
k	k	0.07	0.03	k	k
y	y	0.56	0.62	y	y

p-value = 0.655 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
d	d	0.28	0.11	d	d
l	l	0.25	0.09	l	l
v	v	0.48	0.80	v	v

p-value = 0.001 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
d	d	0.16	0.09	d	d
l	l	0.31	0.12	l	l
y	y	0.52	0.78	y	y

p-value = 0.011 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
d	d	0.90	0.86	d	d
v	v	0.02	0.03	v	v
y	y	0.08	0.11	y	y

p-value = 0.667 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
i	i	0.02	0.03	i	i
j	j	0.07	0.09	j	j
k	k	0.92	0.88	k	k

p-value = 0.645 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
i	i	0.00	0.08	i	i
j	j	0.23	0.30	j	j
l	l	0.77	0.62	l	l

p-value = 0.042 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
i	i	0.00	0.03	i	i
j	j	0.05	0.11	j	j
v	v	0.95	0.86	v	v

p-value = 0.148 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
i	i	0.10	0.16	i	i
j	j	0.00	0.02	j	j
y	y	0.90	0.83	y	y

p-value = 0.298 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
i	i	0.20	0.38	i	i
k	k	0.79	0.61	k	k
l	l	0.02	0.02	l	l

p-value = 0.064 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
i	i	0.72	0.66	i	i
k	k	0.13	0.14	k	k
v	v	0.15	0.20	v	v

p-value = 0.742 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
i	i	0.64	0.70	i	i
k	k	0.13	0.08	k	k
y	y	0.23	0.22	y	y

p-value = 0.678 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
i	i	0.07	0.16	i	i
l	l	0.00	0.06	l	l
v	v	0.93	0.78	v	v

p-value = 0.018 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
i	i	0.05	0.06	i	i
l	l	0.02	0.02	l	l
y	y	0.93	0.92	y	y

p-value = 1.000 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
i	i	0.87	0.89	i	i
v	v	0.02	0.00	v	v
y	y	0.11	0.11	y	y

p-value = 1.000 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
j	j	0.25	0.62	j	j
k	k	0.69	0.36	k	k
l	l	0.07	0.02	l	l

p-value < 0.001 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
j	j	0.72	0.69	j	j
k	k	0.16	0.12	k	k
v	v	0.11	0.19	v	v

p-value = 0.653 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
j	j	0.51	0.31	j	j
k	k	0.44	0.62	k	k
y	y	0.05	0.06	y	y

p-value = 0.057 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
j	j	0.18	0.22	j	j
l	l	0.00	0.03	l	l
v	v	0.82	0.75	v	v

p-value = 0.465 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
j	j	0.05	0.11	j	j
l	l	0.23	0.59	l	l
y	y	0.72	0.30	y	y

p-value < 0.001 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
j	j	0.80	0.77	j	j
v	v	0.15	0.22	v	v
y	y	0.05	0.02	y	y

p-value = 0.238 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
k	k	0.16	0.06	k	k
l	l	0.52	0.34	l	l
v	v	0.31	0.59	v	v

p-value = 0.004 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
k	k	0.15	0.08	k	k
l	l	0.54	0.25	l	l
y	y	0.31	0.67	y	y

p-value < 0.001 (typeface has an effect)

Candara			Georgia		
character		freq.	freq.	character	
k	k	0.87	0.83	k	k
v	v	0.00	0.08	v	v
y	y	0.13	0.09	y	y

p-value = 0.075 (typeface has no effect)

Candara			Georgia		
character		freq.	freq.	character	
l	l	0.90	0.92	l	l
v	v	0.00	0.02	v	v
y	y	0.10	0.06	y	y

p-value = 0.744 (typeface has no effect)

Century Schoolbook			Georgia		
character		freq.	freq.	character	
d	d	0.90	0.84	d	d
i	i	0.00	0.05	i	i
j	j	0.10	0.11	j	j

p-value = 0.361 (typeface has no effect)

Century Schoolbook			Georgia		
character		freq.	freq.	character	
d	d	0.34	0.38	d	d
i	i	0.24	0.28	i	i
v	v	0.42	0.34	v	v

p-value = 0.694 (typeface has no effect)

Century Schoolbook			Georgia		
character		freq.	freq.	character	
d	d	0.12	0.19	d	d
j	j	0.19	0.25	j	j
v	v	0.69	0.56	v	v

p-value = 0.356 (typeface has no effect)

Century Schoolbook			Georgia		
character		freq.	freq.	character	
i	i	0.02	0.03	i	i
j	j	0.07	0.11	j	j
v	v	0.92	0.86	v	v

p-value = 0.519 (typeface has no effect)

Century Schoolbook			Verdana		
character		freq.	freq.	character	
d	d	0.05	0.11	d	d
o	o	0.03	0.03	o	o
x	x	0.92	0.86	x	x

p-value = 0.404 (typeface has no effect)

Courier New			Verdana		
character		freq.	freq.	character	
h	h	0.22	0.21	h	h
o	o	0.46	0.13	o	o
w	w	0.32	0.67	w	w

p-value < 0.001 (typeface has an effect)

Futura			PT Sans		
character		freq.	freq.	character	
n	n	0.04	0.02	n	n
q	q	0.36	0.31	q	q
z	z	0.60	0.68	z	z

p-value = 0.512 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.79	0.89	g	g
h	h	0.04	0.05	h	h
m	m	0.16	0.06	m	m

p-value = 0.154 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.87	0.89	g	g
h	h	0.03	0.02	h	h
n	n	0.10	0.10	n	n

p-value = 0.833 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.06	0.59	g	g
h	h	0.91	0.35	h	h
q	q	0.03	0.06	q	q

p-value < 0.001 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.66	0.78	g	g
h	h	0.07	0.13	h	h
r	r	0.27	0.10	r	r

p-value = 0.021 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.13	0.46	g	g
h	h	0.45	0.32	h	h
y	y	0.42	0.22	y	y

p-value < 0.001 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.13	0.49	g	g
h	h	0.06	0.06	h	h
z	z	0.81	0.44	z	z

p-value < 0.001 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.97	0.98	g	g
m	m	0.03	0.02	m	m
n	n	0.00	0.00	n	n

p-value = 1.000 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.01	0.35	g	g
m	m	0.96	0.57	m	m
q	q	0.03	0.08	q	q

p-value < 0.001 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.91	0.79	g	g
m	m	0.03	0.06	m	m
r	r	0.06	0.14	r	r

p-value = 0.134 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.03	0.44	g	g
m	m	0.57	0.37	m	m
y	y	0.40	0.19	y	y

p-value < 0.001 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.36	0.51	g	g
m	m	0.00	0.13	m	m
z	z	0.64	0.37	z	z

p-value < 0.001 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.00	0.40	g	g
n	n	1.00	0.52	n	n
q	q	0.00	0.08	q	q

p-value < 0.001 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.94	0.87	g	g
n	n	0.00	0.03	n	n
r	r	0.06	0.10	r	r

p-value = 0.283 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.03	0.48	g	g
n	n	0.57	0.40	n	n
y	y	0.40	0.13	y	y

p-value < 0.001 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.45	0.62	g	g
n	n	0.03	0.03	n	n
z	z	0.52	0.35	z	z

p-value = 0.205 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.04	0.32	g	g
q	q	0.01	0.14	q	q
r	r	0.94	0.54	r	r

p-value < 0.001 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.09	0.37	g	g
q	q	0.01	0.13	q	q
y	y	0.90	0.51	y	y

p-value < 0.001 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.01	0.21	g	g
q	q	0.01	0.05	q	q
z	z	0.97	0.75	z	z

p-value < 0.001 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.04	0.44	g	g
r	r	0.51	0.43	r	r
y	y	0.45	0.13	y	y

p-value < 0.001 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.40	0.70	g	g
r	r	0.04	0.05	r	r
z	z	0.55	0.25	z	z

p-value = 0.001 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
g	g	0.46	0.59	g	g
y	y	0.03	0.03	y	y
z	z	0.51	0.38	z	z

p-value = 0.348 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
h	h	0.69	0.73	h	h
m	m	0.30	0.27	m	m
n	n	0.01	0.00	n	n

p-value = 0.704 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
h	h	0.03	0.00	h	h
m	m	0.37	0.14	m	m
q	q	0.60	0.86	q	q

p-value = 0.002 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
h	h	0.40	0.22	h	h
m	m	0.16	0.08	m	m
r	r	0.43	0.70	r	r

p-value = 0.007 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
h	h	0.03	0.02	h	h
m	m	0.10	0.10	m	m
y	y	0.87	0.89	y	y

p-value = 0.833 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
h	h	0.21	0.13	h	h
m	m	0.01	0.03	m	m
z	z	0.78	0.84	z	z

p-value = 0.516 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
h	h	0.01	0.02	h	h
n	n	0.21	0.13	n	n
q	q	0.78	0.86	q	q

p-value = 0.677 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
h	h	0.45	0.27	h	h
n	n	0.00	0.00	n	n
r	r	0.55	0.73	r	r

p-value = 0.047 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
h	h	0.00	0.02	h	h
n	n	0.16	0.08	n	n
y	y	0.84	0.90	y	y

p-value = 0.182 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
h	h	0.16	0.10	h	h
n	n	0.00	0.02	n	n
z	z	0.84	0.89	z	z

p-value = 0.299 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
h	h	0.04	0.05	h	h
q	q	0.58	0.65	q	q
r	r	0.37	0.30	r	r

p-value = 0.629 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
h	h	0.40	0.35	h	h
q	q	0.10	0.19	q	q
y	y	0.49	0.46	y	y

p-value = 0.348 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
h	h	0.03	0.00	h	h
q	q	0.13	0.21	q	q
z	z	0.84	0.79	z	z

p-value = 0.212 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
h	h	0.06	0.22	h	h
r	r	0.16	0.13	r	r
y	y	0.78	0.65	y	y

p-value = 0.025 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
h	h	0.16	0.37	h	h
r	r	0.01	0.03	r	r
z	z	0.82	0.60	z	z

p-value = 0.015 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
h	h	0.51	0.43	h	h
y	y	0.18	0.14	y	y
z	z	0.31	0.43	z	z

p-value = 0.484 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
m	m	0.00	0.02	m	m
n	n	0.00	0.00	n	n
q	q	1.00	0.98	q	q

p-value = 1.000 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
m	m	0.21	0.08	m	m
n	n	0.00	0.00	n	n
r	r	0.79	0.92	r	r

p-value = 0.045 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
m	m	0.00	0.02	m	m
n	n	0.01	0.02	n	n
y	y	0.99	0.97	y	y

p-value = 1.000 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
m	m	0.00	0.03	m	m
n	n	0.00	0.00	n	n
z	z	1.00	0.97	z	z

p-value = 0.496 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
m	m	0.06	0.14	m	m
q	q	0.90	0.76	q	q
r	r	0.04	0.10	r	r

p-value = 0.119 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
m	m	0.55	0.38	m	m
q	q	0.03	0.08	q	q
y	y	0.42	0.54	y	y

p-value = 0.127 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
m	m	0.01	0.03	m	m
q	q	0.42	0.33	q	q
z	z	0.57	0.63	z	z

p-value = 0.524 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
m	m	0.04	0.08	m	m
r	r	0.03	0.03	r	r
y	y	0.93	0.89	y	y

p-value = 0.893 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
m	m	0.04	0.19	m	m
r	r	0.01	0.02	r	r
z	z	0.94	0.79	z	z

p-value = 0.014 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
m	m	0.52	0.54	m	m
y	y	0.39	0.27	y	y
z	z	0.09	0.19	z	z

p-value = 0.129 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
n	n	0.01	0.06	n	n
q	q	0.94	0.75	q	q
r	r	0.04	0.19	r	r

p-value = 0.005 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
n	n	0.57	0.38	n	n
q	q	0.01	0.14	q	q
y	y	0.42	0.48	y	y

p-value = 0.006 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
n	n	0.04	0.03	n	n
q	q	0.36	0.33	q	q
z	z	0.60	0.63	z	z

p-value = 0.819 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
n	n	0.01	0.10	n	n
r	r	0.00	0.13	r	r
y	y	0.99	0.78	y	y

p-value < 0.001 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
n	n	0.00	0.03	n	n
r	r	0.07	0.11	r	r
z	z	0.93	0.86	z	z

p-value = 0.341 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
n	n	0.46	0.41	n	n
y	y	0.43	0.37	y	y
z	z	0.10	0.22	z	z

p-value = 0.188 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
q	q	0.09	0.30	q	q
r	r	0.55	0.43	r	r
y	y	0.36	0.27	y	y

p-value = 0.010 (typeface has an effect)

Futura			PT Serif		
character		freq.	freq.	character	
q	q	0.45	0.40	q	q
r	r	0.01	0.05	r	r
z	z	0.54	0.56	z	z

p-value = 0.593 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
q	q	0.37	0.51	q	q
y	y	0.04	0.05	y	y
z	z	0.58	0.44	z	z

p-value = 0.359 (typeface has no effect)

Futura			PT Serif		
character		freq.	freq.	character	
r	r	0.51	0.38	r	r
y	y	0.36	0.32	y	y
z	z	0.13	0.30	z	z

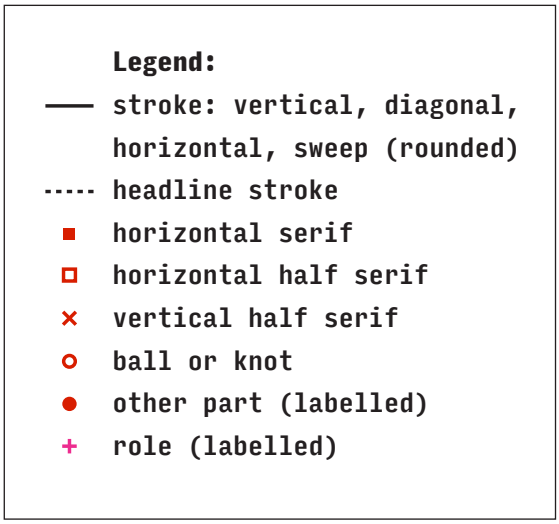
p-value = 0.073 (typeface has no effect)

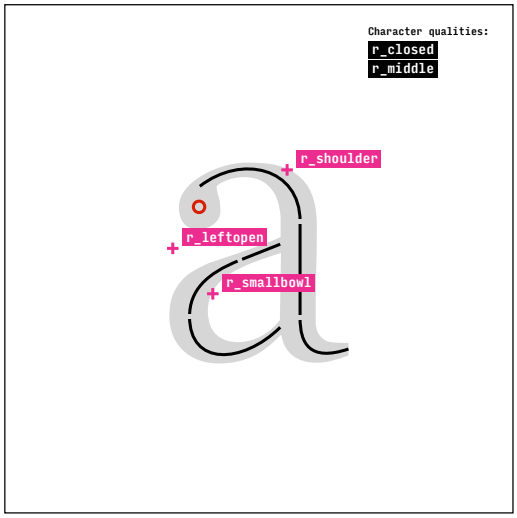
PT Sans			PT Serif		
character		freq.	freq.	character	
n	n	0.02	0.03	n	n
q	q	0.31	0.33	q	q
z	z	0.68	0.63	z	z

p-value = 0.780 (typeface has no effect)

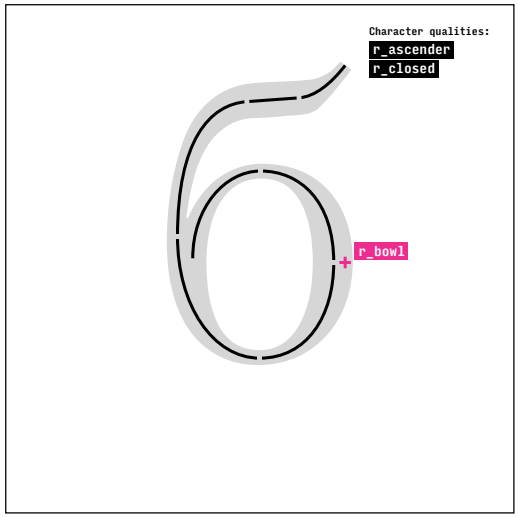
Appendix 7: concept diagrams

Diagrams for all character concepts used by the model described in chapter 6. The stroke parts are marked with their approximate midlines. The non-stroke parts are marked with red symbols, each symbol representing a different part. Less frequent parts are marked with a circle and labelled. The roles are marked with a cross in magenta and labelled.



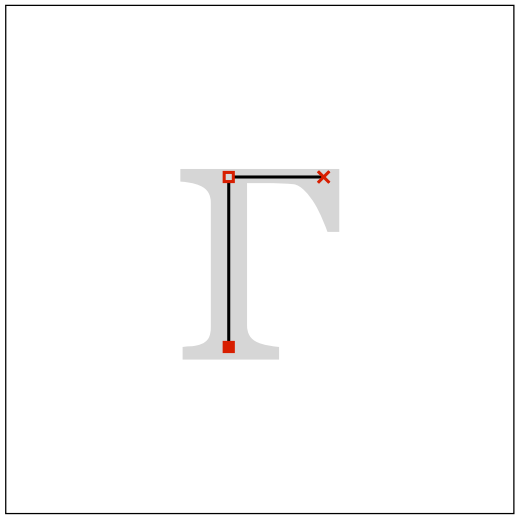


cyrl.serif.a

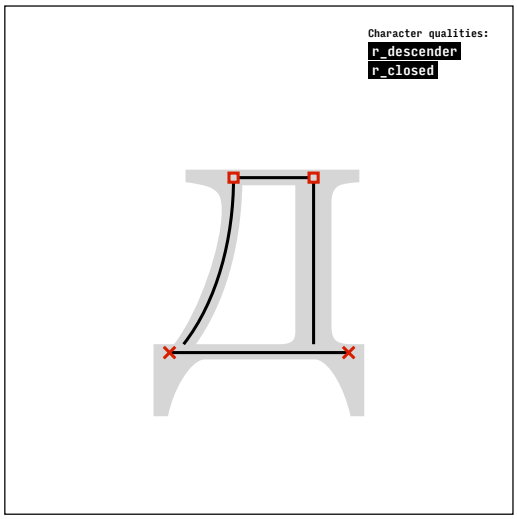


cyrl.serif.be

cyrl.serif.ve

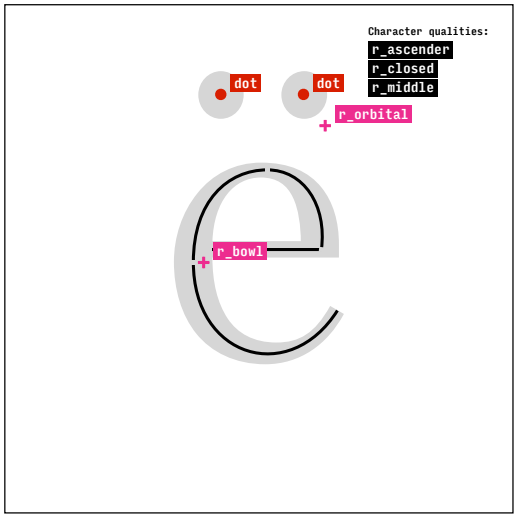


cyrl.serif.ghe

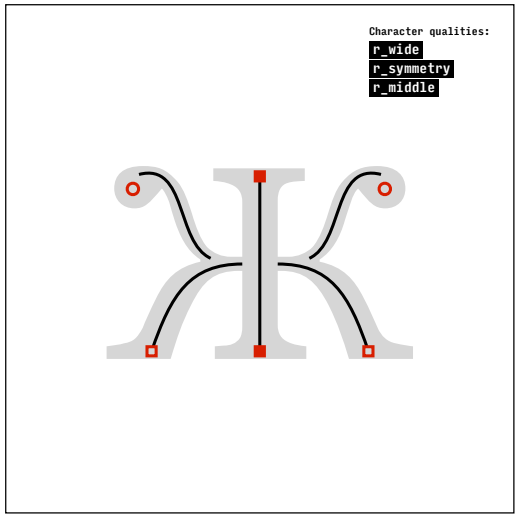


cyrl.serif.de

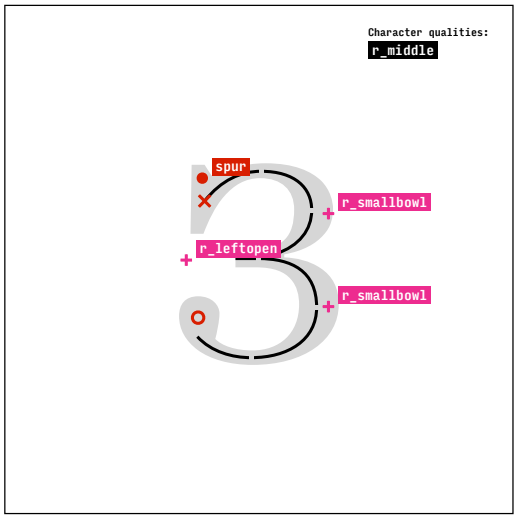
cyrl.serif.ie



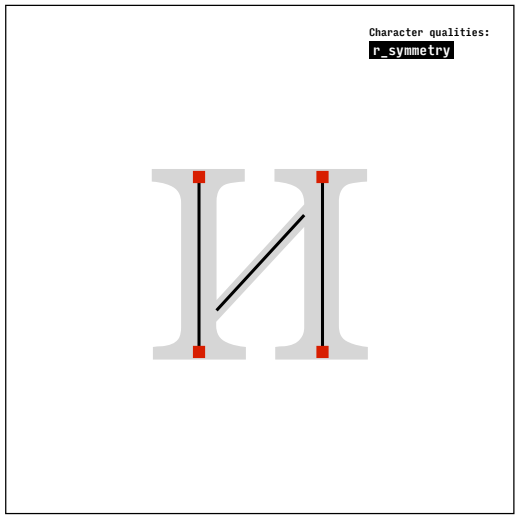
cyrl.serif.io



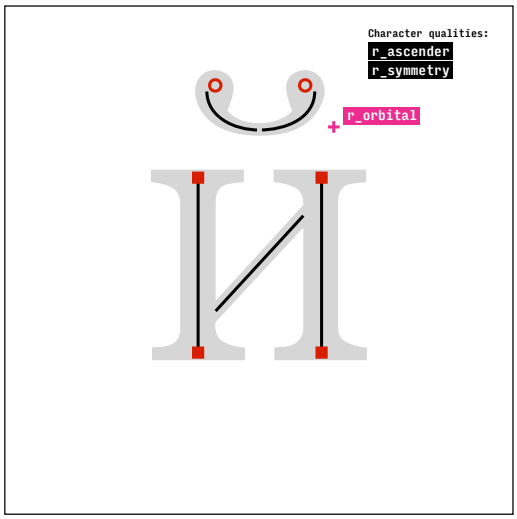
cyrl.serif.zhe



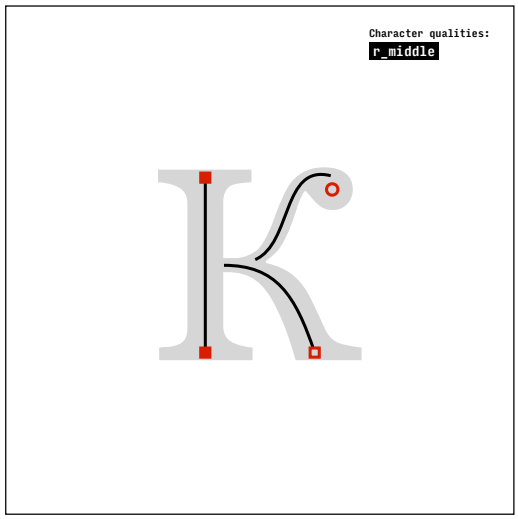
cyrl.serif.ze



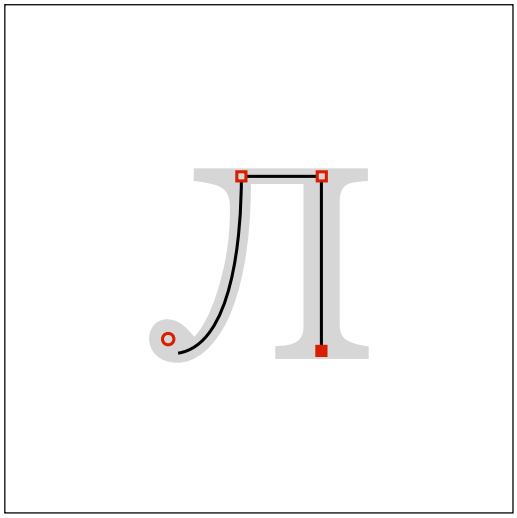
cyrl.serif.i



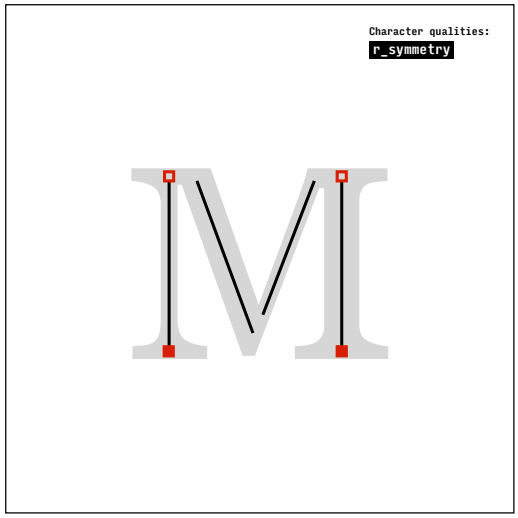
cyrl.serif.short-i



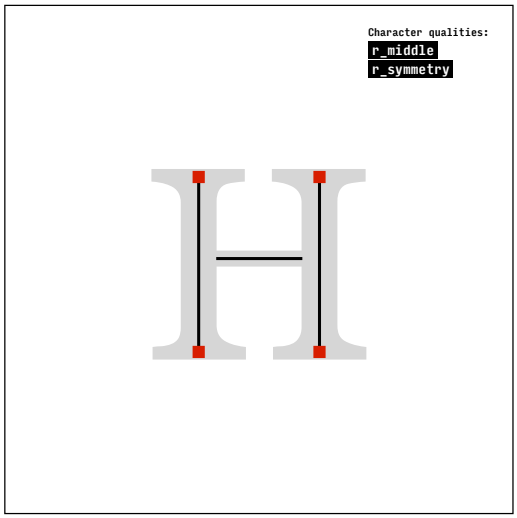
cyrl.serif.ka



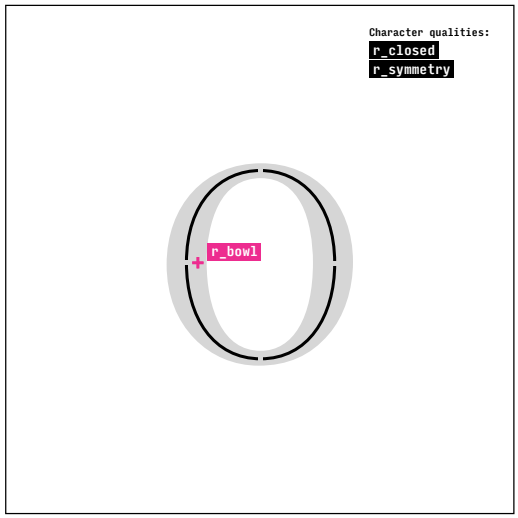
cyrl.serif.el



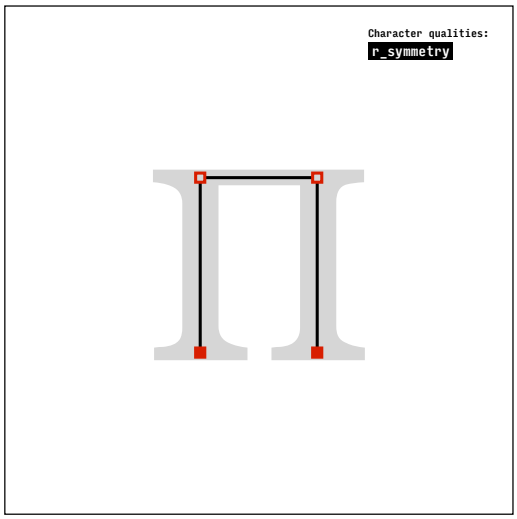
cyrl.serif.em



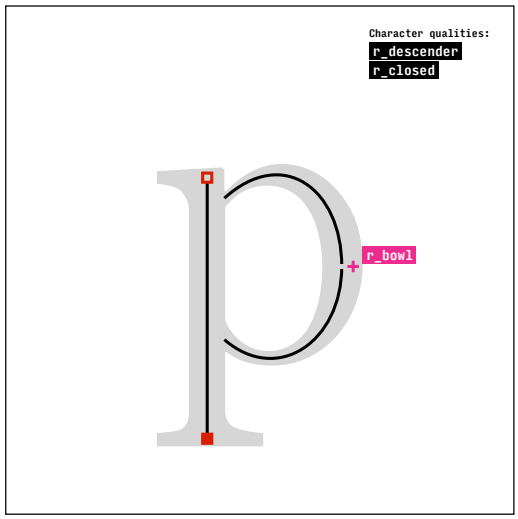
cyrl.serif.en



cyrl.serif.o

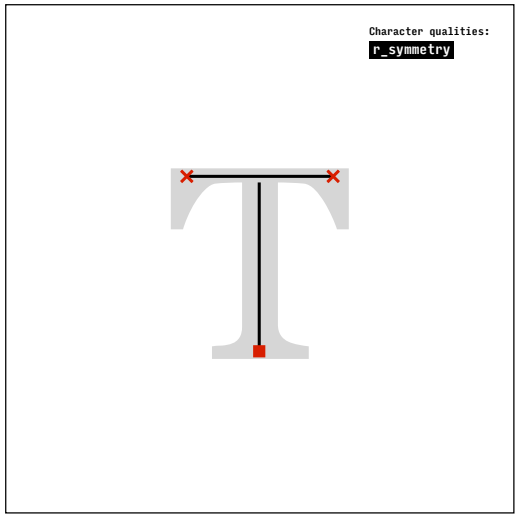


cyrl.serif.pe

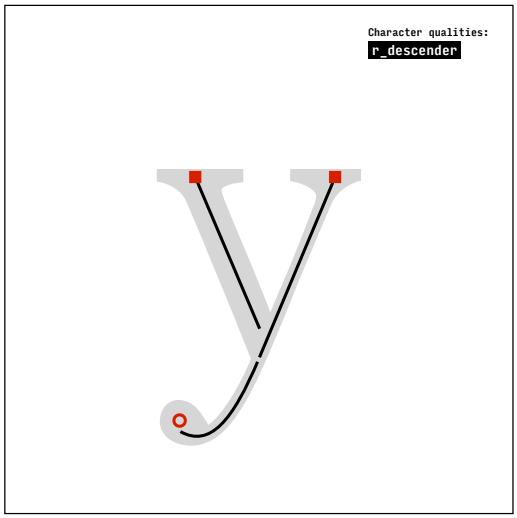


cyrl.serif.er

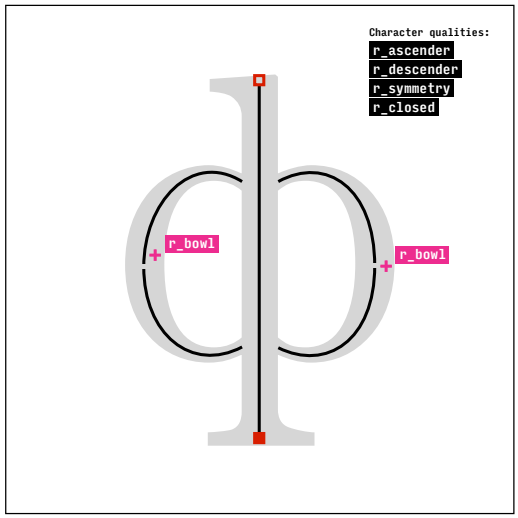
cyrl.serif.es



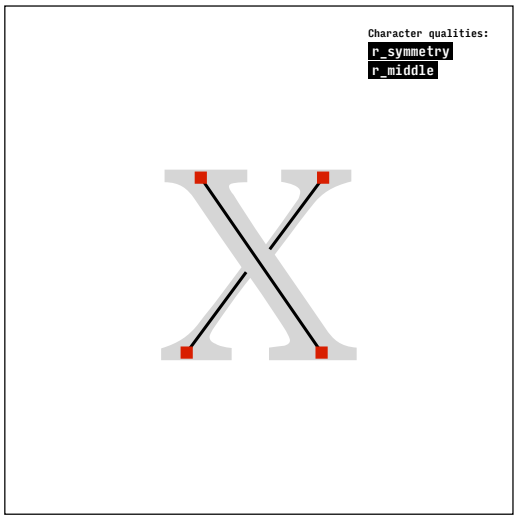
cyrl.serif.te



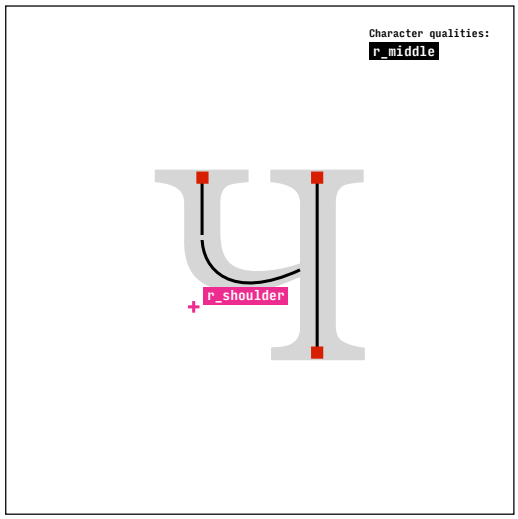
cyrl.serif.u



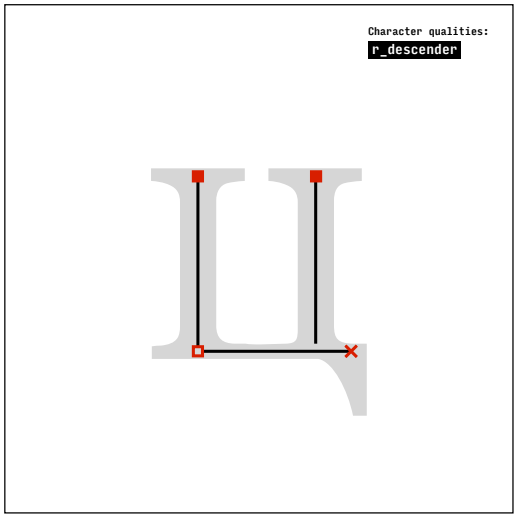
cyrl.serif.ef



cyrl.serif.ha



cyrl.serif.che

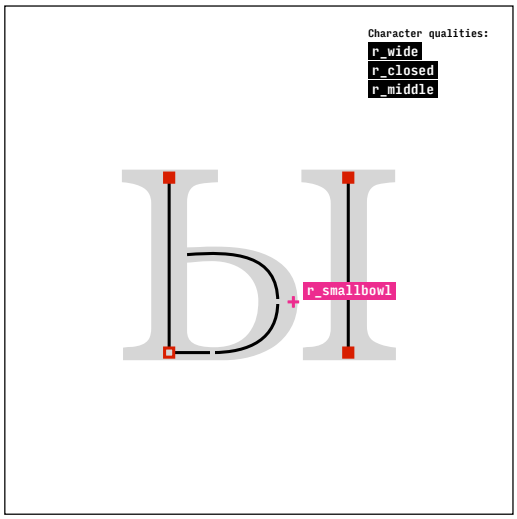


cyrl.serif.tse

cyrl.serif.sha

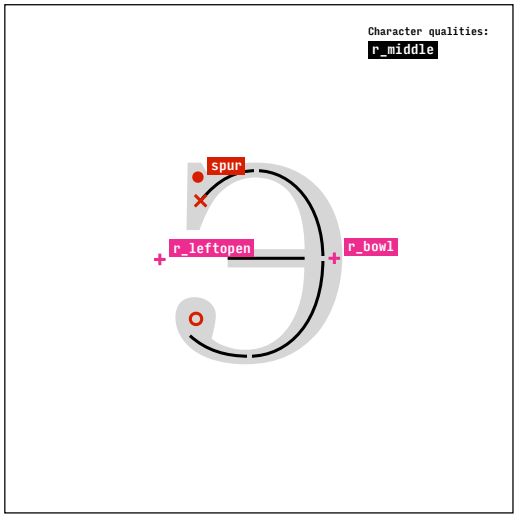
cyrl.serif.shcha

cyrl.serif.hard-sign



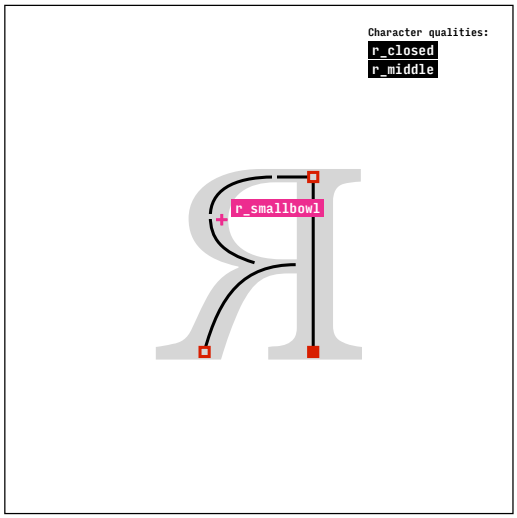
cyrl.serif.yeru

cyrl.serif.soft-sign

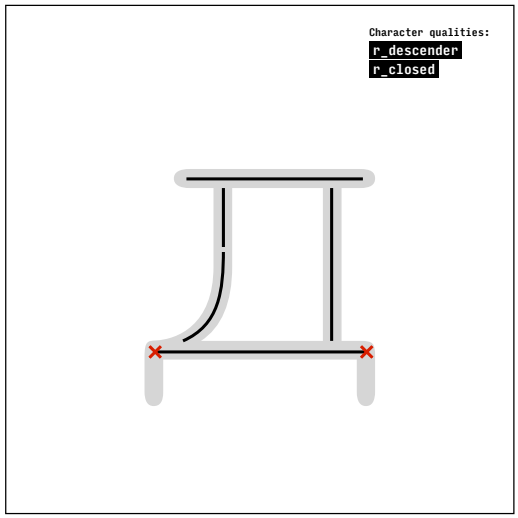


cyrl.serif.e

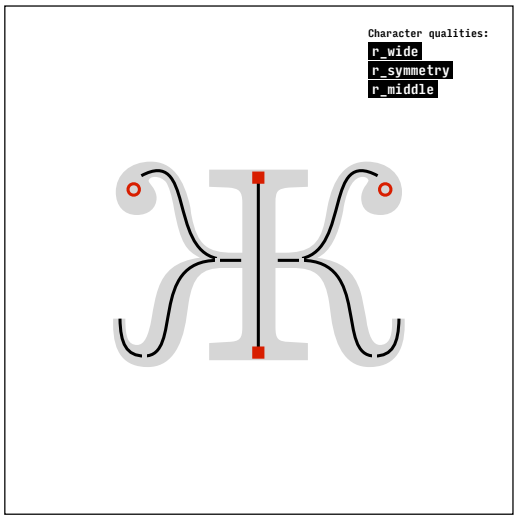
cyrl.serif.yu



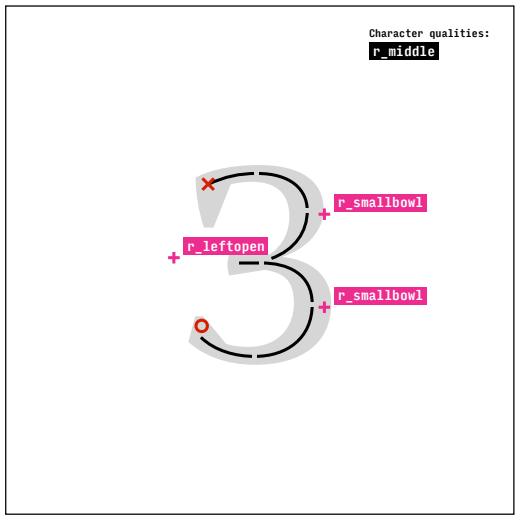
cyrl.serif.ya



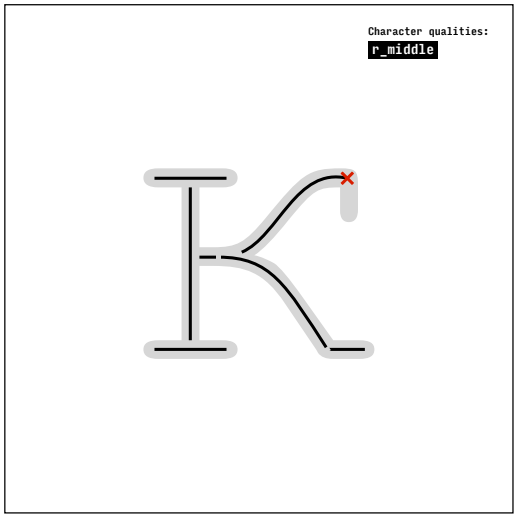
cyrl.serif.de.courier



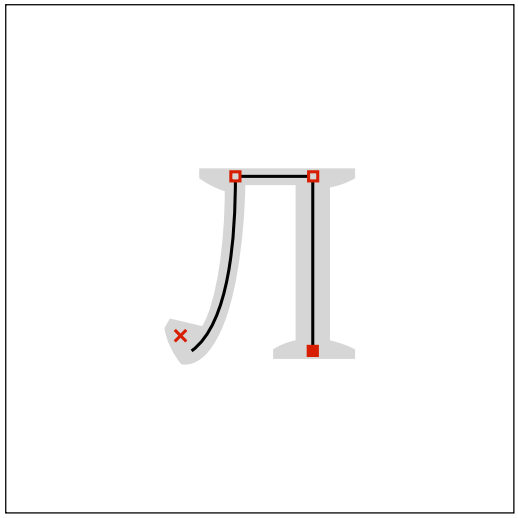
cyrl.serif.zhe.withflick



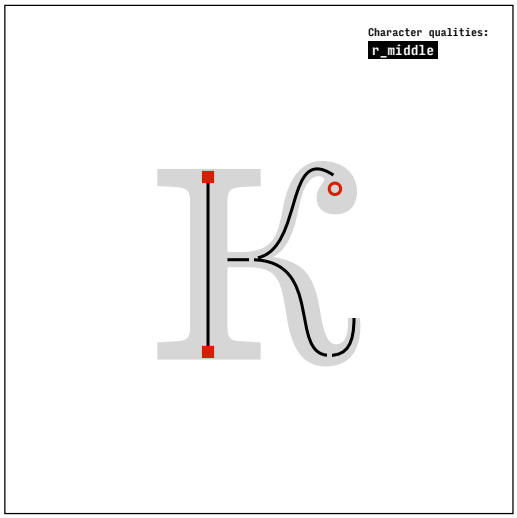
cyrl.serif.ze.halfseriffed



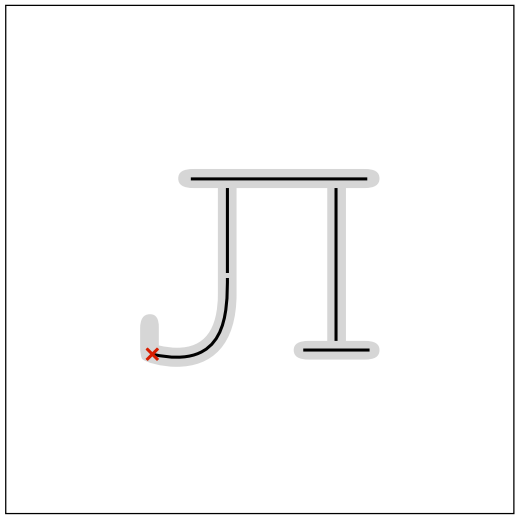
cyrl.serif.ka.courier



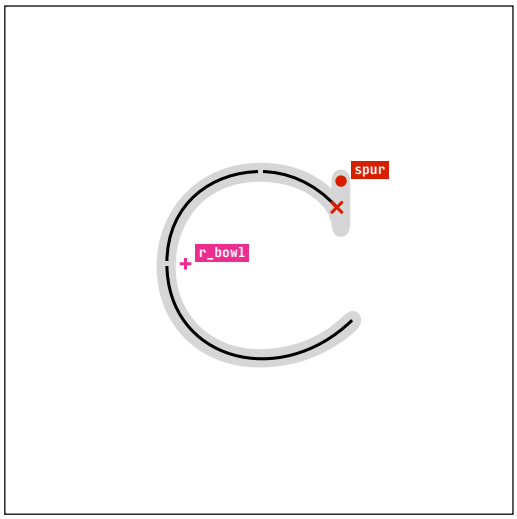
cyrl.serif.el.halfseriffed



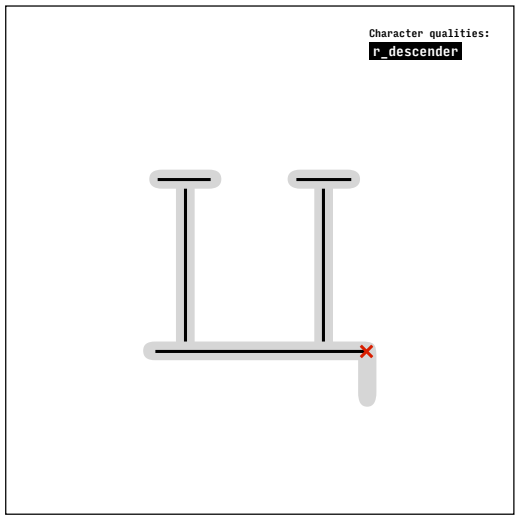
cyrl.serif.ka.withflick



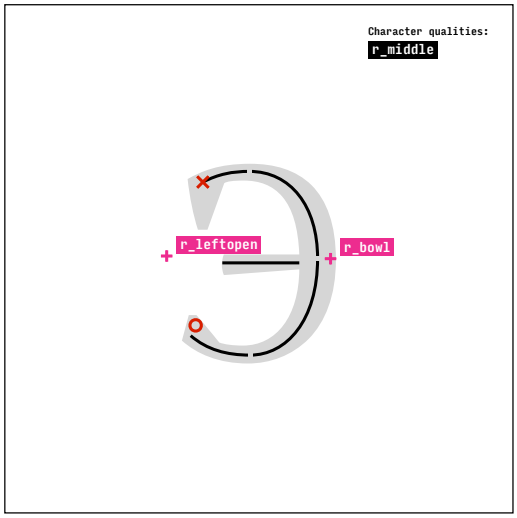
cyrl.serif.el.courier



cyrl.serif.es.courier



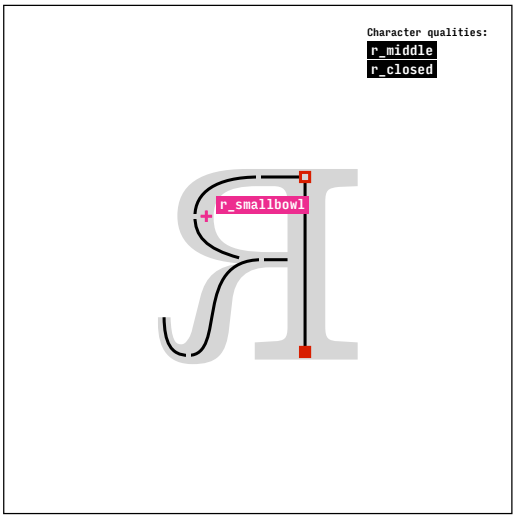
cyrl.serif.tse.courier



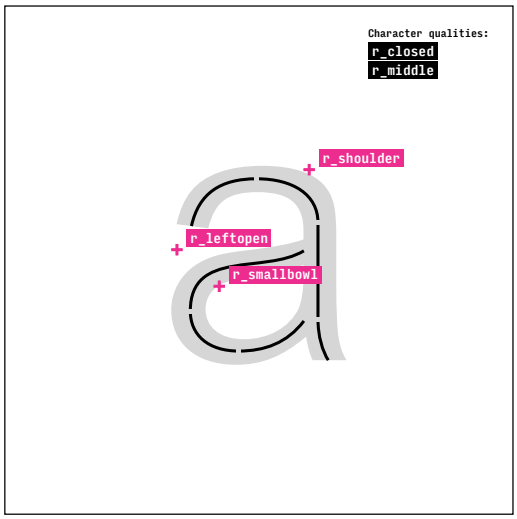
cyrl.serif.e.halfseriffed

cyrl.serif.yu.courier

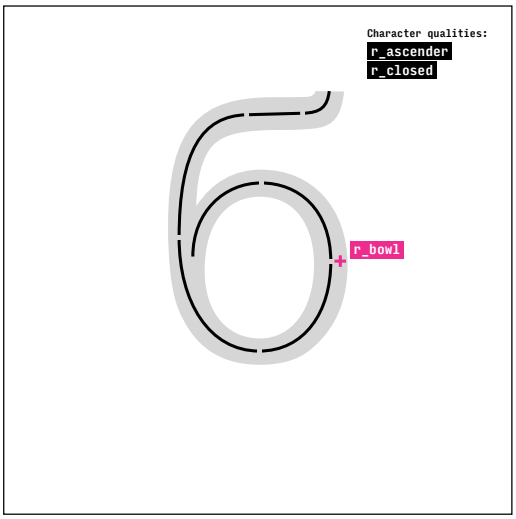
cyrl.serif.ya.courier



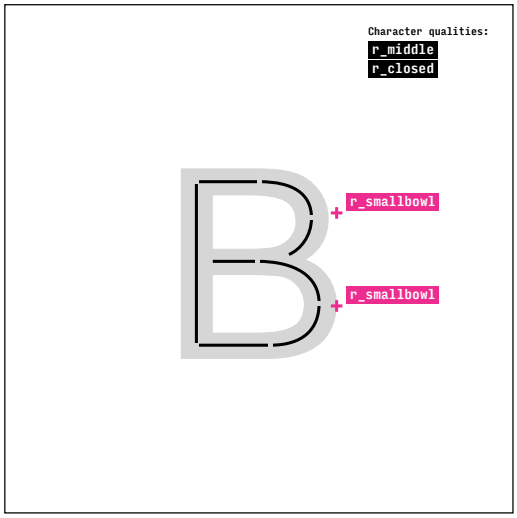
cyrl.serif.ya.withflick



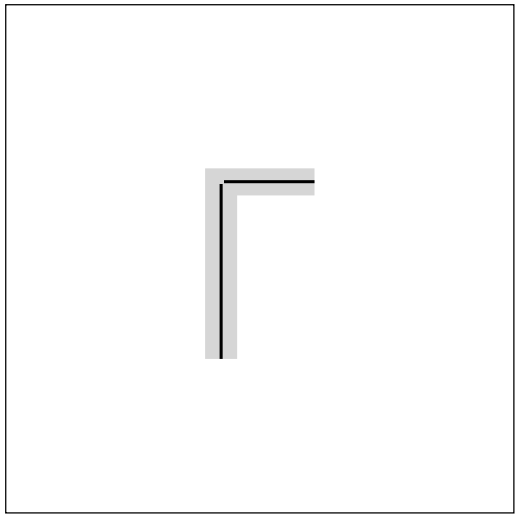
cyrl.sans.a



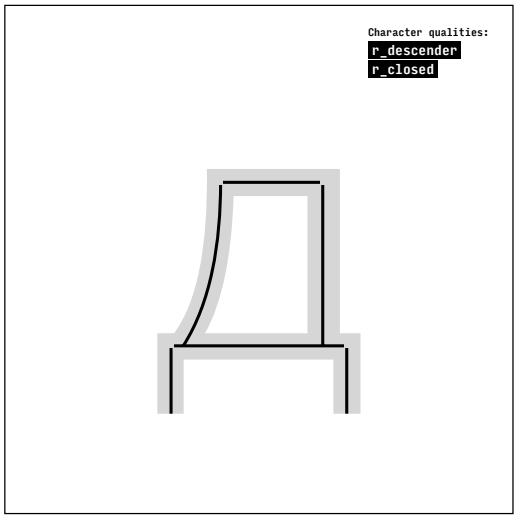
cyrl.sans.be



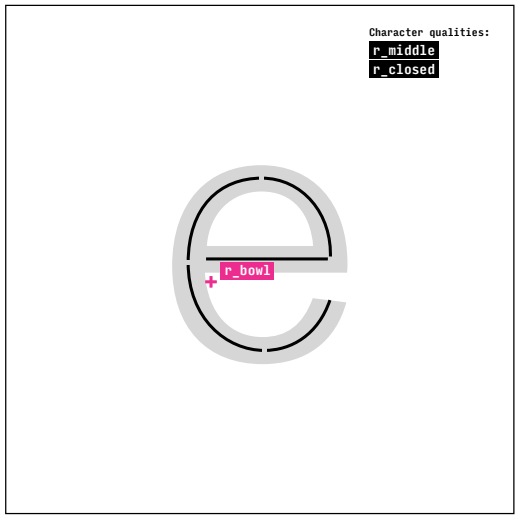
cyrl.sans.ve



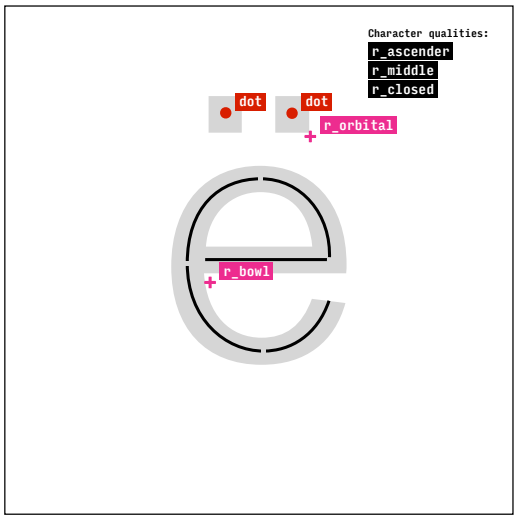
cyrl.sans.ghe



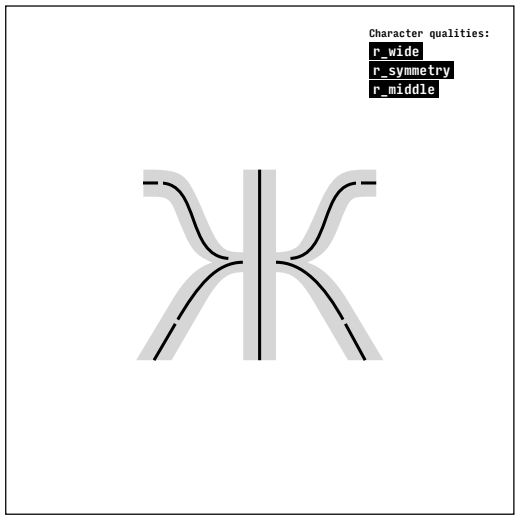
cyrl.sans.de



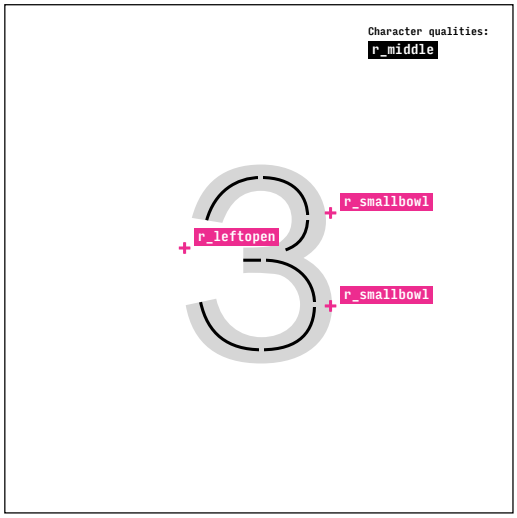
cyrl.sans.ie



cyrl.sans.io



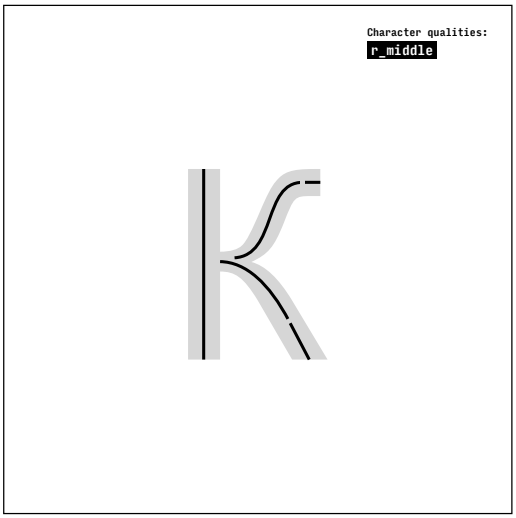
cyrl.sans.zhe



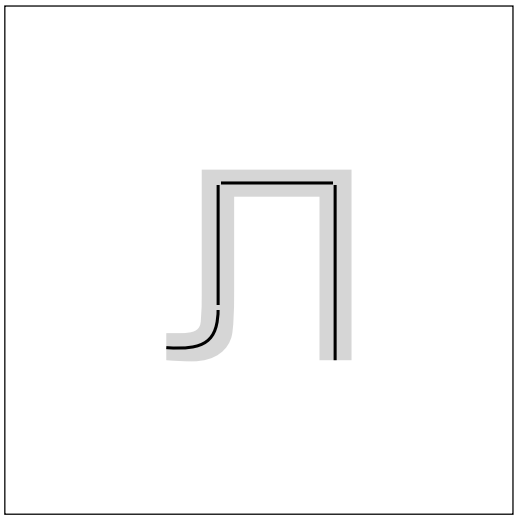
cyrl.sans.ze

cyrl.sans.i

cyrl.sans.short-i

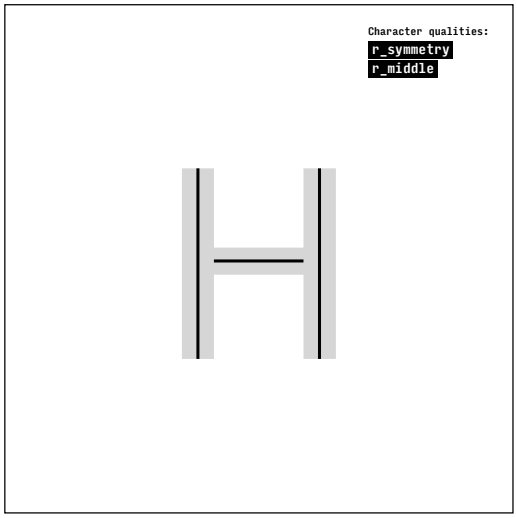


cyrl.sans.ka

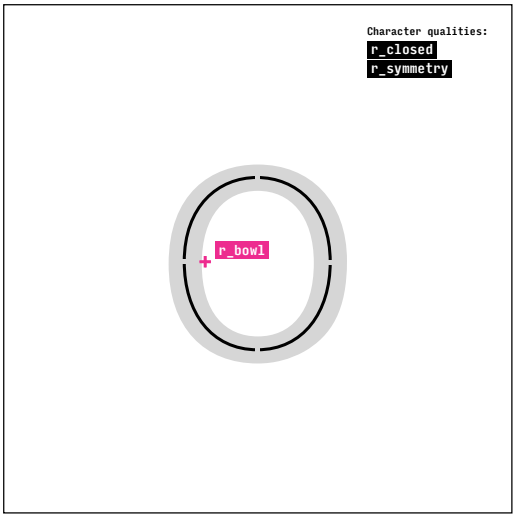


cyrl.sans.el

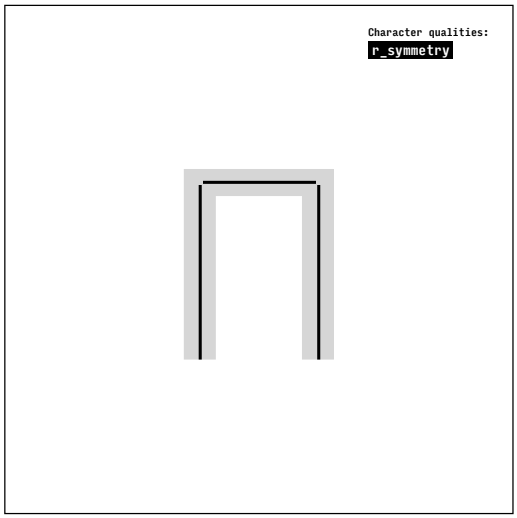
cyrl.sans.em



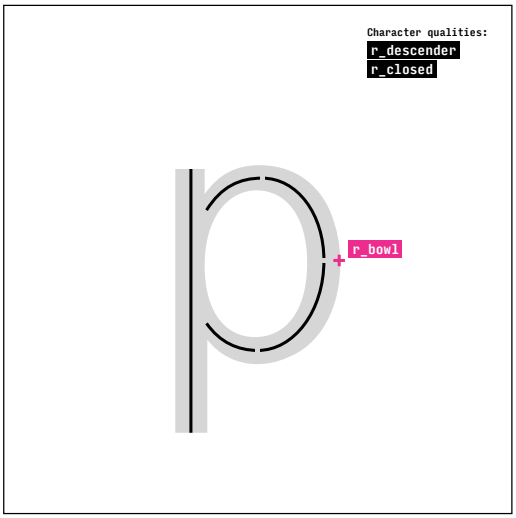
cyrl.sans.en



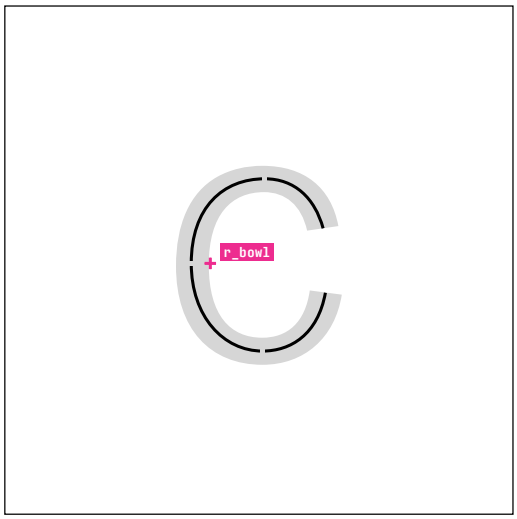
cyrl.sans.o



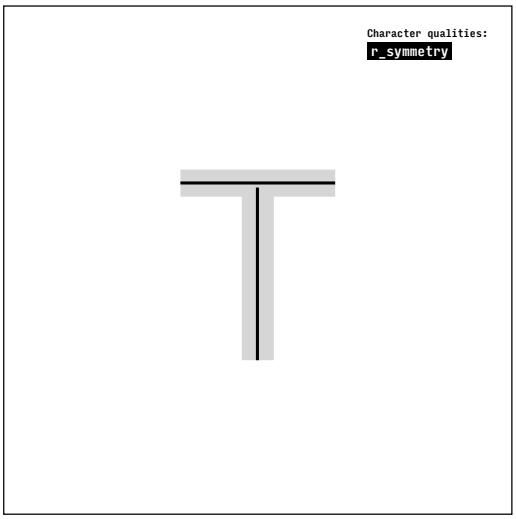
cyrl.sans.pe



cyrl.sans.er



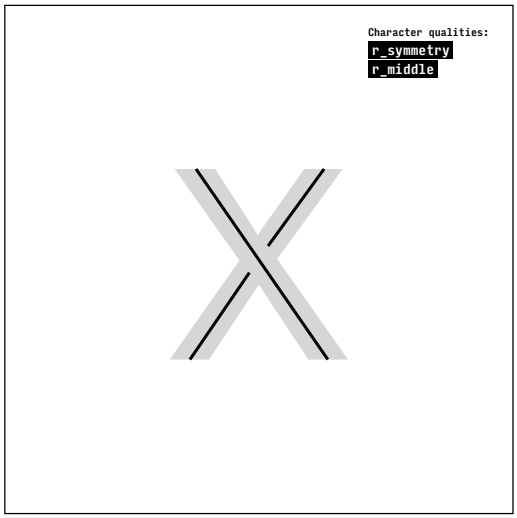
cyrl.sans.es



cyrl.sans.te

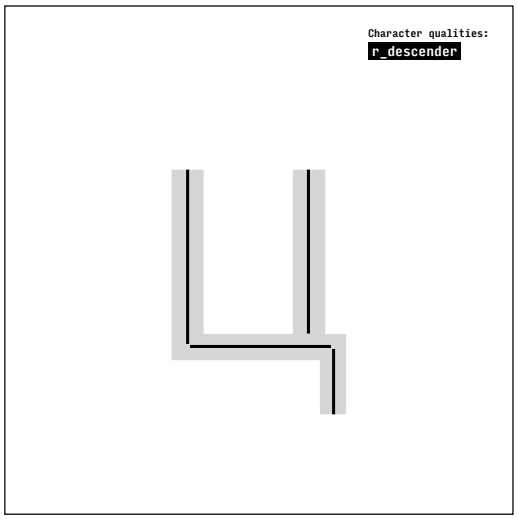
cyrl.sans.u

cyrl.sans.ef

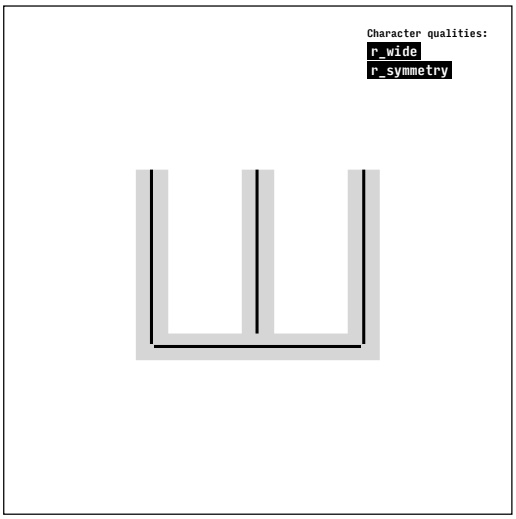


cyrl.sans.ha

cyrl.sans.che



cyrl.sans.tse



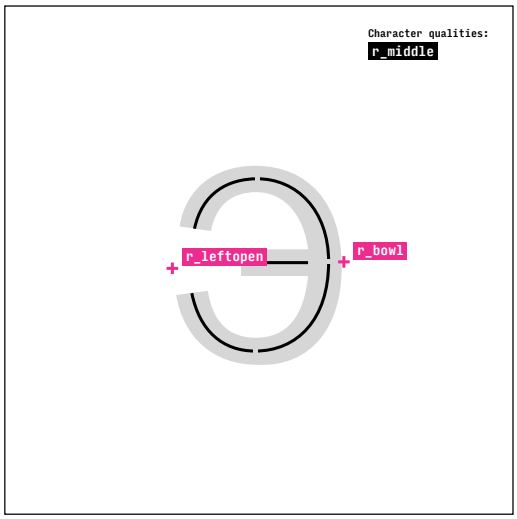
cyrl.sans.sha

cyrl.sans.shcha

cyrl.sans.hard-sign

cyrl.sans.yeru

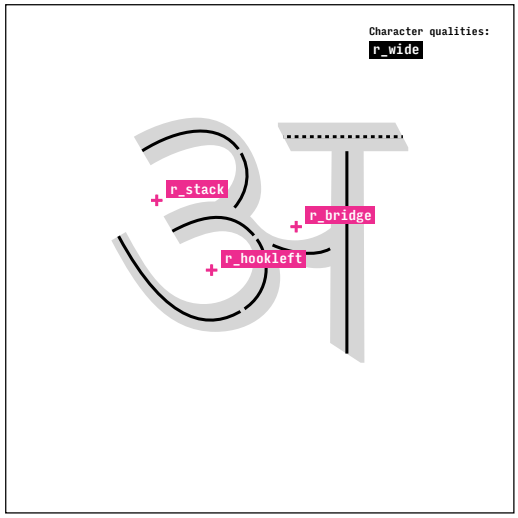
cyrl.sans.soft-sign



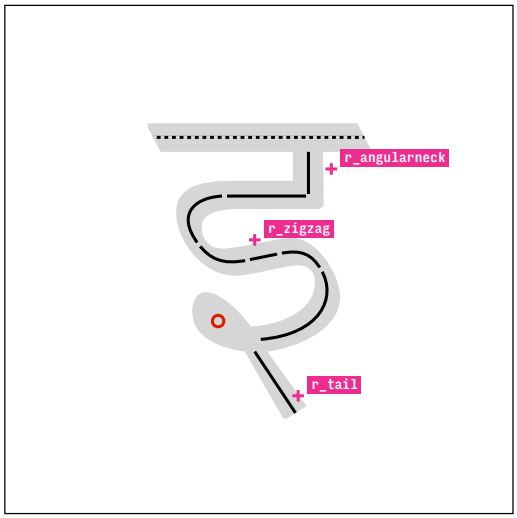
cyrl.sans.e

cyrl.sans.yu

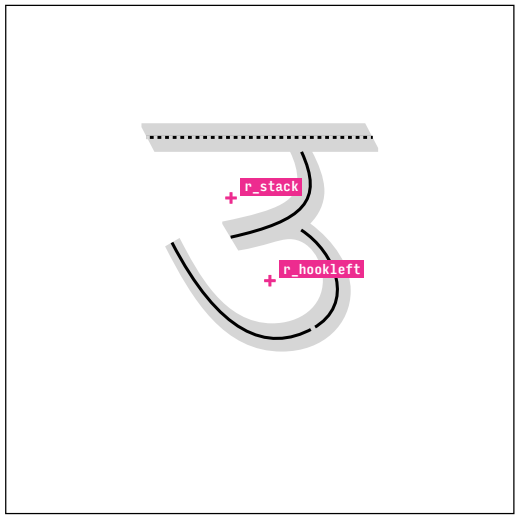
cyrl.sans.ya



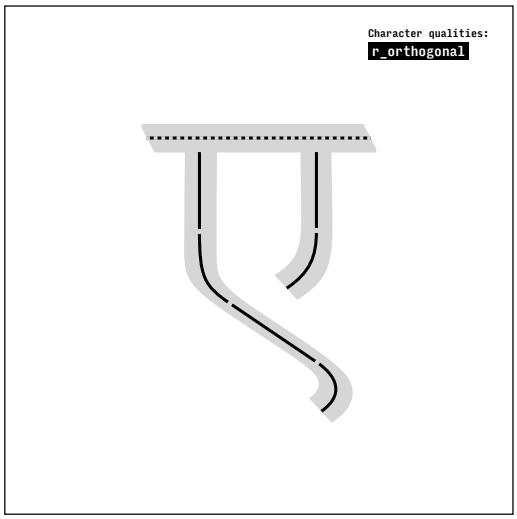
deva.high.A



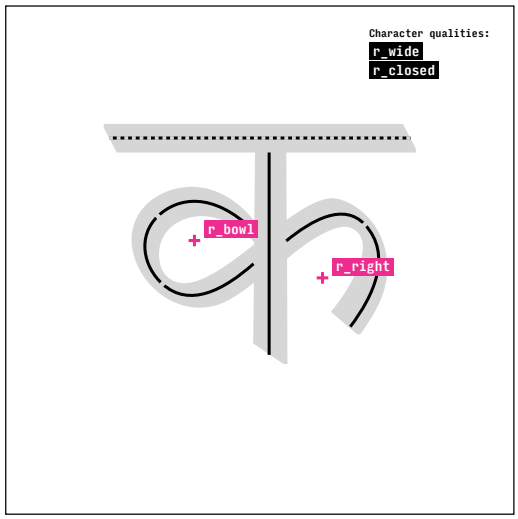
deva.high.I



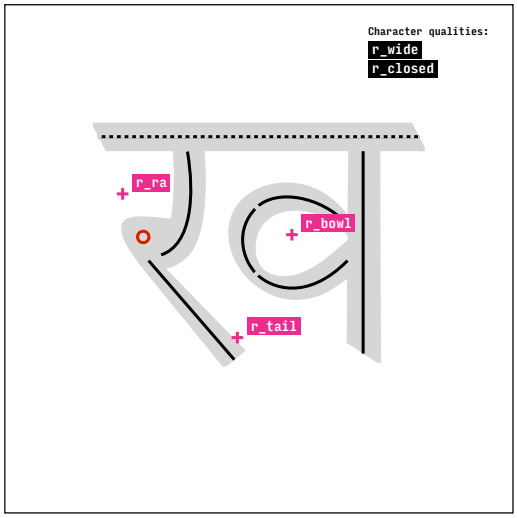
deva.high.U



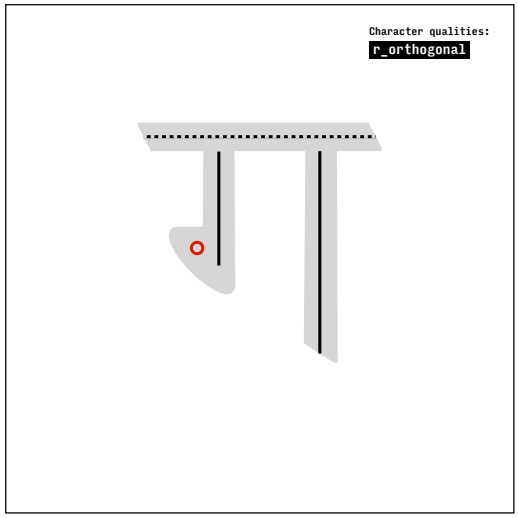
deva.high.E



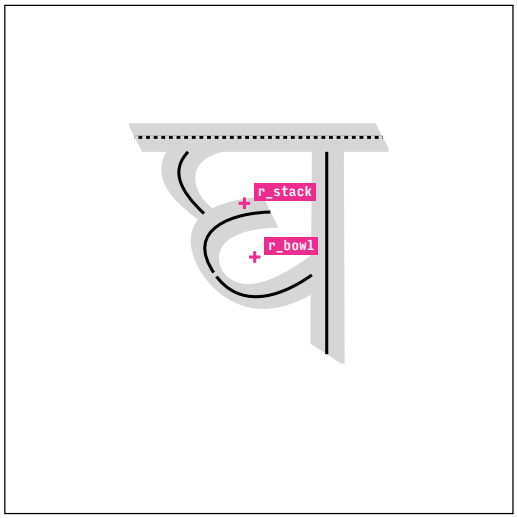
deva.high.Ka



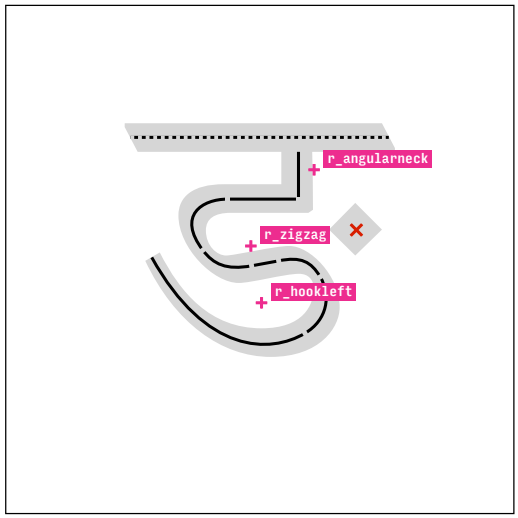
deva.high.Kha



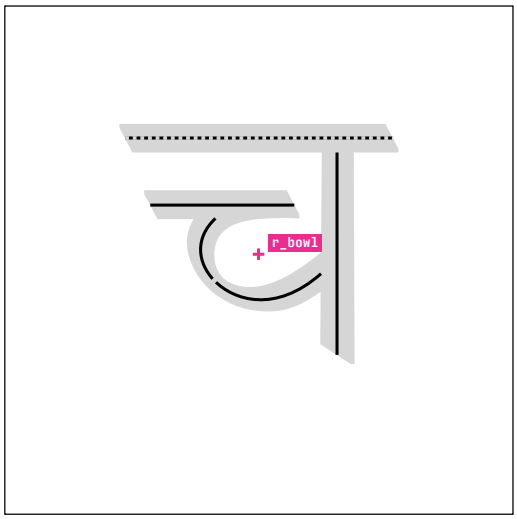
deva.high.Ga



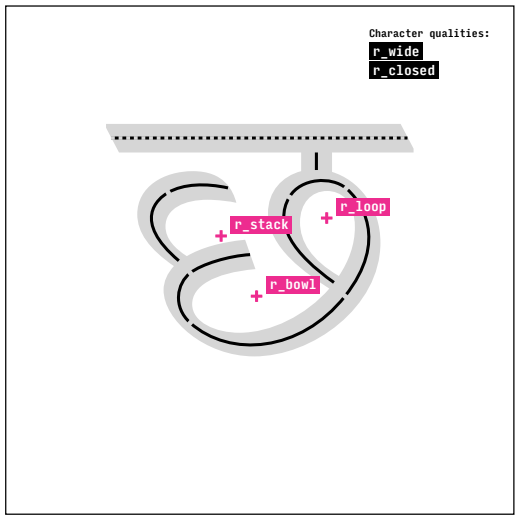
deva.high.Gha



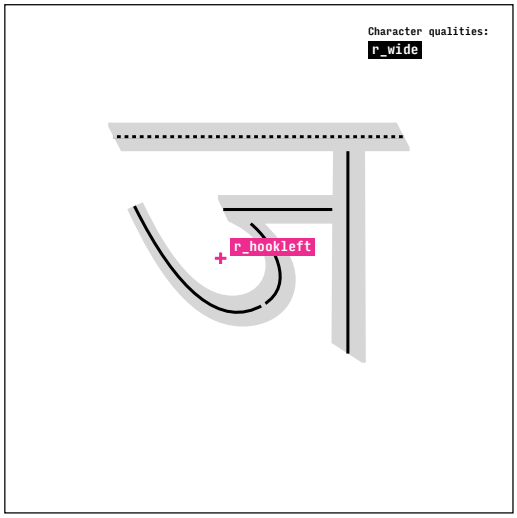
deva.high.Nga



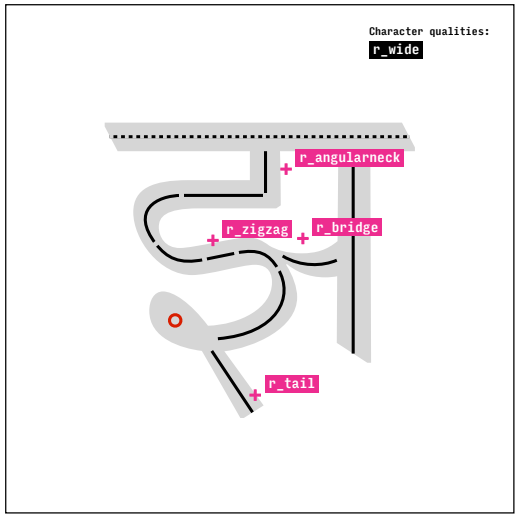
deva.high.Ca



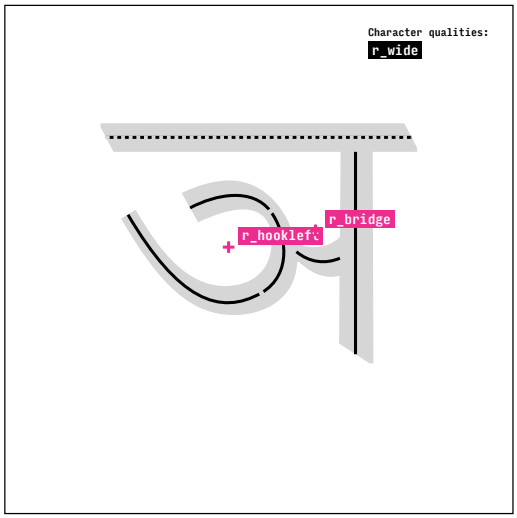
deva.high.Cha



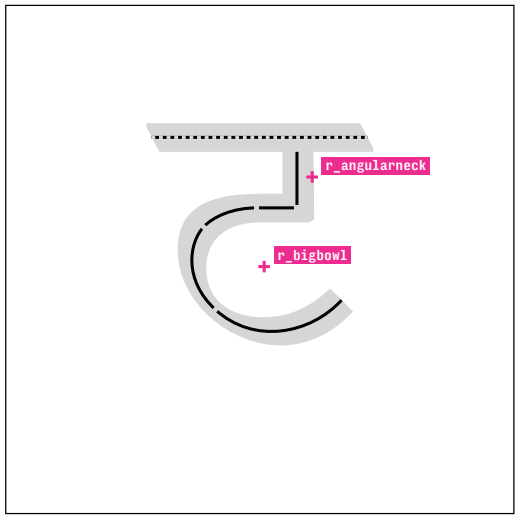
deva.high.Ja



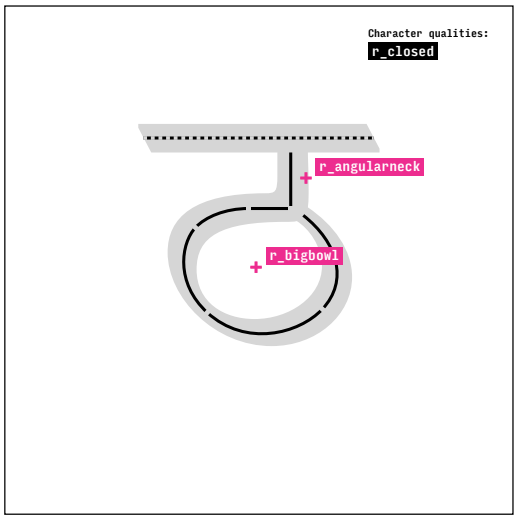
deva.high.Jha



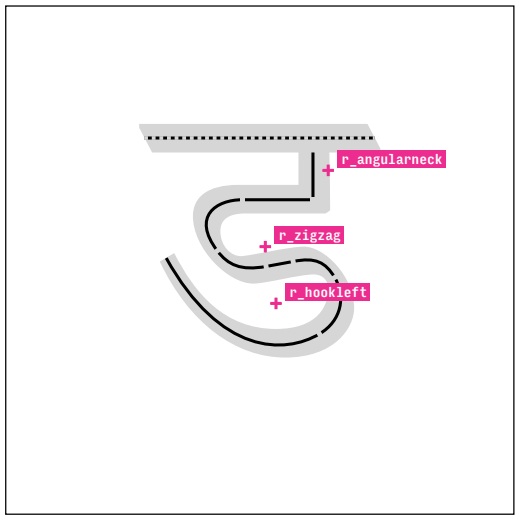
deva.high.Nya



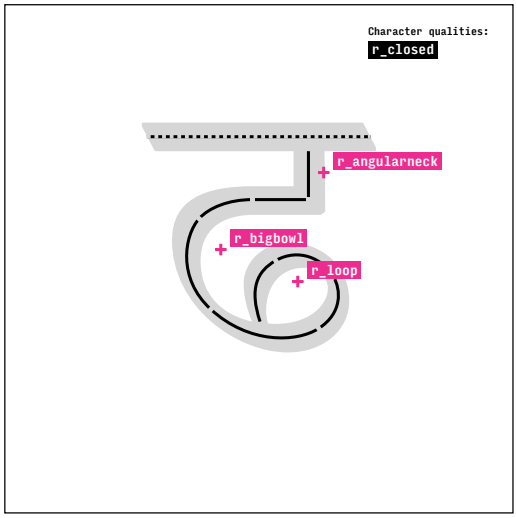
deva.high.Tta



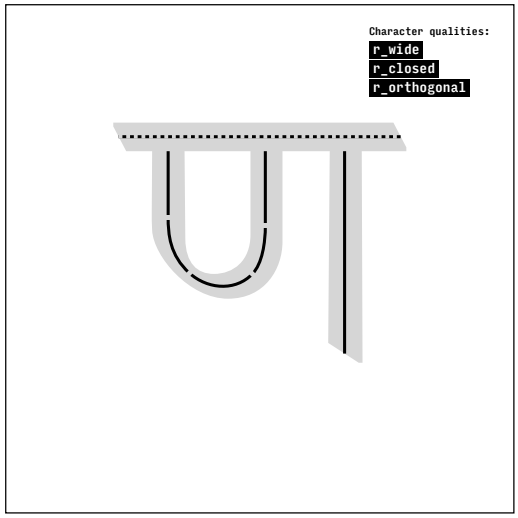
deva.high.Ttha



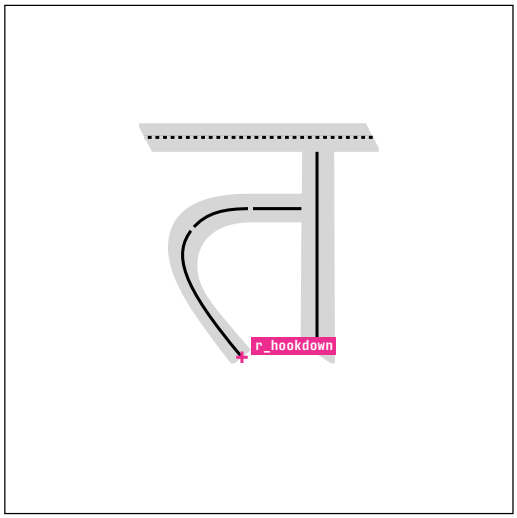
deva.high.Dda



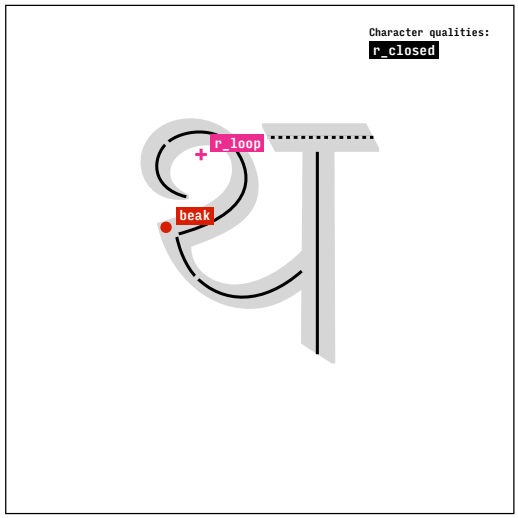
deva.high.Ddha



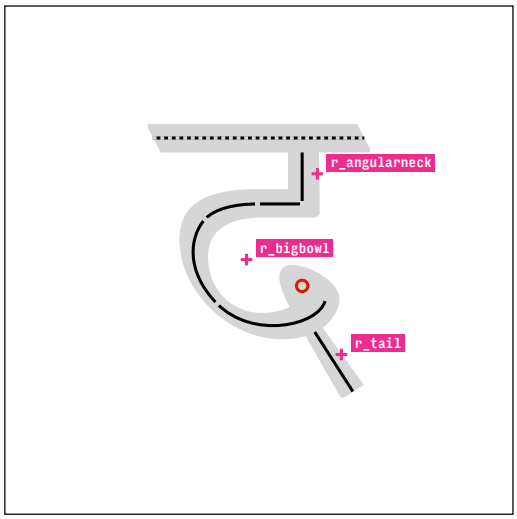
deva.high.Nna



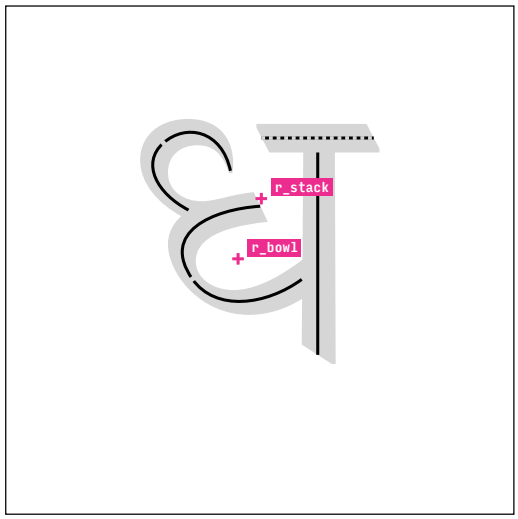
deva.high.Ta



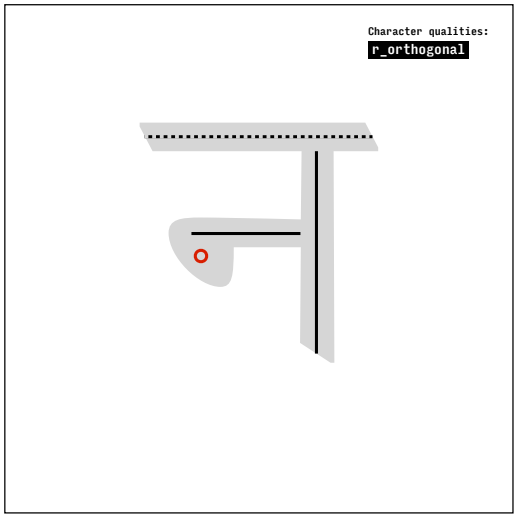
deva.high.Tha



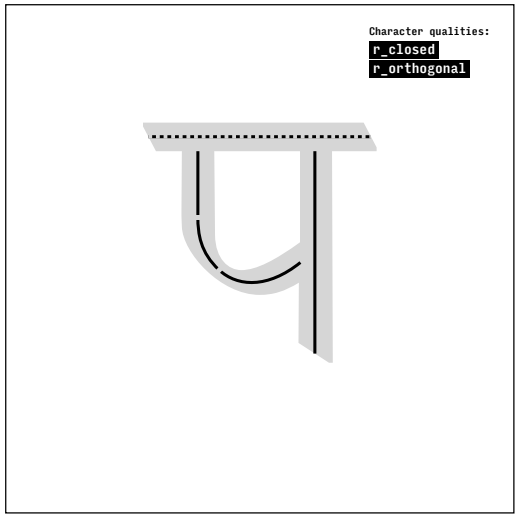
deva.high.Da



deva.high.Dha

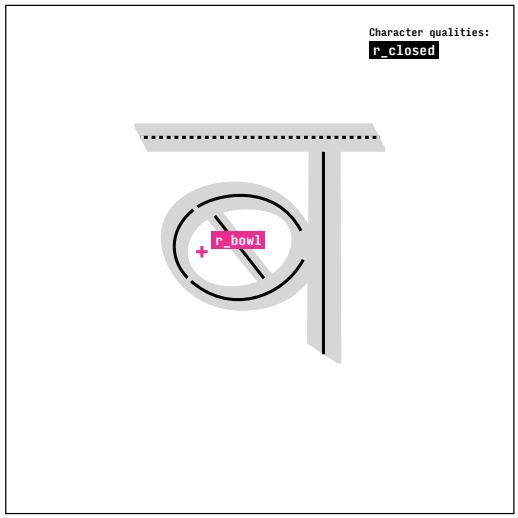


deva.high.Na

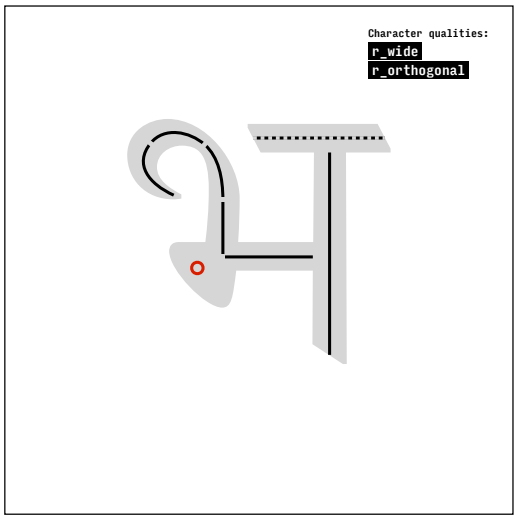


deva.high.Pa

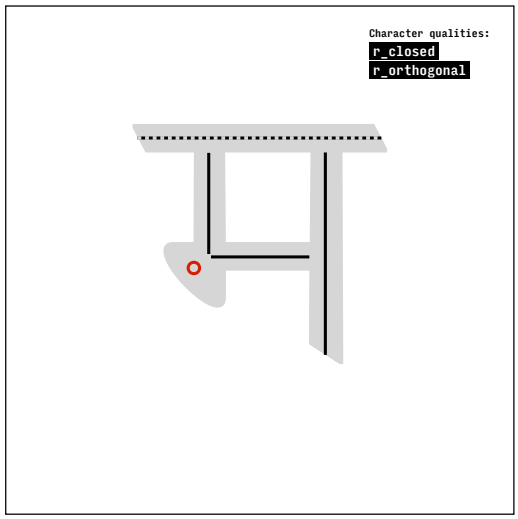
deva.high.Pha



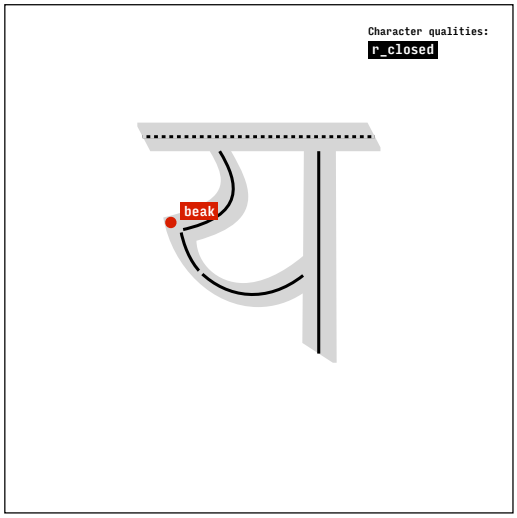
deva.high.Ba



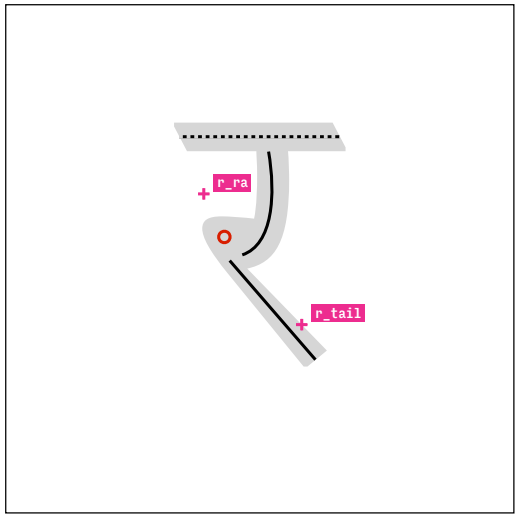
deva.high.Bha



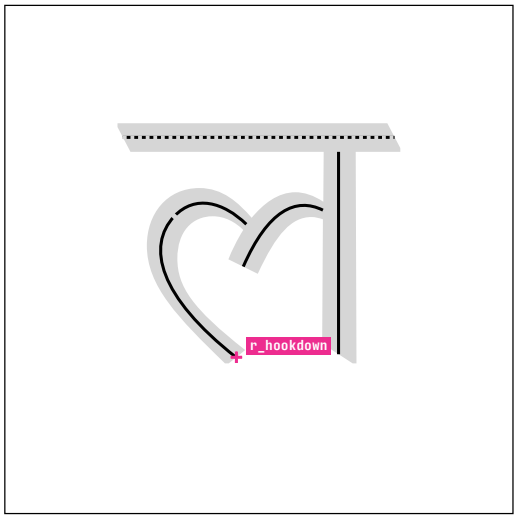
deva.high.Ma



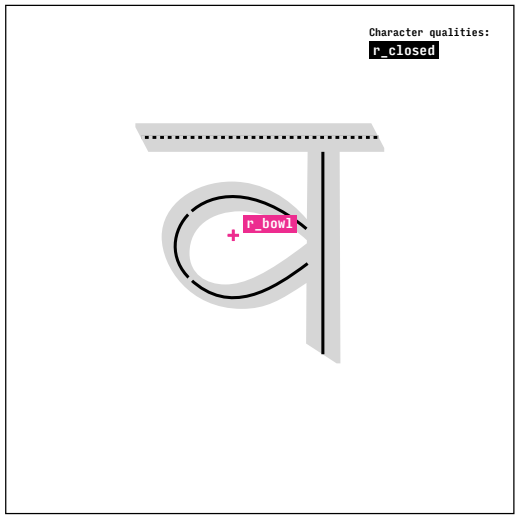
deva.high.Ya



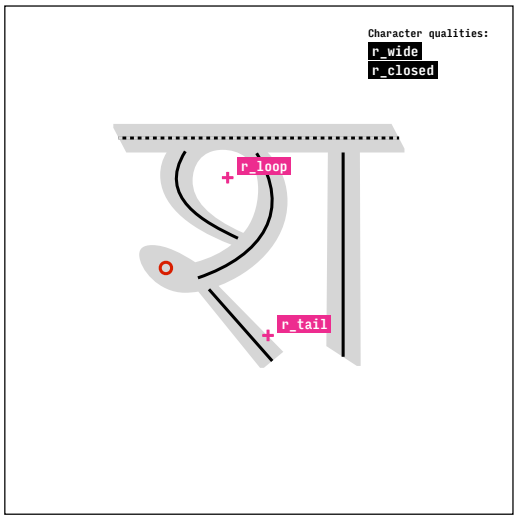
deva.high.Ra



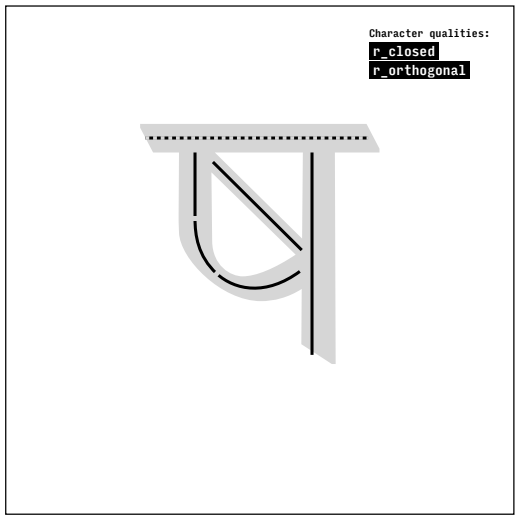
deva.high.La



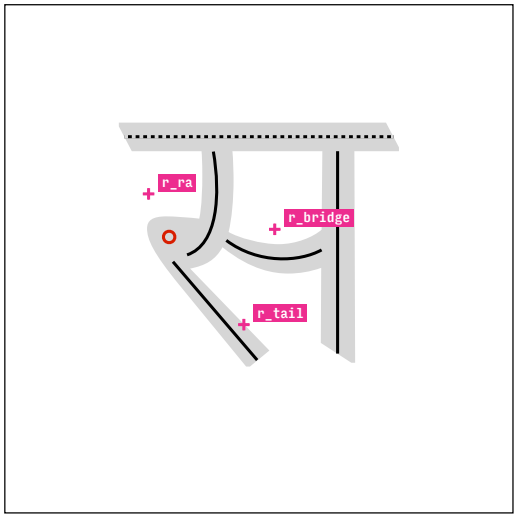
deva.high.Va



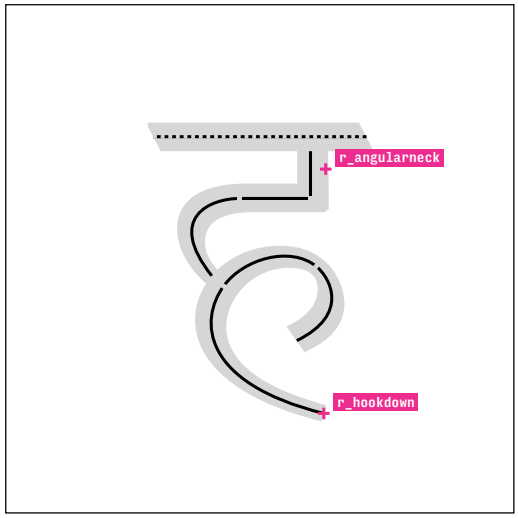
deva.high.Sha



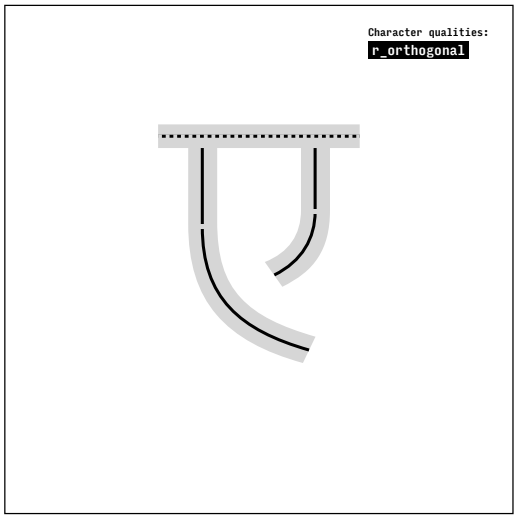
deva.high.Ssa



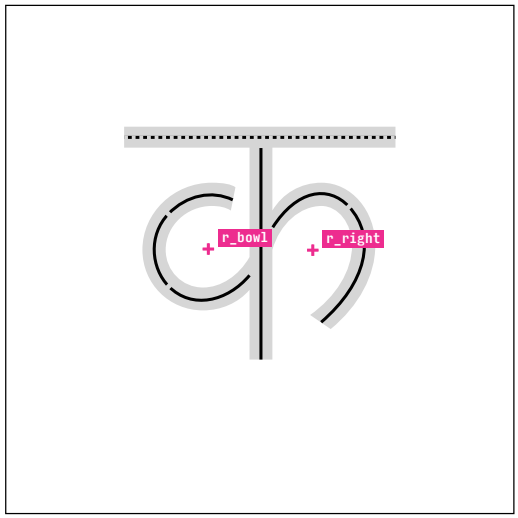
deva.high.Sa



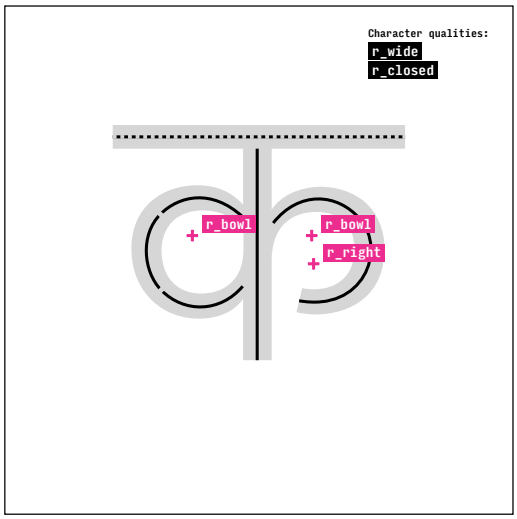
deva.high.Ha



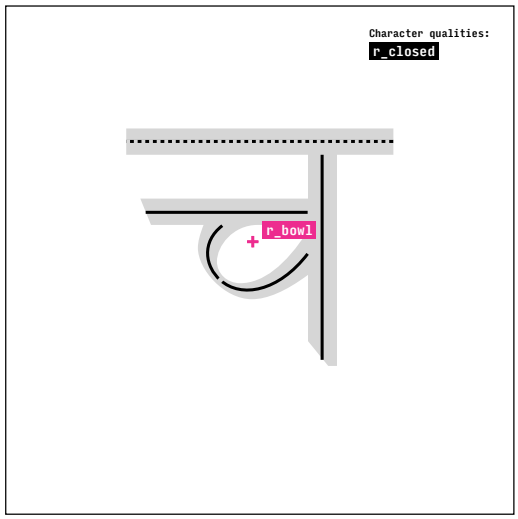
deva.low.E.simple



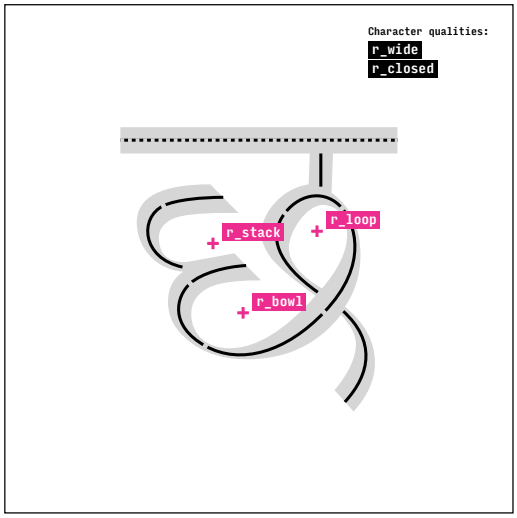
deva.low.Ka.open



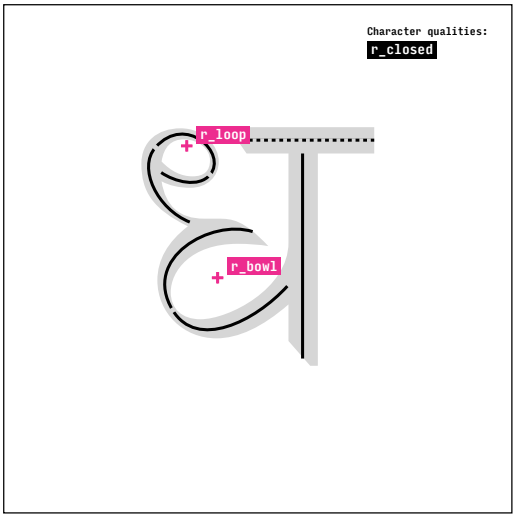
deva.low.Ka.symmetrical



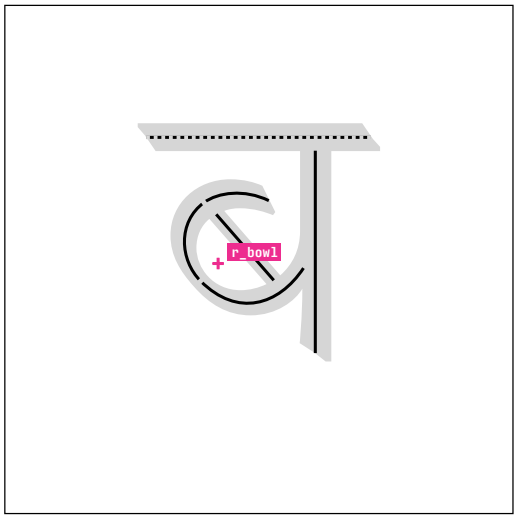
deva.high.Ca.closed



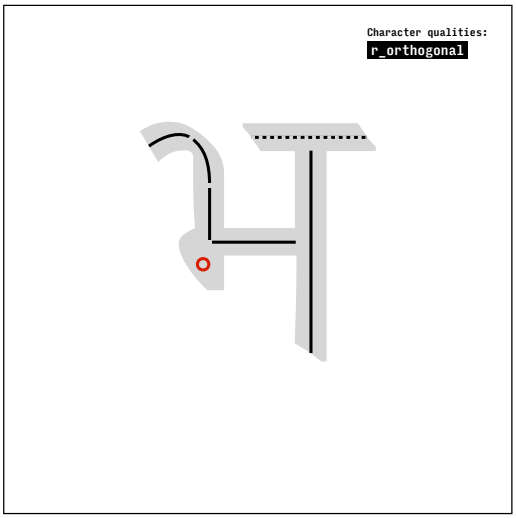
deva.high.Cha.withtail



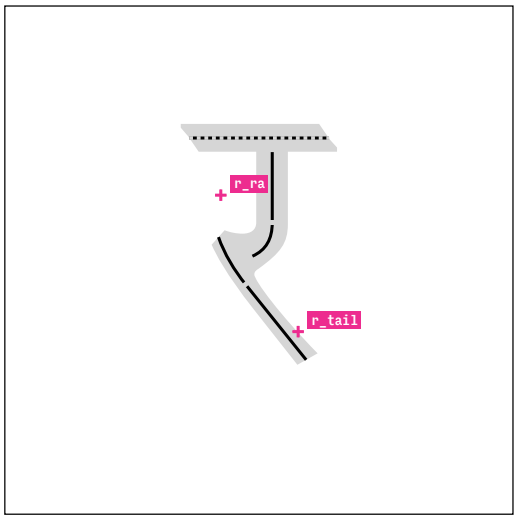
deva.high.Dha.withloop



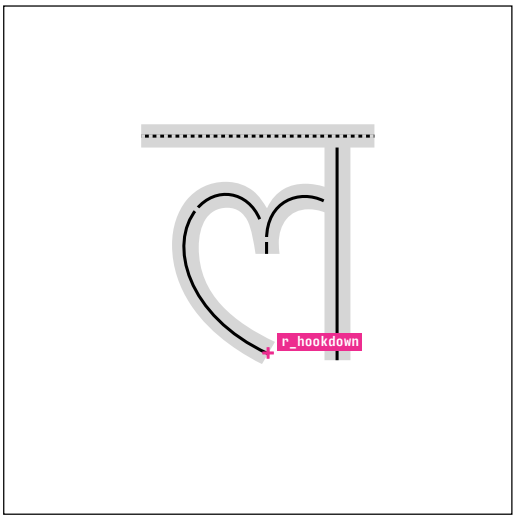
deva.high.Ba.open



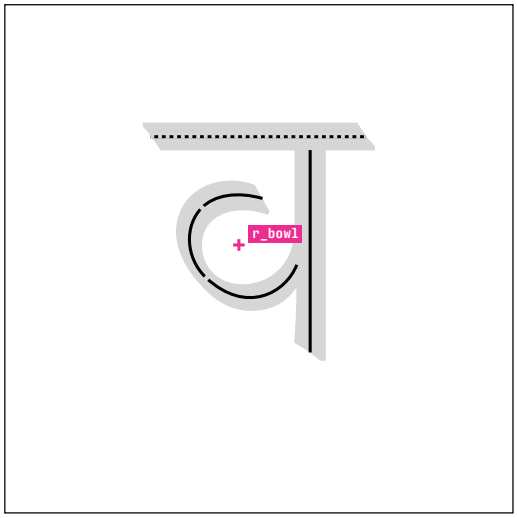
deva.high.Bha.withoutloop



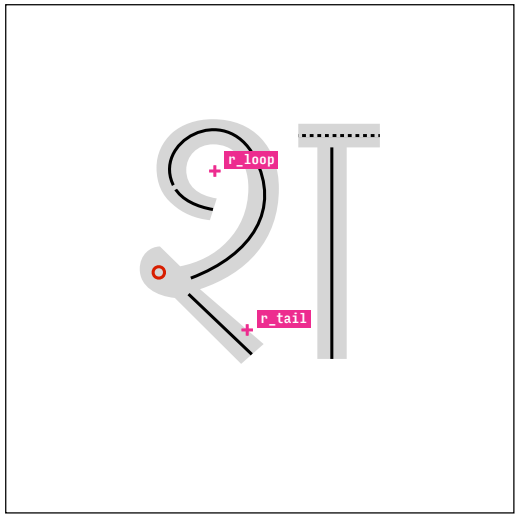
deva.high.Ra.withoutknot



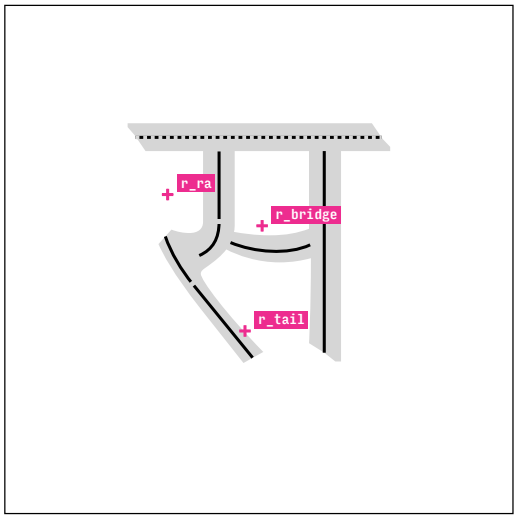
deva.low.La.verticalmiddle



deva.high.Va.open

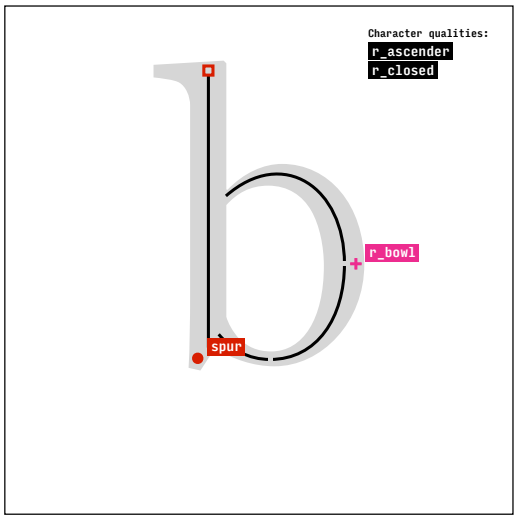


deva.low.Sha.open



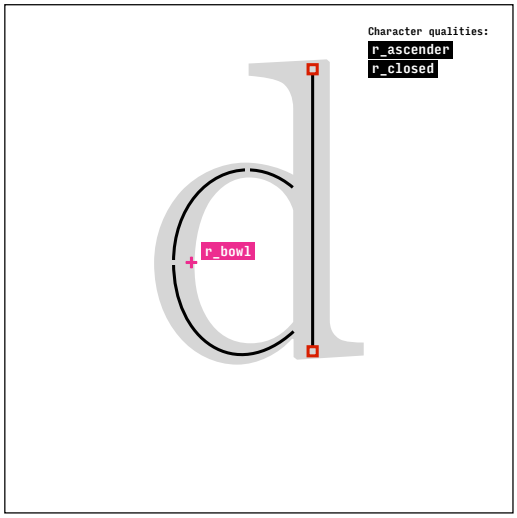
deva.high.Sa.withoutknot

latn.serif.a



latn.serif.b

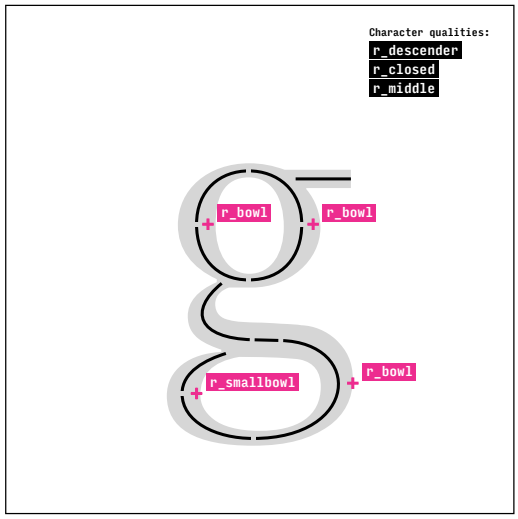
latn.serif.c



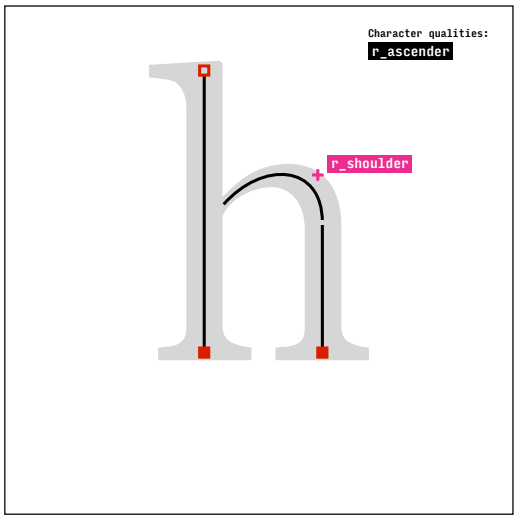
latn.serif.d

latn.serif.e

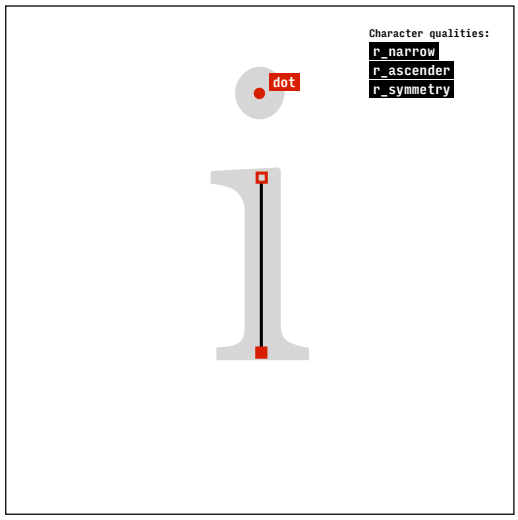
latn.serif.f



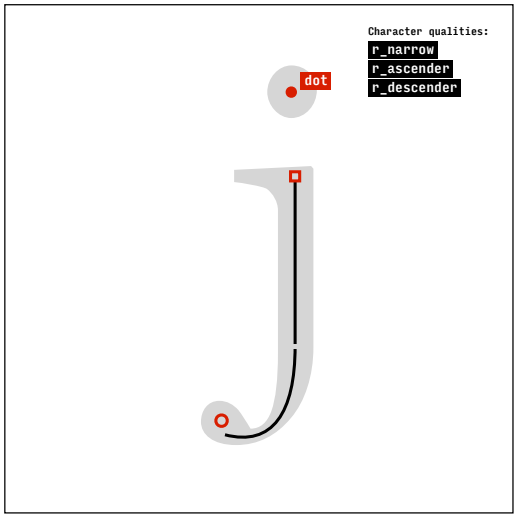
latn.serif.g



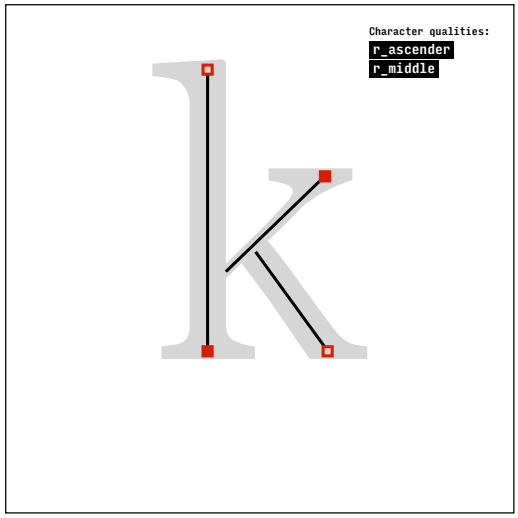
latn.serif.h



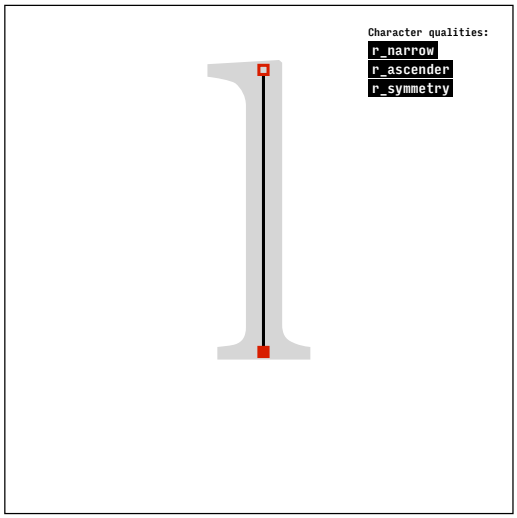
latn.serif.i



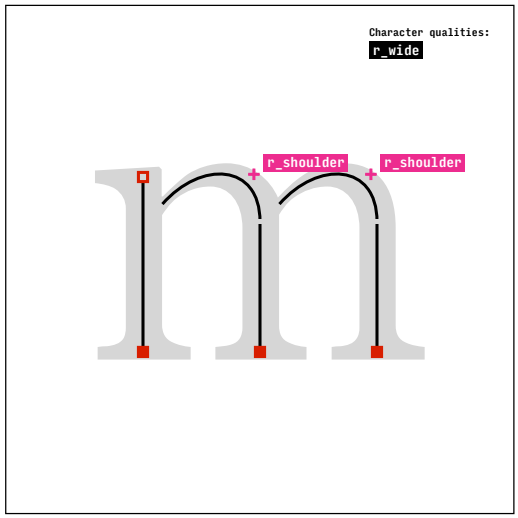
latn.serif.j



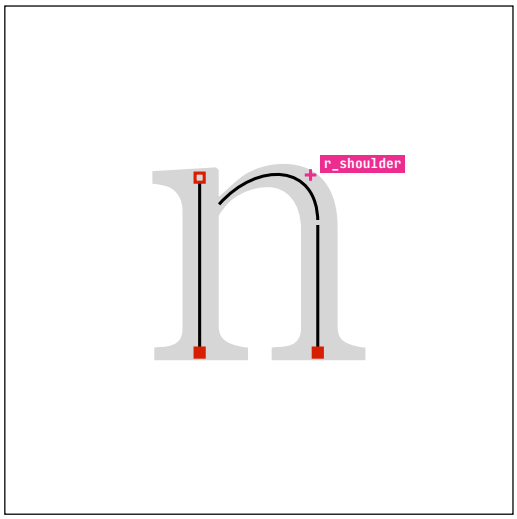
latn.serif.k



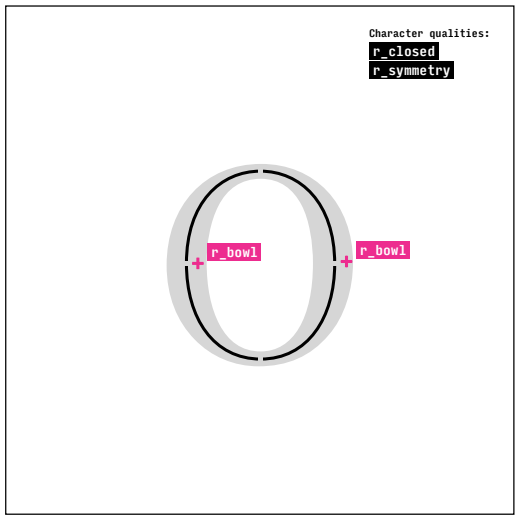
latn.serif.l



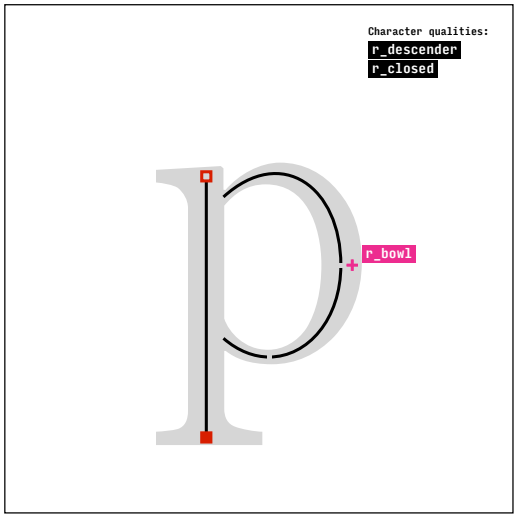
latn.serif.m



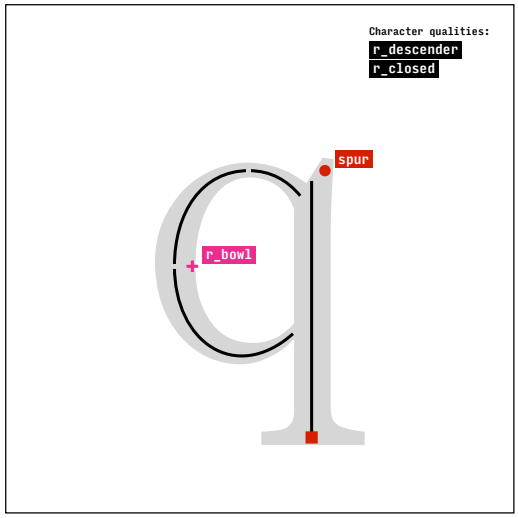
latn.serif.n



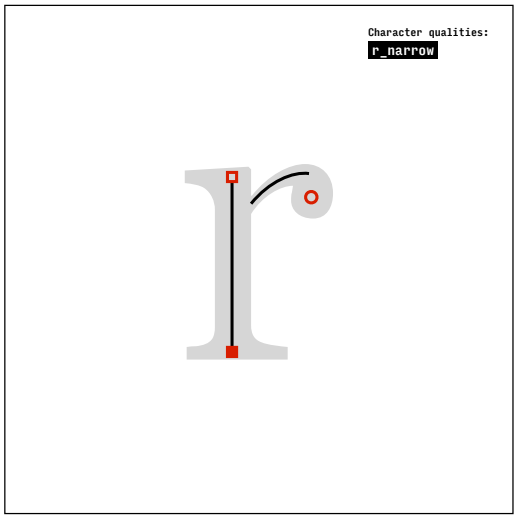
latn.serif.o



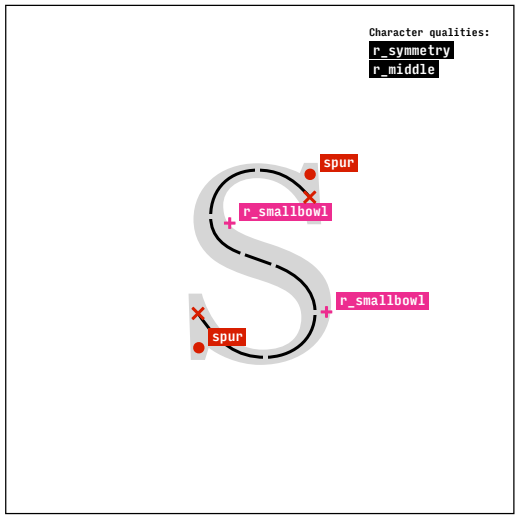
latn.serif.p



latn.serif.q

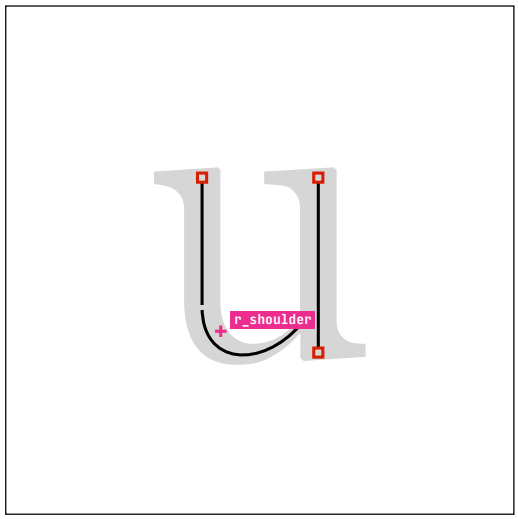


latn.serif.r

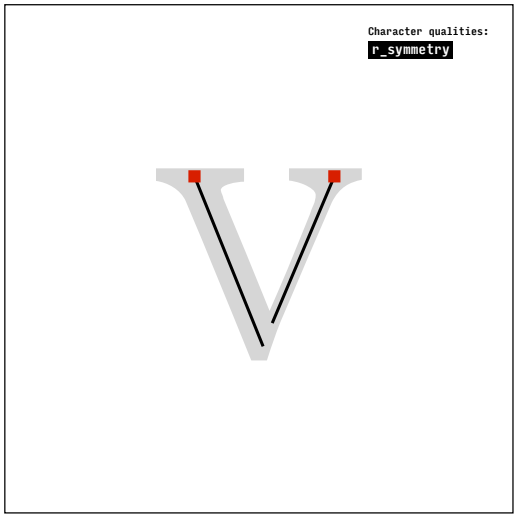


latn.serif.s

latn.serif.t

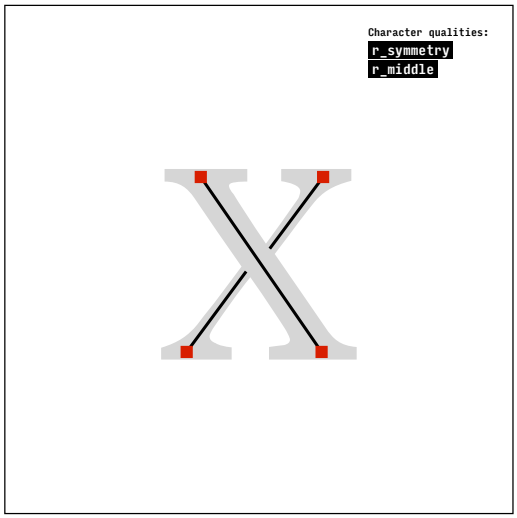


latn.serif.u

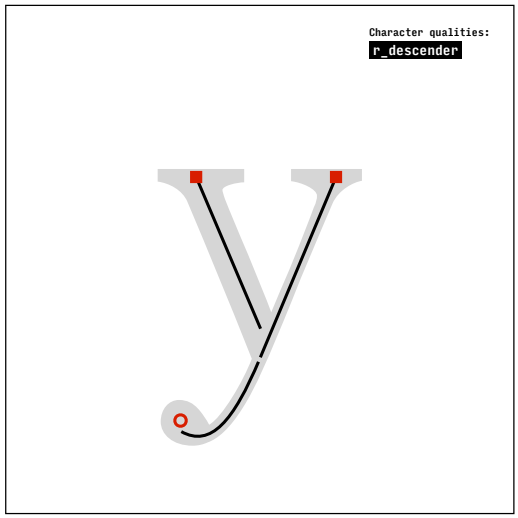


latn.serif.v

latn.serif.w

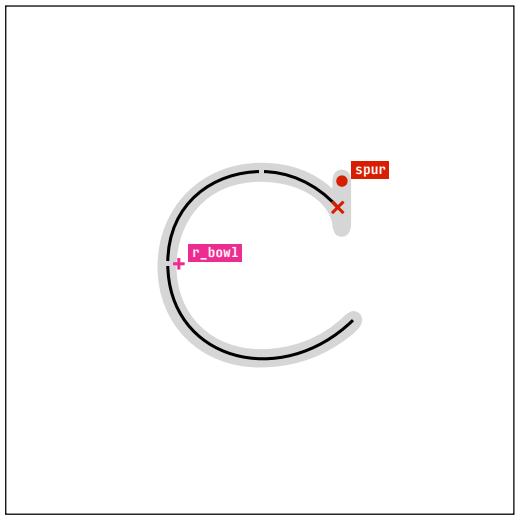


latn.serif.x



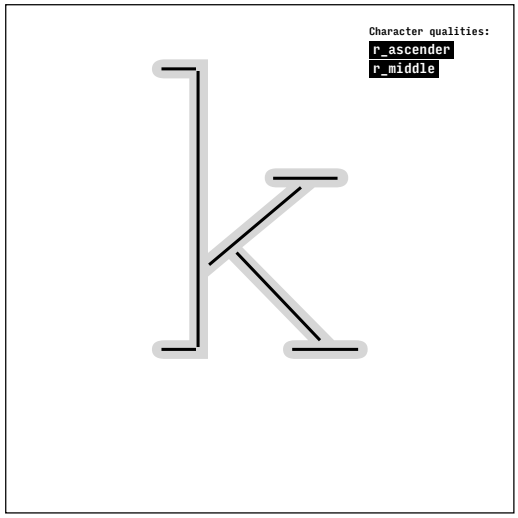
latn.serif.y

latn.serif.z

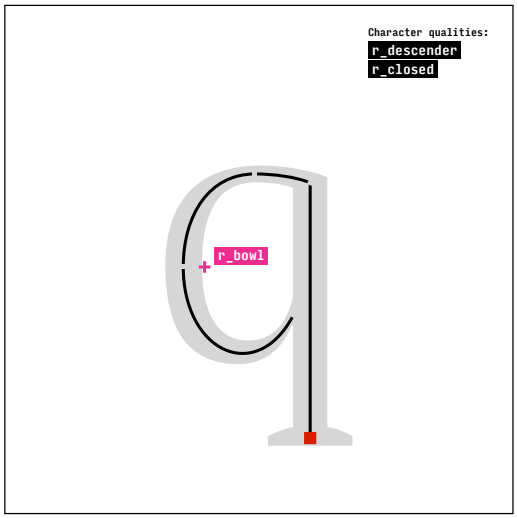


latn.serif.c.courier

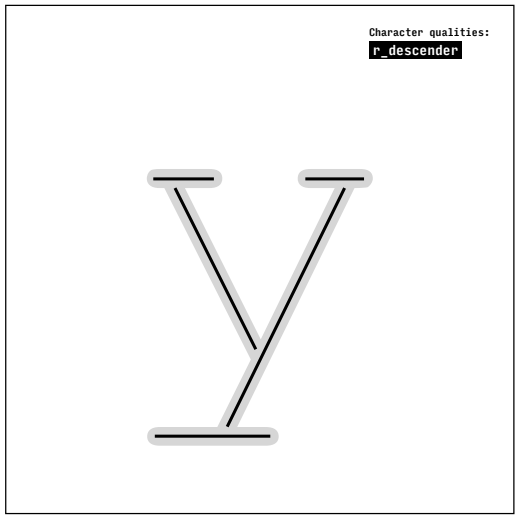
latn.serif.f.courier



latn.serif.k.courier

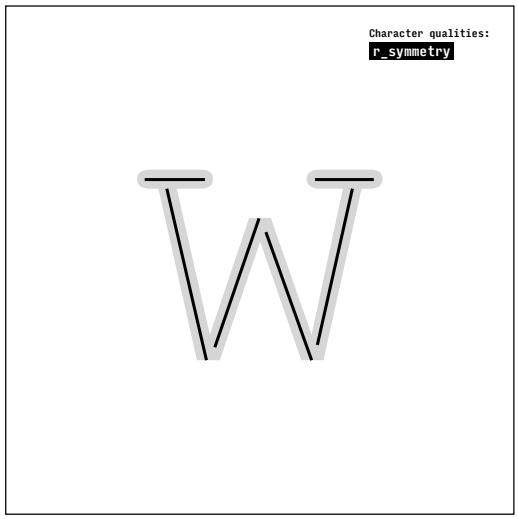


latn.serif.q.spurless



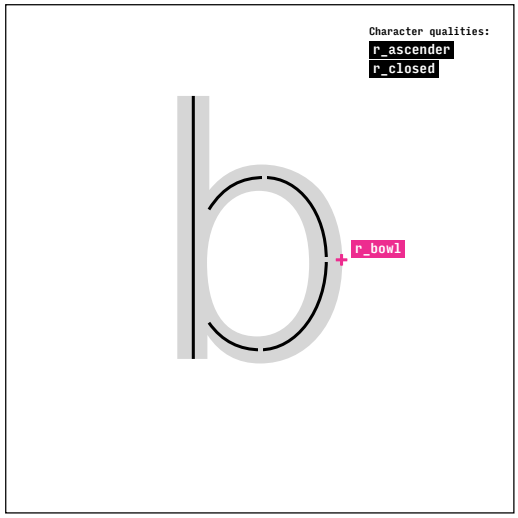
latn.serif.y.courier

latn.serif.h.courier

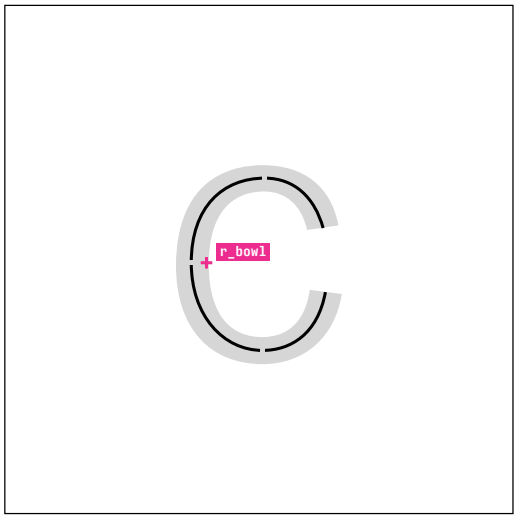


latn.serif.w.courier

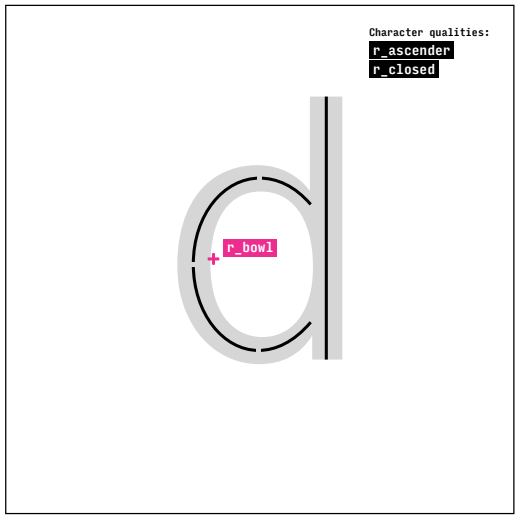
latn.sans.a



latn.sans.b



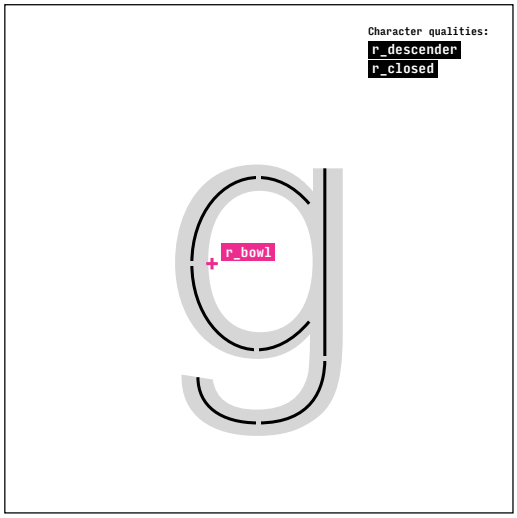
latn.sans.c



latn.sans.d

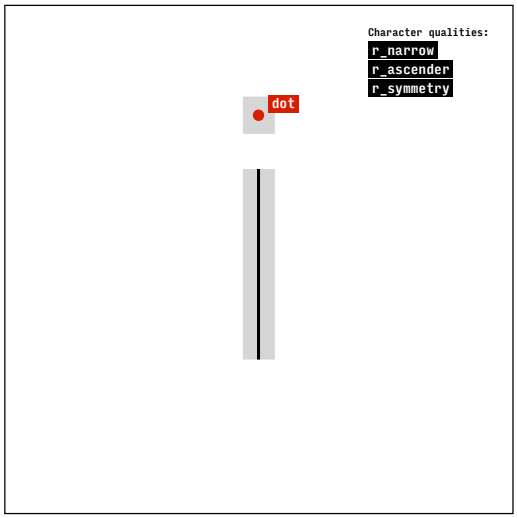
latn.sans.e

latn.sans.f

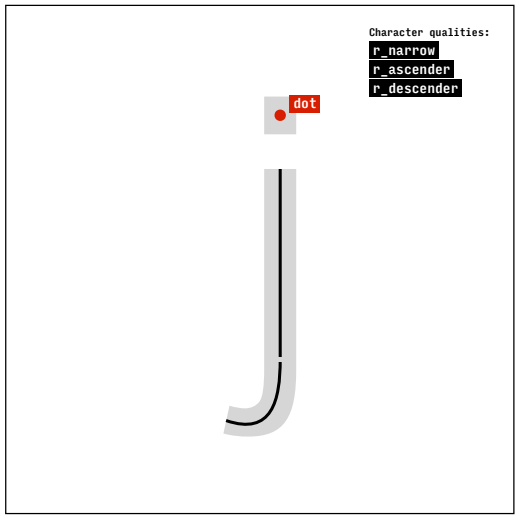


latn.sans.g

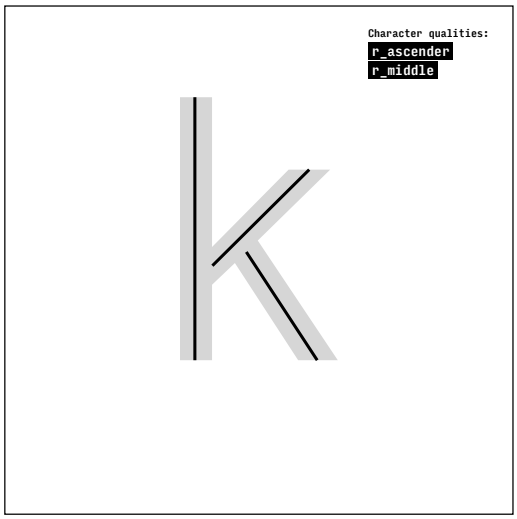
latn.sans.h



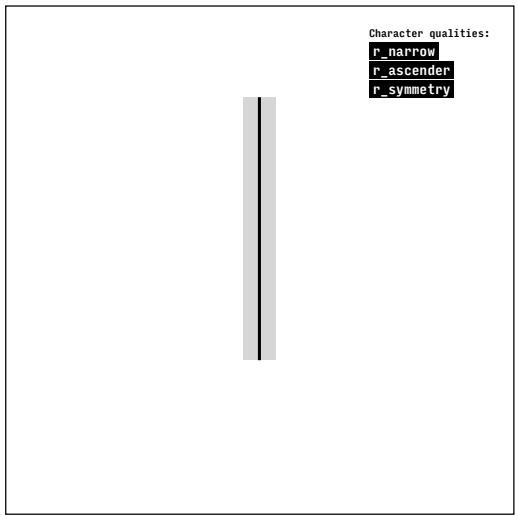
latn.sans.i



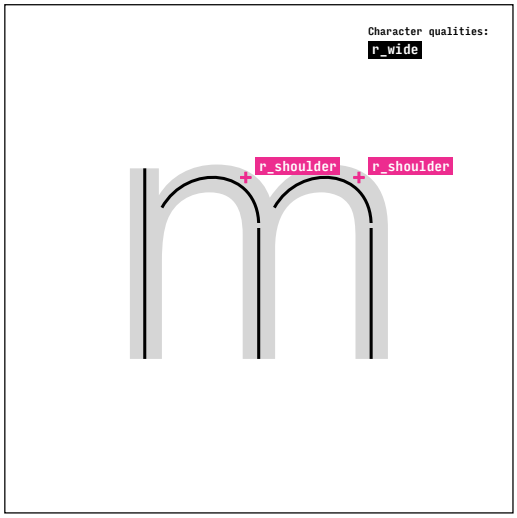
latn.sans.j



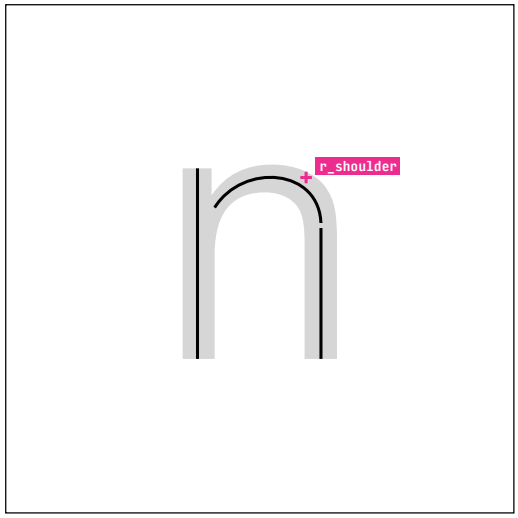
latn.sans.k



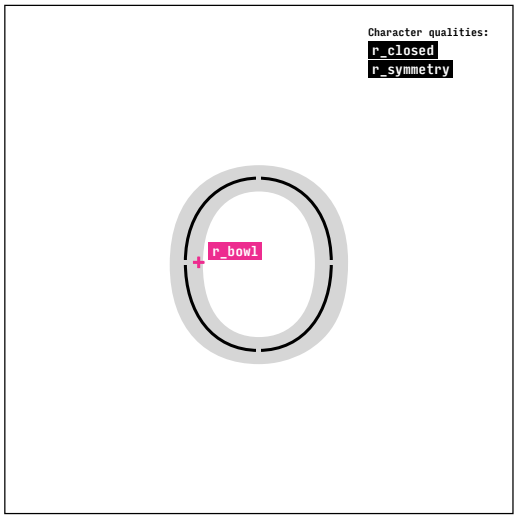
latn.sans.l



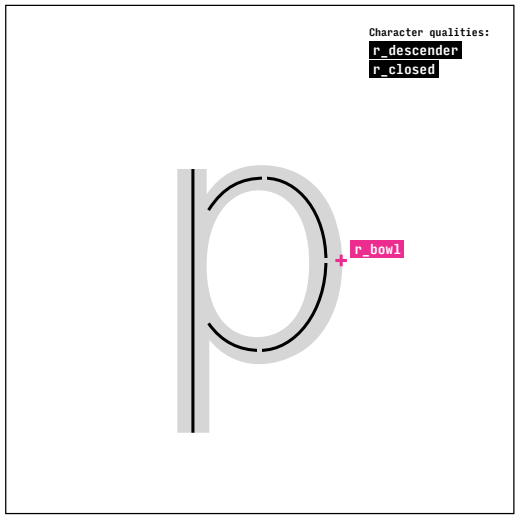
latn.sans.m



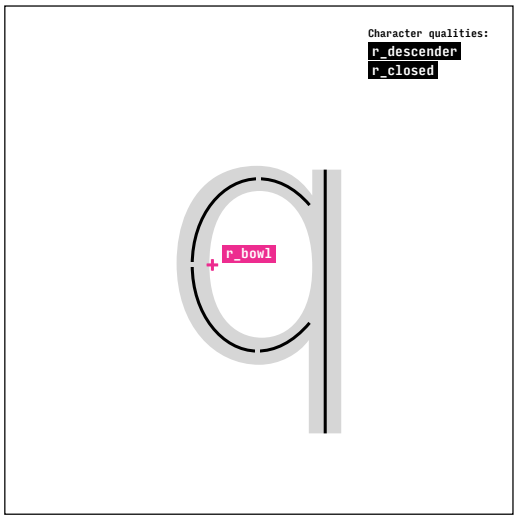
latn.sans.n



latn.sans.o



latn.sans.p

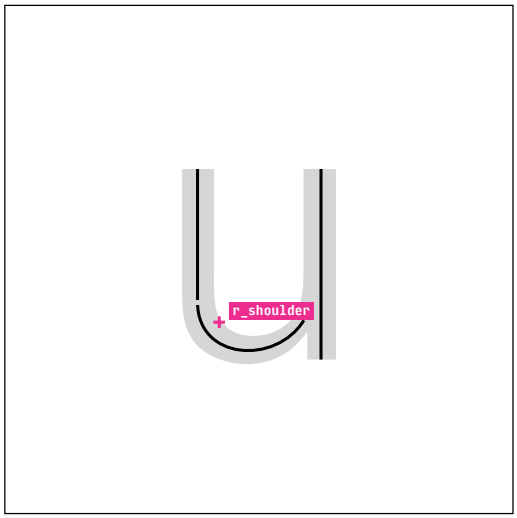


latn.sans.q

latn.sans.r

latn.sans.s

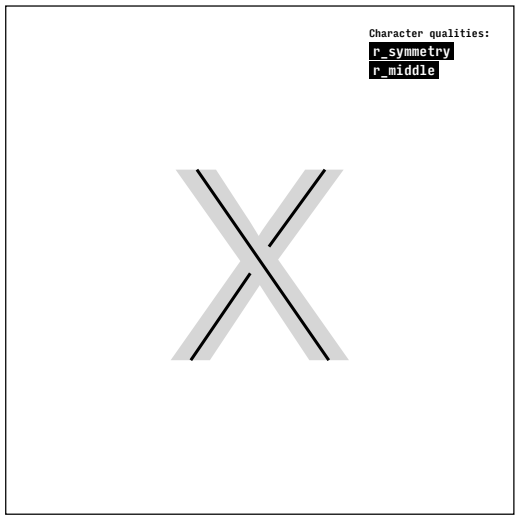
latn.sans.t



latn.sans.u

latn.sans.v

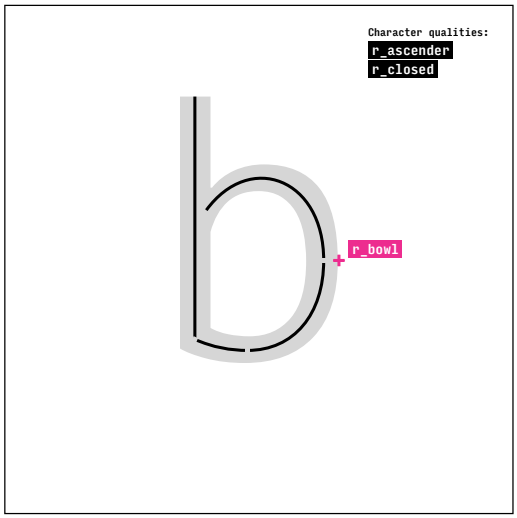
latn.sans.w



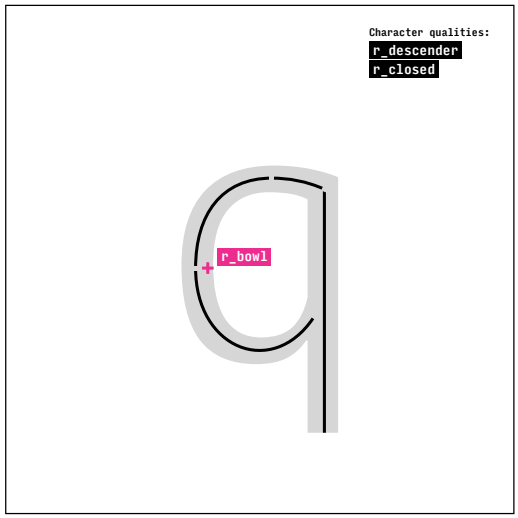
latn.sans.x

latn.sans.y

latn.sans.z



latn.sans.b.spurless



latn.sans.q.spurless

Appendix 8: typeface mappings

Typeface mappings in YAML format define which concepts should be used to represent particular character categories for the given typeface. These are required by the model described in chapter 6.

```
typeface:
  name: Adobe Devanagari
  characters:
    studied:
        deva.A: deva.high.A
        deva.Kha: deva.high.Kha
        deva.Gha: deva.high.Gha
        deva.Ta: deva.high.Ta
        deva.Ba: deva.high.Ba
        deva.Ra: deva.high.Ra
        deva.Ssa: deva.high.Ssa
        deva.Sa: deva.high.Sa

typeface:
  name: Arial
  characters:
    studied:
        latn.a: latn.sans.a
        latn.e: latn.sans.e
        latn.l: latn.sans.l
        latn.p: latn.sans.p
        latn.s: latn.sans.s
        latn.t: latn.sans.t
        latn.y: latn.sans.y
        latn.z: latn.sans.z
        cyrl.zhe: cyrl.sans.zhe
        cyrl.short-i: cyrl.sans.short-i
        cyrl.em: cyrl.sans.em
        cyrl.er: cyrl.sans.er
        cyrl.es: cyrl.sans.es
        cyrl.u: cyrl.sans.u
        cyrl.ef: cyrl.sans.ef
        cyrl.yu: cyrl.sans.yu
```

```
typeface:
  name: Calibri
  characters:
    studied:
        latn.d: latn.sans.d
        latn.i: latn.sans.i
        latn.j: latn.sans.j
        latn.o: latn.sans.o
        latn.p: latn.sans.p
        latn.r: latn.sans.r
        latn.v: latn.sans.v
        latn.x: latn.sans.x

typeface:
  name: Cambria
  characters:
    studied:
        latn.b: latn.serif.b
        latn.f: latn.serif.f
        latn.i: latn.serif.i
        latn.m: latn.serif.m
        latn.n: latn.serif.n
        latn.p: latn.serif.p
        latn.q: latn.serif.q
        latn.z: latn.serif.z

typeface:
  name: Candara
  characters:
    studied:
        latn.b: latn.sans.b
        latn.d: latn.sans.d
        latn.i: latn.sans.i
        latn.j: latn.sans.j
        latn.k: latn.sans.k
        latn.l: latn.sans.l
        latn.v: latn.sans.v
        latn.y: latn.sans.y
```

```
typeface:
  name: Century Schoolbook
  characters:
    studied:
        latn.d: latn.serif.d
        latn.i: latn.serif.i
        latn.j: latn.serif.j
        latn.o: latn.serif.o
        latn.p: latn.serif.p
        latn.r: latn.serif.r
        latn.v: latn.serif.v
        latn.x: latn.serif.x
        cyrl.a: cyrl.serif.a
        cyrl.ghe: cyrl.serif.ghe
        cyrl.de: cyrl.serif.de
        cyrl.io: cyrl.serif.io
        cyrl.zhe: cyrl.serif.zhe.withflick
        cyrl.ka: cyrl.serif.ka.withflick
        cyrl.pe: cyrl.serif.pe
        cyrl.ya: cyrl.serif.ya.withflick

typeface:
  name: Courier New
  characters:
    studied:
        latn.c: latn.serif.c.courier
        latn.e: latn.serif.e
        latn.f: latn.serif.f.courier
        latn.h: latn.serif.h.courier
        latn.k: latn.serif.k.courier
        latn.o: latn.serif.o
        latn.w: latn.serif.w.courier
        latn.y: latn.serif.y.courier
        cyrl.de: cyrl.serif.de.courier
        cyrl.ka: cyrl.serif.ka.courier
        cyrl.el: cyrl.serif.el.courier
        cyrl.o: cyrl.serif.o
        cyrl.es: cyrl.serif.es.courier
        cyrl.tse: cyrl.serif.tse.courier
        cyrl.yu: cyrl.serif.yu.courier
        cyrl.ya: cyrl.serif.ya.courier
```

```
typeface:
  name: Devanagari MT
  characters:
    studied:
        deva.U: deva.high.U
        deva.Gha: deva.high.Gha
        deva.Ca: deva.high.Ca.closed
        deva.Cha: deva.high.Cha.withtail
        deva.Tta: deva.high.Tta
        deva.Dda: deva.high.Dda
        deva.Dha: deva.high.Dha.withloop
        deva.Ha: deva.high.Ha

typeface:
  name: Ek Mukta
  characters:
    studied:
        deva.E: deva.low.E.simple
        deva.Ka: deva.low.Ka.symmetrical
        deva.Ja: deva.high.Ja
        deva.Ddha: deva.high.Ddha
        deva.Na: deva.high.Na
        deva.Bha: deva.high.Bha
        deva.Ma: deva.high.Ma
        deva.Sha: deva.low.Sha.open

typeface:
  name: Futura
  characters:
    studied:
        latn.g: latn.sans.g
        latn.h: latn.sans.h
        latn.m: latn.sans.m
        latn.n: latn.sans.n
        latn.q: latn.sans.q
        latn.r: latn.sans.r
        latn.y: latn.sans.y
        latn.z: latn.sans.z
```

```
typeface:
  name: Georgia
  characters:
    studied:
        latn.b: latn.serif.b
        latn.d: latn.serif.d
        latn.i: latn.serif.i
        latn.j: latn.serif.j
        latn.k: latn.serif.k
        latn.l: latn.serif.l
        latn.v: latn.serif.v
        latn.y: latn.serif.y
        cyrl.be: cyrl.serif.be
        cyrl.ie: cyrl.serif.ie
        cyrl.ze: cyrl.serif.ze
        cyrl.i: cyrl.serif.i
        cyrl.em: cyrl.serif.em
        cyrl.en: cyrl.serif.en
        cyrl.ef: cyrl.serif.ef
        cyrl.che: cyrl.serif.che

typeface:
  name: ITF Devanagari
  characters:
    studied:
        deva.Ca: deva.high.Ca
        deva.Jha: deva.high.Jha
        deva.Nya: deva.high.Nya
        deva.Ttha: deva.high.Ttha
        deva.Bha: deva.high.Bha.withoutloop
        deva.Ma: deva.high.Ma
        deva.Ra: deva.high.Ra.withoutknot
        deva.Va: deva.high.Va.open
```

```
typeface:
  name: Kohinoor Devanagari
  characters:
    studied:
        deva.I: deva.high.I
        deva.Ga: deva.high.Ga
        deva.Dda: deva.high.Dda
        deva.Na: deva.high.Na
        deva.Pa: deva.high.Pa
        deva.Ba: deva.high.Ba.open
        deva.Sa: deva.high.Sa.withoutknot
        deva.Ha: deva.high.Ha

typeface:
  name: Lohit Devanagari
  characters:
    studied:
        deva.Ka: deva.low.Ka.open
        deva.Ga: deva.high.Ga
        deva.Nna: deva.high.Nna
        deva.Tha: deva.high.Tha
        deva.Pa: deva.high.Pa
        deva.Pha: deva.high.Pha
        deva.Ya: deva.high.Ya
        deva.La: deva.high.La

typeface:
  name: Murty Hindi
  characters:
    studied:
        deva.Ddha: deva.high.Ddha
        deva.Nna: deva.high.Nna
        deva.Ta: deva.high.Ta
        deva.Tha: deva.high.Tha
        deva.Da: deva.high.Da
        deva.Ya: deva.high.Ya
        deva.Sha: deva.high.Sha
        deva.Ssa: deva.high.Ssa
```

```
typeface:
  name: Nirmala UI
  characters:
    studied:
        deva.U: deva.high.U
        deva.Nga: deva.high.Nga
        deva.Cha: deva.high.Cha
        deva.Tta: deva.high.Tta
        deva.Ttha: deva.high.Ttha
        deva.Da: deva.high.Da
        deva.Dha: deva.high.Dha
        deva.La: deva.low.La.verticalmiddle

typeface:
  name: PT Sans
  characters:
    studied:
        latn.b: latn.sans.b
        latn.f: latn.sans.f
        latn.i: latn.sans.i
        latn.n: latn.sans.n
        latn.p: latn.sans.p
        latn.q: latn.sans.q
        latn.w: latn.sans.w
        latn.z: latn.sans.z
        cyrl.ghe: cyrl.sans.ghe
        cyrl.ie: cyrl.sans.ie
        cyrl.o: cyrl.sans.o
        cyrl.pe: cyrl.sans.pe
        cyrl.te: cyrl.sans.te
        cyrl.ha: cyrl.sans.ha
        cyrl.shcha: cyrl.sans.shcha
        cyrl.e: cyrl.sans.e
```

```
typeface:
  name: PT Serif
  characters:
    studied:
        latn.g: latn.serif.g
        latn.h: latn.serif.h
        latn.m: latn.serif.m
        latn.n: latn.serif.n
        latn.q: latn.serif.q.spurless
        latn.r: latn.serif.r
        latn.y: latn.serif.y
        latn.z: latn.serif.z
        cyrl.ve: cyrl.serif.ve
        cyrl.ze: cyrl.serif.ze.halfseriffed
        cyrl.el: cyrl.serif.el.halfseriffed
        cyrl.er: cyrl.serif.er
        cyrl.te: cyrl.serif.te
        cyrl.ha: cyrl.serif.ha
        cyrl.u: cyrl.serif.u
        cyrl.e: cyrl.serif.e.halfseriffed

typeface:
  name: Times New Roman
  characters:
    studied:
        latn.a: latn.serif.a
        latn.c: latn.serif.c
        latn.f: latn.serif.f
        latn.l: latn.serif.l
        latn.m: latn.serif.m
        latn.n: latn.serif.n
        latn.t: latn.serif.t
        latn.u: latn.serif.u
        cyrl.a: cyrl.serif.a
        cyrl.be: cyrl.serif.be
        cyrl.io: cyrl.serif.io
        cyrl.short-i: cyrl.serif.short-i
        cyrl.tse: cyrl.serif.tse
        cyrl.sha: cyrl.serif.sha
        cyrl.shcha: cyrl.serif.shcha
        cyrl.hard-sign: cyrl.serif.hard-sign
```

```
typeface:
  name: Verdana
  characters:
    studied:
        latn.d: latn.sans.d
        latn.g: latn.sans.g
        latn.h: latn.sans.h
        latn.o: latn.sans.o
        latn.s: latn.sans.s
        latn.u: latn.sans.u
        latn.w: latn.sans.w
        latn.x: latn.sans.x
        cyrl.a: cyrl.sans.a
        cyrl.ve: cyrl.sans.ve
        cyrl.i: cyrl.sans.i
        cyrl.ha: cyrl.sans.ha
        cyrl.che: cyrl.sans.che
        cyrl.hard-sign: cyrl.sans.hard-sign
        cyrl.yeru: cyrl.sans.yeru
        cyrl.soft-sign: cyrl.sans.soft-sign
```

Appendix 9: Python code

Python code for a class Model defining the basis of the model described in chapter 6 and a routine used to assess the model.

```
"""
Coherence model: classes used for modelling character similarity and coherence

This library implements four classes: Feature, Concept, Typeface and Model.
The main interfaces is the Model class which uses the other classes.
"""

import defcon
import itertools
import json
import logging
import math
import os
import numpy as np
import pandas as pd
from myaml import load_yamljson, save_yamljson
from collections import OrderedDict

logging.basicConfig(format='%(levelname)s: %(message)s', level=logging.DEBUG)


# ----------------------------------------------------------------------------
# General functions and helpers
# ----------------------------------------------------------------------------

def get_all_triplets(characters):
    """
    Get a sorted list of triplet combinations
    for all characters from input
    """

    triplets = []
    for triplet in list(itertools.combinations(characters, 3)):
        triplets.append(sorted(list(triplet)))

    return triplets


prefix_to_script = {
    "cyrl": "cyrillic",
    "deva": "devanagari",
    "latn": "latin",
}
```

```
script_to_prefix = {
    “cyrillic”: “cyrl”,
    “devanagari”: “deva”,
    “latin”: “latn”,
}


# ------------------------------------------------------------------------
# Model
# ------------------------------------------------------------------------

class Feature(object):
    “””
    Representation of a feature
    “””

    def __init__(self, name, **kwargs):
        “””
        Input: name – feature name
               id – unique id of a feature (there can be multiple features
                    of the same name in the same character concept)
               salience – salience score of the feature
               type – type of the feature: “part”, “role” (default)
                      this is used only for filtering
        “””
        super(Feature, self).__init__()
        if kwargs is None:
            kwargs = {}
        self.name = name
        self.id = kwargs.get(“id”, name)
        self.salience = kwargs.get(“salience”, 0.0)
        if self.salience < 0:
            self.salience = 0.0  # non-negative salience values only
        self.type = kwargs.get(“type”, “role”)

    def __str__(self):
        o = “<Feature ‘%s’ (%s, %s)>” % (self.name, self.type, self.salience)
        return o

    def __repr__(self):
        return self.__str__()
```

```
        def as_ordered_dict(self):
            """
            Input: None
            Output:  object represented as a OrderedDict
            """

            data = OrderedDict()
            for k in sorted(self.__dict__.keys()):
                v = self.__dict__[k]
                if v is not None:
                    data[k] = v
            return data


    class Concept(object):
        """
        Representation of a character concept
        """

        def __init__(self, name, features=[]):
            super(Concept, self).__init__()
            self.name = name
            self.features = features

        def __str__(self):
            fns = ["%s (%s)" % (f.name, f.salience) for f in self.features]
            return "<Concept '%s' with features: %s>" % (self.name, ", ".join(fns))

        def __repr__(self):
            return self.__str__()

        def as_ordered_dict(self):
            """
            Input: None
            Output: Concept object represented as a OrderedDict
            """

            data = OrderedDict()
            data["name"] = self.name
            data["features"] = [f.as_ordered_dict() for f in self.features]
            return data
```

```
class Typeface(object):
    """
    Object representing a typeface,
    its characters and corresponding character concepts,
    and other parameters
    """

    def __init__(self, path=None):
        """
        Create a Typeface object and load it's settings from a file

        Input:  path to YAML/JSON file with typeface specification
        Output: None
        """

        super(Typeface, self).__init__()
        data = load_yamljson(path)["typeface"]
        if data is None:
            data = {}
        self.name = data.get("name", None)
        self.font = data.get("font", None)
        self.ufo = data.get("ufo", None)
        self.fontsize = data.get("fontsize", 100.0)
        self.fontshift = data.get("fontshift", 0.0)
        self.concepts_path = data.get("concepts_path", None)
        self.characters = data["characters"].get("studied", OrderedDict())
        self.other_characters = data["characters"].get("other", OrderedDict())
        if self.name:
            self.token = self.name.lower().replace(" ", "-")
        else:
            self.token = None

    def __str__(self):
        output = []
        output.append("Typeface: %s" % self.name)
        output.append("  made of characters:")
        for cn, ccn in self.characters.items():
            output.append("  %s -> %s" % (cn, ccn))
        return "\n".join(output)

    def __repr__(self):
        return self.__str__()
```

```
    def as_ordered_dict(self):
        """
        Input: None
        Output: Typeface object represented as a OrderedDict
        """

        data = OrderedDict()
        data["name"] = self.name
        data["font"] = self.font
        data["ufo"] = self.ufo
        data["fontsize"] = self.fontsize
        data["fontshift"] = self.fontshift
        data["concepts_path"] = self.concepts_path
        data["characters"] = OrderedDict()
        data["characters"]["studied"] = self.characters
        data["characters"]["other"] = self.other_characters
        return data

    def save(self, path=None):
        """
        Save typeface data to a file. Overwrite existing file.

        Input:  path to file
        Output: None
        """

        if path is None:
            path = self.path
        data = OrderedDict()
        data["typeface"] = self.as_ordered_dict()
        save_yamljson(data, path)
```

```
class Model(object):
    """
    Represents coherence model for one or more typefaces
    and associated operations (calculations of similarity etc.)
    """

    def __init__(self, **kwargs):
        """
        Load the database of concepts and mappings for typefaces studied.

        Input:  typeface_paths - paths to YAML files with typeface mappings
                script - script used to filter concepts from typeface mappings
                method - method used for concepts' comparisons, currently
                         three methods are supported:
                         "generalized", "contextual", and "random"
                load_parts - if part-features should be loaded into the model
                load_roles - if role-features should be loaded into the model
                ignore_features - list of features to ignore
        Output: None
        """

        super(Model, self).__init__()

        # load typefaces
        self.typefaces = OrderedDict()
        self.script = kwargs.get("script", "latin")
        concepts_paths = []
        loaded_concepts = []
        if kwargs is None:
            kwargs = {}
        self.method = kwargs.get("method", "contextual")
        self.load_parts = kwargs.get("load_parts", True)
        self.load_roles = kwargs.get("load_roles", True)
        self.ignore_features = kwargs.get("ignore_features", [])
        saliences = kwargs.get("saliences", pd.Series())
        typeface_paths = kwargs.get("typeface_paths", [])
        if isinstance(typeface_paths, str):
            typeface_paths = [typeface_paths]
        for path in typeface_paths:
            t = Typeface(path=path)
            self.typefaces[t.token] = t
            cpath = os.path.join(os.path.dirname(path), t.concepts_path)
            concepts_paths.append(cpath)
```

```
    loaded_concepts += t.characters.values()
# filter required concepts by script
required_concepts = []
if self.script != “all”:
    for ccn in loaded_concepts:
        if ccn.startswith(script_to_prefix[self.script]):
            required_concepts.append(ccn)
else:
    required_concepts = loaded_concepts
# load required concepts from concept UFOs
# sets the features as well
self.concepts = OrderedDict()
f_index = []
for concepts_path in set(concepts_paths):
    f_index += self.load_from_ufo(concepts_path, required_concepts)
# report characters using non-existing concepts
for tn in self.typefaces:
    for cn, ccn in self.typefaces[tn].characters.items():
        if (ccn in required_concepts) and (ccn not in self.concepts):
            ccn_new = “.”.join(ccn.split(“.”)[:-1])
            logging.info(
                “””
                Concept ‘%s’ (used in ‘%s’ in ‘%s’)
                not in the database. Using default ‘%s’ instead.
                ””” % (ccn, cn, tn, ccn_new))
            self.typefaces[tn].characters[cn] = ccn_new
# set similarity function
self.set_comparison_method(self.method)
# build matrices
# set concepts matrix (local saliences)
c_index = sorted(self.concepts.keys())
f_index = sorted(f_index)
self.concepts_matrix = pd.DataFrame(0.0,
                                    columns=c_index,
                                    index=f_index)
for cn, cc in self.concepts.items():
    for f in cc.features:
        # get local saliences from features used in concepts
        # there could be multiple copies of the same feature
        self.concepts_matrix[cn][f.name] += f.salience
# set global saliences
if saliences.empty:
    self.reset_global_saliences()
else:
    self.set_global_saliences(saliences)
```

```
    def __str__(self):
        return "Model for typefaces: %s" % ", ".join(self.typefaces)

    def __repr__(self):
        return self.__str__()

    def reset_global_saliences(self):
        """
        Randomly initialize global saliences to values in [0.0, 1.0)
        """

        rvals = np.random.random_sample(len(self.concepts_matrix.index))
        self.gsaliences = pd.Series(rvals, index=self.concepts_matrix.index)

    def set_global_saliences(self, s):
        """
        Set global saliences from a Series
        """

        # ensure integrity
        self.gsaliences = pd.Series(0.0, index=self.concepts_matrix.index)
        shared_ix = list(set(self.gsaliences.index).intersection(s.index))
        self.gsaliences[shared_ix] = s[shared_ix]

        for f in self.gsaliences.index:
            if f not in s.index:
                logging.info("No salience for feature '%s', set to 0.0." % f)

    def load_global_saliences(self, path):
        """
        Load global saliences from a CSV file
        """

        saliences = pd.read_csv(path, index_col=[0], header=[0]).iloc[:, 0]
        self.set_global_saliences(saliences.astype(float))

    def save_global_saliences(self, path):
        """
        Save global saliences to a CSV file
        """
        self.gsaliences.to_csv(path)
```

```
    # ------------------------------------------------------------------------
    # Loading concepts and features
    # ------------------------------------------------------------------------

    def __get_width_thresholds(self, font):
        """
        Calculate width thresholds based on statistics for the representative
        characters. This is used to assign roles r_narrow and r_wide.
        """

        width_thresholds = {}

        # Cyrillic
        sc = "cyrl"
        widths = pd.Series([g.width for g in font if g.name.startswith(sc)])
        low = 0  # do not use the r_narrow role for this script
        high = widths.quantile(.85)
        width_thresholds[sc] = (low, high)
        # Devanagari
        sc = "deva"
        widths = pd.Series([g.width for g in font if g.name.startswith(sc)])
        low = 0  # do not use the r_narrow role for this script
        high = widths.quantile(.75)
        width_thresholds[sc] = (low, high)
        # Latin
        sc = "latn"
        widths = pd.Series([g.width for g in font if g.name.startswith(sc)])
        low = widths.quantile(.2)
        high = widths.quantile(.85)
        width_thresholds[sc] = (low, high)

        return width_thresholds

    def load_from_ufo(self, path, required_concepts):
        """
        Load character concepts (features) from the UFO files.

        Input:  path - path to the UFOs (.ufo, .real.ufo)
        Output: index of unique features loaded
        """

        f_index = []
        # load concepts font
```

```
            concept_font = defcon.Font(path)
            x_height = concept_font.info.xHeight
            # load font with representative characters
            real_font = defcon.Font(path.replace(".ufo", ".real.ufo"))
            # get width thresholds
            width_thresholds = self.__get_width_thresholds(real_font)
            # load only concepts that are needed
            # concepts are stored as glyphs in the UFO
            for glyph in concept_font:
                gn = glyph.name
                if glyph.name in required_concepts:
                    # get features for the concept
                    features = self.__get_features(glyph, x_height)
                    ar = self.__get_autoroles(real_font, gn, width_thresholds)
                    features.update(ar)
                    # ignore features from the ignore list
                    features_ = {}
                    for f in features.values():
                        if f.name not in self.ignore_features:
                            features_[f.id] = f
                    # load only features groups requested (parts, roles)
                    ffeatures = []
                    for f in features_.values():
                        if f.type == "part" and self.load_parts:
                            ffeatures.append(f)
                        elif f.type == "role" and self.load_roles:
                            ffeatures.append(f)
                    # create a new concept
                    self.concepts[gn] = Concept(gn, ffeatures)
                    f_index += [f.name for f in ffeatures]

        # return a list of feature names
        return list(set(f_index))

    def __get_features(self, glyph, x_height=0):
        """
        Return a list of features that are represented as contour segments
        (stroke parts) or anchors (non-stroke parts, roles) in the UFO.
        Measure contour lengths to get stroke salience.

        Input:  glyph - Defcon Glyph object
                x_height - x-height from the concept font
        Return: a dict of features (unique ids as keys)
        """
```

```
# settings for Devanagari
headline_scope = (580 - 30, 580 + 30)

def bezier_point(t, start, c1, c2, end):
    return start * (1.0 - t)**3 + 3.0 * c1 * (1.0 - t)**2 * t + \
        3.0 * c2 * (1.0 - t) * t**2 + end * t**3

def bezier_length(pt1, cpt1, cpt2, pt2):
    length = 0
    prev_x, prev_y = pt1.x, pt1.y

    steps = 300  # give a good approximation
    for i in range(0, steps):
        t = i / float(steps)
        x = bezier_point(t, pt1.x, cpt1.x, cpt2.x, pt2.x)
        y = bezier_point(t, pt1.y, cpt1.y, cpt2.y, pt2.y)
        length += math.sqrt((x - prev_x)**2 + (y - prev_y)**2)
        prev_x, prev_y = x, y
    return length

features = {}

# stroke features are represented as contour segments
# using a counter (i) to make unique ids
i = 0
for c in glyph:
    if c.open:
        prev = c.segments[0][-1]
    else:
        prev = c.segments[-1][-1]
    for s in c.segments:
        pt = s[-1]
        if prev != pt:
            if len(s) > 1:
                length = bezier_length(prev, *s)
                # make relative
                length /= x_height
                name = "sweep"
            else:
                dx = prev.x - pt.x
                dy = prev.y - pt.y
                length = math.sqrt(dx**2 + dy**2)
                # make relative
```

```
                    length /= x_height
                    # add directional role
                    if dx and dy:
                        name = “diagonal”
                    elif dy:
                        name = “vertical”
                    # add headline role
                    # is within the bounds for devanagari headline
                    # script deduced from the glyph name
                    elif dx and glyph.name.startswith(“deva.”) \
                            and (pt.y > headline_scope[0]) \
                            and (pt.y < headline_scope[1]):
                        name = “headline”
                    elif dx:
                        name = “horizontal”
                ix = “%s_%s” % (name, i)  # unique id
                f = Feature(name, id=ix, salience=length, type=”part”)
                features[f.id] = f
                # bump up the stroke counter and length counter
                i += 1
            prev = pt
    # features represented as anchors
    for a in glyph.anchors:
        an = a.name.split(“_”)
        if a.name.startswith(“r_”):
            # roles
            # names of their anchors are unique -> id
            name = “_”.join(an[0:2])
            f = Feature(name, id=a.name, salience=1.0)
            features[f.id] = f
        else:
            # non-stroke parts
            # names of their anchors are unique -> id
            name = an[0]
            f = Feature(name, id=a.name, salience=1.0, type=”part”)
            features[f.id] = f

    return features
```

```
def __get_autoroles(self, real_font, concept_name, width_thresholds={}):
        """
        Analyze the representative glyph to discover automated roles:
        - r_ascender, r_descender: whether there are extenders. The salience is
                                   a proportion of the "extender zone" occupied
                                   by this extender (i.e. ascender, descender)
        - r_narrow, r_wide: 1.0 if narrow/wide compared to width thresholds

        This has to work on the "real" font and contours of the representative
        characters to get the measurements right.

        Input:  real_font - Defcon Font object
                concept_name - name of the concept
                width_thresholds -  dict of upper and lower width thresholds
                                    to assign r_narrow and r_wide roles
        Output: a dict of features (unique ids as keys)
        """

        # get basic font info
        x_height = real_font.info.xHeight
        overshoot_height = 0.1 * x_height  # 10% of x-height
        ascender = real_font.info.ascender
        descender = real_font.info.descender
        real_glyph = real_font[concept_name]
        roles = {}

        # get width thresholds to assign roles r_narrow and r_wide
        # fallback values are set to extremes (avoids use of the roles)
        width_low, width_high = -100000, 100000
        for prefix in width_thresholds:
            if real_glyph.name.startswith(prefix):
                width_low, width_high = width_thresholds[prefix]
                break

        if not real_glyph.name.startswith("deva."):
            # r_ascender and r_descender
            # Latin and Cyrillic only
            # find height of ascender & descender
            # have to work with real glyph contour
            ascender_min_y, ascender_max_y = None, x_height
            descender_min_y, descender_max_y = 0, None
            for c in real_glyph:
                c_y = [s[-1].y for s in c.segments]
```

```
            min_y = min(c_y)
            max_y = max(c_y)
            if max_y > (x_height + overshoot_height):
                if not ascender_min_y:
                    ascender_min_y = ascender
                ascender_min_y = min(ascender_min_y, *c_y)
                ascender_max_y = max(ascender_max_y, *c_y)
            if min_y < -overshoot_height:
                if not descender_max_y:
                    descender_max_y = descender
                descender_min_y = min(descender_min_y, *c_y)
                descender_max_y = max(descender_max_y, *c_y)
        if ascender_min_y:
            ascender_min_y = max(x_height, ascender_min_y)
        else:
            ascender_min_y = x_height
        if descender_max_y:
            descender_max_y = min(0, descender_max_y)
        else:
            descender_max_y = 0
        zone = ascender - x_height
        a_proportion = (ascender_max_y - ascender_min_y) / zone
        f = Feature("r_ascender", salience=a_proportion)
        roles[f.id] = f
        d_proportion = abs((descender_min_y - descender_max_y) / descender)
        f = Feature("r_descender", salience=d_proportion)
        roles[f.id] = f
    # r_narrow
    if real_glyph.width <= width_low:
        f = Feature("r_narrow", salience=1.0)
        roles[f.id] = f
    # r_wide
    if real_glyph.width >= width_high:
        f = Feature("r_wide", salience=1.0)
        roles[f.id] = f

    return roles
```

```
    # ------------------------------------------------------------------------
    # Similarity and predictions
    # ------------------------------------------------------------------------

    def set_comparison_method(self, method):
        """
        Set a method for making comparisons between concepts.

        The compare_concepts(a, b, c) method should return
        a feature vector that (when combined with global saliences)
        defines the response frequency for a particular pair (a, b)
        in the context of a third character (c).
        """

        self.compare_concepts = eval("self.compare_concepts__" + method)

    def compare_concepts__random(self, a, b, c):
        """
        Return a zero vector (uncomparable concepts).
        This is used to establish a baseline.
        Triggers triplet response frequencies: (0.3333, 0.3333, 0.3333)
        """

        return np.zeros(a.shape)

    def compare_concepts__generalized(self, a, b, c):
        """
        Generalized similarity vector.
        Ignores the third character.
        """

        with np.errstate(divide="ignore", invalid="ignore"):
            return np.where(a * b != 0, 2 * np.minimum(a, b) / (a + b), 0.0)

    def compare_concepts__contextual(self, a, b, c):
        """
        Contextual similarity vector where the third characters' (c)
        distinct features contribute towards the similarity of (a, b)

        The judgement is made of two steps:
        1. consider similarity between 'a' and 'b'
            a. get shared salience of features that are both in 'a' and 'b'
        2. consider differences between 'c' and 'a' and between 'c' and 'b':
            a. get features from 'c' that are neither in 'a' nor in 'b'
            b. get features shared between 'a' and 'b' that are not in 'c'
        """
```

```
        # saliences for features in A and B
        sharedAB = 2 * np.minimum(a, b)
        # saliences for features in A and B that are not in C
        uniqueAB = sharedAB * (c == 0)
        # saliences for features in C that are not in A or B
        uniqueC = c * (a == 0) * (b == 0)
        combination = sharedAB + uniqueAB + uniqueC
        total = a + b + c
        with np.errstate(divide="ignore", invalid="ignore"):
            return np.where(total != 0, combination / total, 0.0)

    def get_triplet_prediction(self, typeface_name, triplet_names,
                               index=[0, 1, 2]):
        """
        Get a triplet response frequencies for particular typeface
        and character triplet in the OOO-test

        Input:  typeface_name - typeface name
                triplet_names - triplet of character names
                index - index of the resulting Series
        Output: response - Series with corresponding response frequencies
        """

        assert typeface_name in self.typefaces
        for cn in triplet_names:
            assert cn in self.typefaces[typeface_name].characters

        # get concept vectors for the characters in this typeface
        concepts = []
        for cn in triplet_names:
            ccn = self.typefaces[typeface_name].characters[cn]
            concepts.append(np.array(self.concepts_matrix[ccn]))
        a, b, c = concepts
        # get comparison vectors
        A = self.compare_concepts(b, c, a)
        B = self.compare_concepts(a, c, b)
        C = self.compare_concepts(a, b, c)
        # combine comparison vectors with global-saliences vector

        A = A.dot(self.gsaliences)
        B = B.dot(self.gsaliences)
        C = C.dot(self.gsaliences)
        assert isinstance(A, float)
```

```
        assert isinstance(B, float)
        assert isinstance(C, float)
        total = A + B + C
        if total == 0:
            # cannot decide about this triplet
            A = B = C = 0.3333
        else:
            # normalize frequencies
            A /= total
            B /= total
            C /= total
        return pd.Series([A, B, C], index=index)

    def make_similarity_matrix(self, typeface_name, characters_names=None):
        """
        Make a similarity matrix for all character combinations for
        characters from a specified typeface

        Input:  typeface_name - typeface name
                characters_names - a list of character names
        Output: similarity_matrix - DataFrame
        """

        assert typeface_name in self.typefaces
        if characters_names is not None:
            for cn in characters_names:
                assert cn in self.typefaces[typeface_name].characters
        else:
            characters_names = []
            for cn in self.typefaces[typeface_name].characters:
                if cn.startswith(script_to_prefix[self.script]):
                    characters_names.append(cn)

        characters_index = sorted(list(set(characters_names)))
        similarity_matrix = pd.DataFrame(0.0, index=characters_index,
                                         columns=characters_index)
        triplets = get_all_triplets(characters_index)

        for t in triplets:
            cn1, cn2, cn3 = t
            r1, r2, r3 = self.get_triplet_prediction(typeface_name, t)
            similarity_matrix[cn1][cn2] += r3
            similarity_matrix[cn1][cn3] += r2
```

```
            similarity_matrix[cn2][cn3] += r1
            # and symmetrically
            similarity_matrix[cn2][cn1] += r3
            similarity_matrix[cn3][cn1] += r2
            similarity_matrix[cn3][cn2] += r1
        # clear diagonal
        for cn in characters_index:
            similarity_matrix[cn][cn] = np.nan
        # get mean
        similarity_matrix /= (len(characters_index) - 1)
        return similarity_matrix

    def make_predictions(self, columns=None, index=[0, 1, 2]):
        """
        Make a triplet response matrix for all triplet combinations provided.

        Input:  columns - a list of tuples,
                          columns have to be in the following format:
                          (typeface_name, triplet of character names)
                index - index of the resulting matrix
        Output: response_matrix - DataFrame with triplet response frequencies
                                  with identical columns
        """

        assert columns is not None
        responses = pd.DataFrame(columns=columns, index=index)
        for typeface, triplet in columns:
            p = self.get_triplet_prediction(typeface, triplet, index=index)
            responses[typeface, triplet] = p
        return responses

    # ------------------------------------------------------------------------
    # Model training
    # ------------------------------------------------------------------------

    def train(self, observed, epochs=1000, step=0.05, batchsize=40, loss=None):
        """
        Fit global feature saliences to the observed data to minimize
        the loss function and updates the model's global saliences.

        Input:  observed - triplet response matrix with observed frequencies,
                           columns have to be in the following format:
                           (typeface_name, triplet of character names)
```

```
        epochs – number of iterations
        step – change step (aka alpha)
        batchsize – mini-batch size
        loss – loss function used
    Output: loss_history – list with the progressively changing loss value
    """

    def next_batch(observed, batchsize):
        for i in np.arange(0, observed.shape[0], batchsize):
            # yield a tuple of the current batched data and labels
            observed_b = observed[i:i + batchsize]
            compared_b = compared[i:i + batchsize]
            yield (compared_b, observed_b)

    def compare(typeface_name, triplet_names):
        """
        In order to make a prediction of similarity judgements,
        get a comparison vector for each character in the triplet trial.
        Input:  typeface_name – a string with typeface name from the model
                triplet_names – a tuple with three character names
        Output: response_vectors – a tuple with three comparison vectors
        """

        # get concept names for the characters in this typeface
        # get local feature saliences for each concept
        concept_saliences = []
        assert typeface_name in self.typefaces
        for cn in triplet_names:
            assert cn in self.typefaces[typeface_name].characters
            ccn = self.typefaces[typeface_name].characters[cn]
            concept_saliences.append(np.array(self.concepts_matrix[ccn]))
        a, b, c = concept_saliences

        # get comparison vectors for each character
        A = self.compare_concepts(b, c, a)
        B = self.compare_concepts(a, c, b)
        C = self.compare_concepts(a, b, c)

        # normalize the vectors
        total = A + B + C
        with np.errstate(divide="ignore", invalid="ignore"):
            A = np.where(total != 0, A / total, 0)
            B = np.where(total != 0, B / total, 0)
            C = np.where(total != 0, C / total, 0)
```

```
    return np.array([A, B, C])

gsaliences = self.gsaliences.values
compared = []
for tn, triplet in observed.columns:
    compared.append(compare(tn, triplet))
compared = np.array(compared)
# the dot product between comparisons and global saliences
# causes predictions to transpose, this is a compensation
observed = observed.values.T
# SGD
losshistory = []
for epoch in np.arange(0, epochs):
    lo = 1.0
    for batch_x, observed_b in next_batch(observed, batchsize):
        predicted_b = batch_x.dot(gsaliences)
        error = predicted_b - observed_b
        lo = loss(observed_b, predicted_b)
        batch_x = batch_x.reshape(-1, batch_x.shape[-1])
        error = error.reshape(-1,)
        momentum = batch_x.T.dot(error) / error.shape[0]
        gsaliences += -step * momentum
        # saliences have to be positive
        gsaliences = np.maximum(0, gsaliences)
    losshistory.append(lo)

# update global saliences
self.gsaliences = pd.Series(gsaliences,
                            index=self.gsaliences.index)

return losshistory
```

```
"""
Model training routine (CLI script)
"""

import numpy as np
import pandas as pd
import os
import sys

from chardict import all_typefaces, all_script_typefaces
from model import Model
from sklearn.metrics import cohen_kappa_score


# ----------------------------------------------------------------------------
# directory used to save results

RESDIR = "results"


# ----------------------------------------------------------------------------
# Loss function used


def mean_squared(observed, predicted):
    """
    Loss as a mean of squared differences
    between values from the two DataFrames.
    """

    return ((predicted - observed)**2).mean()


# ----------------------------------------------------------------------------
# Evaluate model


def r_squared(observed, predicted):
    """
    Calculate coefficient of determination (R^2)
    for observed and predicted values.
    """
```

```
    observed = observed.values
    predicted = predicted.values
    SSres = ((observed - predicted)**2).sum()
    emean = observed.mean()
    SStot = ((observed - emean)**2).sum()
    r2 = 1 - SSres / SStot
    return r2


def evaluate(observed_f, predicted_f, loss=mean_squared):
    """
    Compare predicted to observed frequencies and compile a report.

    Input:  observed_f  - DataFrame with observed frequencies
            predicted_f - DataFrame with predicted frequencies
                          (same shape and columns as observed)
    Output: report  - DataFrame
            overall - dict with overall evaluation metrics
    """

    observed_f = observed_f.fillna(0)
    predicted_f = predicted_f.fillna(0)

    report = pd.DataFrame(columns=observed_f.columns,
                          index=["Loss", "Accuracy", "OOOO", "Kappa"])
    for typeface, triplet in observed_f.columns:
        # compare the odd one out
        ex_ooo = observed_f[typeface, triplet].idxmax()
        pr_ooo = predicted_f[typeface, triplet].idxmax()
        report[typeface, triplet]["OOOO"] = ex_ooo == pr_ooo

    # overall statistics
    count = report.loc["OOOO"].count()
    overall = {}
    # OOOO success rate in %
    overall["OOOO"] = report.loc["OOOO"].astype(int).sum() / count
    overall["OOOO"] = round(overall["OOOO"] * 100, 4)
    # Cohen's kappa
    kappa = cohen_kappa_score(observed_f.idxmax(), predicted_f.idxmax())
    overall["Kappa"] = round(kappa, 4)
    # calculate overall loss and R^2
    ll = loss(observed_f, predicted_f).mean()
```

```
    r2 = r_squared(observed_f, predicted_f)
    overall["Loss"] = round(ll, 4)
    overall["Accuracy"] = round(r2, 4)
    report.insert(0, "overall", pd.Series(overall))
    return report, overall


# ------------------------------------------------------------------------
# Train model
# ------------------------------------------------------------------------


def train_model(load_parts=True,
                load_roles=True,
                ignore_features=[],
                script="latin",
                typefaces=[
                ],
                group="sans",
                method="contextual",
                loss=mean_squared,
                iters=50):

    print("Fitting using:", method, "method")

    # metaparameters for the SGD
    step = 0.01
    epochs = 1000
    batchsize = 40

    # get observed values
    path_f = "../data/csv/frequencies/frequencies_compact.csv"
    all_frequencies = pd.read_csv(path_f, index_col=[0], header=[0, 1, 2])
    all_frequencies = all_frequencies.dropna(axis=1, how="all")
    # fix columns for easier work (convert triplet tokens (strings) to tuples)
    ix = [(a, b, tuple(eval(c))) for a, b, c in list(all_frequencies.columns)]
    all_frequencies.columns = pd.MultiIndex.from_tuples(ix)

    # filter observed values
    # depending on the script and typefaces
    if script != "all":
        # filter by script and typefaces
        df = all_frequencies[script]
```

```
        columns = list(df.columns.remove_unused_levels())
        columns_filtered = [(a, b) for a, b in columns if a in typefaces]
        columns_filtered = pd.MultiIndex.from_tuples(columns_filtered)
        df = df[columns_filtered]
    elif typefaces:
        # filter only by typefaces
        # simply remove script from the column index everywhere
        df = all_frequencies
        columns = [(a, b) for _, a, b in df.columns]
        df.columns = pd.MultiIndex.from_tuples(columns)
        # and then filter by typefaces
        columns_filtered = [(a, b) for a, b in columns if a in typefaces]
        columns_filtered = pd.MultiIndex.from_tuples(columns_filtered)
        df = df[columns_filtered]
    else:
        # no filtering
        # simply remove script from the column index everywhere
        df = all_frequencies
        columns = [(a, b) for _, a, b in df.columns]
        df.columns = pd.MultiIndex.from_tuples(columns)

    # prepare data for baselines
    majority = pd.DataFrame(0.0, columns=df.columns, index=df.index)
    random = pd.DataFrame(0.0, columns=df.columns, index=df.index)
    for col in df.columns:
        majority[col][df[col].idxmax()] = 1.0
        r = np.random.random(3)
        random[col] = r/r.sum()

    # set up the model
    paths_t = ["typeface-mappings/%s.yaml" % t for t in typefaces]
    m = Model(typeface_paths=paths_t,
              script=script,
              load_parts=load_parts,
              load_roles=load_roles,
              ignore_features=ignore_features
              )

    # training
    print("Training on: %s script and '%s' dataset (%d iterations)"
          % (script, group, iters))
    # set up a report to collect evaluations from all model trainings
    ix = ["Train count", "Eval count", "Features count", "Features STD",
          "Loss", "Accuracy", "OOOO", "Kappa"]
```

```
report = pd.DataFrame(columns=range(iters), index=ix)
for i in range(iters):
    print(“.”, end=””)

    # shuffle the columns and split data to training and evaluation sets
    df = df.reindex(np.random.permutation(df.columns), axis=1)
    split_at = round(.75 * len(df.columns))
    data_train = df[df.columns[:split_at]]
    de = df[df.columns[split_at:]]

    # get predictions
    if method == “random”:
        # use random baseline (equal responses)
        predictions = random[de.columns]
    elif method == “majority”:
        # use majority baseline (majority opinion)
        predictions = majority[de.columns]
    elif method == “oracle”:
        # use oracle baseline (exactly as observed)
        predictions = de
    else:
        # use model
        m.set_comparison_method(method)
        m.reset_global_saliences()
        m.train(observed=data_train,
                epochs=epochs,
                step=step,
                batchsize=batchsize,
                loss=loss)
        predictions = m.make_predictions(columns=de.columns)

    # evaluate model
    _, evaluation = evaluate(df[de.columns], predictions, loss=loss)
    evaluation[“Train count”] = len(df.columns)-len(de.columns)
    evaluation[“Eval count”] = len(de.columns)
    # save reports and saliences
    prefix = “%s_%s_%s” % (script, group, method)
    mdir = os.path.join(RESDIR, prefix)
    if not os.path.exists(mdir):
        os.makedirs(mdir)
    if method not in [“oracle”, “majority”, “random”]:
        saliences_path = os.path.join(mdir, str(i) + “_saliences.csv”)
        m.save_global_saliences(saliences_path)
```

```
            evaluation[“Features count”] = len(m.gsaliences)
            evaluation[“Features STD”] = m.gsaliences.std()
        else:
            evaluation[“Features count”] = “”
            evaluation[“Features STD”] = “”
        predictions.to_csv(os.path.join(mdir, str(i) + “_predictions.csv”))
        report[i] = pd.Series(evaluation)
    prefix = “%s_%s_%s__” % (script, group, method)
    df.to_csv(os.path.join(RESDIR, prefix + “data.csv”))
    report.to_csv(os.path.join(RESDIR, prefix + “evaluation.csv”))
    print()


# ------------------------------------------------
# Main: run training from CLI

# CLI arguments -> scripts to process
if len(sys.argv) > 1:
    scripts = [a.lower() for a in sys.argv[1:]]
else:
    scripts = [“all”]

loss = mean_squared

for script in scripts:
    if script == “all”:
        for method in [
                “generalized”,
                “contextual”,
                “random”,
                “oracle”,
                “majority”,
                ]:
            # parts only
            train_model(load_parts=True,
                        load_roles=False,
                        script=script,
                        typefaces=all_typefaces,
                        group=”all-parts”,
                        method=method,
                        loss=loss,
                        )
            # roles only
```

```
            train_model(load_parts=False,
                        load_roles=True,
                        script=script,
                        typefaces=all_typefaces,
                        group="all-roles",
                        method=method,
                        loss=loss,
                        )
            # parts & roles
            train_model(script=script,
                        typefaces=all_typefaces,
                        group="all",
                        method=method,
                        loss=loss,
                        )
    else:
        for group in list(all_script_typefaces[script]) + ["all"]:
            if group == "all":
                # use all typefaces for particular script
                typefaces = []
                for gn in all_script_typefaces[script]:
                    typefaces += list(all_script_typefaces[script][gn])
            else:
                # use only typefaces from particular group
                typefaces = all_script_typefaces[script][group]
            # parts & roles & contextual method
            train_model(script=script,
                        typefaces=typefaces,
                        group=group,
                        method="contextual",
                        loss=loss,
                        )
```

Appendix 10: similarity matrices compared

A subset of a similarity matrix published by Simpson et al. (2013) and a matrix generated using the model described in chapter 6.

	latn.a	latn.b	latn.c	latn.d	latn.e	latn.f	latn.g	latn.h	latn.i	latn.j	latn.k	latn.l
latn.a		2.13	2.50	2.57	3.40	1.06	3.30	1.57	1.16	1.13	1.13	1.10
latn.b	2.13		3.03	5.60	2.27	1.83	3.53	3.70	1.43	1.43	2.13	2.77
latn.c	2.50	3.03		3.57	4.43	1.24	2.47	1.30	1.10	1.13	1.30	1.17
latn.d	2.57	5.60	3.57		2.87	1.43	4.10	2.73	1.73	1.27	1.33	2.30
latn.e	3.40	2.27	4.43	2.87		1.33	2.37	1.23	1.20	1.17	1.20	1.07
latn.f	1.06	1.83	1.24	1.43	1.33		1.40	1.80	4.16	3.67	1.57	4.03
latn.g	3.30	3.53	2.47	4.10	2.37	1.40		1.57	1.53	2.33	1.17	1.30
latn.h	1.57	3.70	1.30	2.73	1.23	1.80	1.57		1.93	1.60	2.77	2.53
latn.i	1.16	1.43	1.10	1.73	1.20	4.16	1.53	1.93		5.17	1.90	6.13
latn.j	1.13	1.43	1.13	1.27	1.17	3.67	2.33	1.60	5.17		1.47	4.67
latn.k	1.13	2.13	1.30	1.33	1.20	1.57	1.17	2.77	1.90	1.47		2.80
latn.l	1.10	2.77	1.17	2.30	1.07	4.03	1.30	2.53	6.13	4.67	2.80	
latn.m	1.40	1.23	1.20	1.13	1.13	1.33	1.27	2.53	1.63	1.13	1.20	1.07
latn.n	1.63	1.77	2.00	1.93	1.73	1.20	1.70	5.53	1.50	1.30	1.53	1.37
latn.o	3.13	4.20	5.23	4.90	4.13	1.17	3.47	1.47	1.33	1.00	1.03	1.23
latn.p	2.03	5.07	2.60	5.10	2.40	1.63	4.50	2.47	1.60	1.57	1.53	1.80
latn.q	2.60	4.67	2.77	5.10	2.43	1.47	5.30	2.17	1.55	1.63	1.47	1.83
latn.r	1.43	1.27	1.40	1.33	1.27	3.80	1.27	2.37	2.70	2.30	2.00	3.26
latn.s	2.13	1.40	2.43	1.27	2.20	1.27	1.80	1.23	1.07	1.17	1.40	1.10
latn.t	1.07	1.63	1.40	1.37	1.20	4.80	1.13	1.97	3.90	3.80	2.13	4.50
latn.u	2.40	1.93	2.17	1.83	1.57	1.27	1.73	3.33	1.67	1.30	1.26	1.29
latn.v	1.23	1.40	1.53	1.55	1.23	1.23	1.20	1.40	1.67	1.37	1.97	1.37
latn.w	1.13	1.10	1.23	1.07	1.30	1.13	1.17	1.30	1.30	1.10	2.07	1.33
latn.x	1.10	1.07	1.23	1.23	1.30	1.28	1.20	1.23	1.40	1.26	3.55	1.27
latn.y	1.07	1.20	1.29	1.13	1.37	1.40	2.67	1.50	1.50	2.87	2.03	1.81
latn.z	1.37	1.13	1.47	1.23	1.33	1.30	1.23	1.20	1.13	1.23	1.50	1.10

TABLE 10.1: A subset of similarity matrix published by Simpson et al. (2013) with header captions adapted to the notation used in this thesis (continues on the next page).

latn.m	latn.n	latn.o	latn.p	latn.q	latn.r	latn.s	latn.t	latn.u	latn.v	latn.w	latn.x	latn.y	latn.z
1.40	1.63	3.13	2.03	2.60	1.43	2.13	1.07	2.40	1.23	1.13	1.10	1.07	1.37
1.23	1.77	4.20	5.07	4.67	1.27	1.40	1.63	1.93	1.40	1.10	1.07	1.20	1.13
1.20	2.00	5.23	2.60	2.77	1.40	2.43	1.40	2.17	1.53	1.23	1.23	1.29	1.47
1.13	1.93	4.90	5.10	5.10	1.33	1.27	1.37	1.83	1.55	1.07	1.23	1.13	1.23
1.13	1.73	4.13	2.40	2.43	1.27	2.20	1.20	1.57	1.23	1.30	1.30	1.37	1.33
1.33	1.20	1.17	1.63	1.47	3.80	1.27	4.80	1.27	1.23	1.13	1.28	1.40	1.30
1.27	1.70	3.47	4.50	5.30	1.27	1.80	1.13	1.73	1.20	1.17	1.20	2.67	1.23
2.53	5.53	1.47	2.47	2.17	2.37	1.23	1.97	3.33	1.40	1.30	1.23	1.50	1.20
1.63	1.50	1.33	1.60	1.55	2.70	1.07	3.90	1.67	1.67	1.30	1.40	1.50	1.13
1.13	1.30	1.00	1.57	1.63	2.30	1.17	3.80	1.30	1.37	1.10	1.26	2.87	1.23
1.20	1.53	1.03	1.53	1.47	2.00	1.40	2.13	1.26	1.97	2.07	3.55	2.03	1.50
1.07	1.37	1.23	1.80	1.83	3.26	1.10	4.50	1.29	1.37	1.33	1.27	1.81	1.10
	4.67	1.30	1.33	1.17	2.23	1.20	1.07	2.27	1.33	3.40	1.10	1.20	1.10
4.67		2.40	1.83	1.90	3.13	1.52	1.10	4.53	1.97	1.61	1.03	1.43	1.57
1.30	2.40		3.60	4.10	1.13	2.27	1.13	2.83	1.27	1.13	1.07	1.17	1.27
1.33	1.83	3.60		5.57	1.77	1.26	1.33	1.80	1.13	1.17	1.10	1.97	1.17
1.17	1.90	4.10	5.57		1.47	1.23	1.80	2.30	1.20	1.03	1.10	2.03	1.17
2.23	3.13	1.13	1.77	1.47		1.60	3.37	1.97	1.67	1.03	1.43	1.27	1.30
1.20	1.52	2.27	1.26	1.23	1.60		1.20	1.33	1.13	1.10	1.63	1.33	2.17
1.07	1.10	1.13	1.33	1.80	3.37	1.20		1.60	1.40	1.00	1.47	1.50	1.27
2.27	4.53	2.83	1.80	2.30	1.97	1.33	1.60		4.93	2.73	1.40	3.13	1.57
1.33	1.97	1.27	1.13	1.20	1.67	1.13	1.40	4.93		5.03	2.63	5.33	1.97
3.40	1.61	1.13	1.17	1.03	1.03	1.10	1.00	2.73	5.03		2.23	2.43	1.67
1.10	1.03	1.07	1.10	1.10	1.43	1.63	1.47	1.40	2.63	2.23		3.10	1.80
1.20	1.43	1.17	1.97	2.03	1.27	1.33	1.50	3.13	5.33	2.43	3.10		1.93
1.10	1.57	1.27	1.17	1.17	1.30	2.17	1.27	1.57	1.97	1.67	1.80	1.93	

	latn.a	latn.b	latn.c	latn.d	latn.e	latn.f	latn.g	latn.h	latn.i	latn.j	latn.k	latn.l
latn.a		0.43	0.3	0.43	0.6	0.22	0.46	0.38	0.13	0.16	0.35	0.14
latn.b	0.43		0.44	0.73	0.47	0.39	0.59	0.48	0.24	0.28	0.29	0.32
latn.c	0.3	0.44		0.44	0.53	0.26	0.45	0.35	0.13	0.18	0.13	0.16
latn.d	0.43	0.73	0.44		0.47	0.39	0.59	0.48	0.24	0.29	0.29	0.32
latn.e	0.6	0.47	0.53	0.47		0.15	0.48	0.2	0.06	0.1	0.28	0.08
latn.f	0.22	0.39	0.26	0.39	0.15		0.26	0.5	0.52	0.57	0.29	0.61
latn.g	0.46	0.59	0.45	0.59	0.48	0.26		0.31	0.17	0.33	0.16	0.19
latn.h	0.38	0.48	0.35	0.48	0.2	0.5	0.31		0.3	0.37	0.35	0.4
latn.i	0.13	0.24	0.13	0.24	0.06	0.52	0.17	0.3		0.65	0.34	0.69
latn.j	0.16	0.28	0.18	0.29	0.1	0.57	0.33	0.37	0.65		0.21	0.48
latn.k	0.35	0.29	0.13	0.29	0.28	0.29	0.16	0.35	0.34	0.21		0.43
latn.l	0.14	0.32	0.16	0.32	0.08	0.61	0.19	0.4	0.69	0.48	0.43	
latn.m	0.37	0.34	0.37	0.34	0.23	0.31	0.32	0.52	0.18	0.24	0.2	0.23
latn.n	0.44	0.38	0.41	0.38	0.23	0.41	0.37	0.65	0.24	0.31	0.26	0.3
latn.o	0.44	0.55	0.64	0.55	0.64	0.19	0.57	0.25	0.12	0.13	0.09	0.14
latn.p	0.44	0.6	0.44	0.6	0.47	0.26	0.74	0.33	0.16	0.32	0.17	0.2
latn.q	0.44	0.6	0.44	0.6	0.47	0.26	0.74	0.33	0.16	0.32	0.17	0.2
latn.r	0.26	0.3	0.31	0.3	0.17	0.65	0.31	0.4	0.48	0.52	0.21	0.55
latn.s	0.4	0.17	0.36	0.17	0.46	0.38	0.18	0.17	0.26	0.29	0.34	0.29
latn.t	0.21	0.38	0.25	0.38	0.14	0.73	0.25	0.49	0.54	0.57	0.29	0.63
latn.u	0.44	0.38	0.41	0.38	0.23	0.41	0.37	0.65	0.24	0.31	0.26	0.3
latn.v	0.13	0.13	0.36	0.13	0.2	0.15	0.14	0.2	0.27	0.1	0.46	0.32
latn.w	0.1	0.1	0.28	0.1	0.16	0.12	0.11	0.16	0.23	0.08	0.35	0.27
latn.x	0.33	0.08	0.26	0.08	0.4	0.1	0.09	0.14	0.21	0.06	0.63	0.25
latn.y	0.17	0.18	0.42	0.18	0.24	0.25	0.34	0.29	0.1	0.32	0.28	0.13
latn.z	0.34	0.09	0.27	0.09	0.42	0.11	0.1	0.14	0.19	0.07	0.65	0.22

TABLE 10.2: Similarity matrix generated using the model described in chapter 6 (continues on the next page).

latn.m	latn.n	latn.o	latn.p	latn.q	latn.r	latn.s	latn.t	latn.u	latn.v	latn.w	latn.x	latn.y	latn.z
0.37	0.44	0.44	0.44	0.44	0.26	0.4	0.21	0.44	0.13	0.1	0.33	0.17	0.34
0.34	0.38	0.55	0.6	0.6	0.3	0.17	0.38	0.38	0.13	0.1	0.08	0.18	0.09
0.37	0.41	0.64	0.44	0.44	0.31	0.36	0.25	0.41	0.36	0.28	0.26	0.42	0.27
0.34	0.38	0.55	0.6	0.6	0.3	0.17	0.38	0.38	0.13	0.1	0.08	0.18	0.09
0.23	0.23	0.64	0.47	0.47	0.17	0.46	0.14	0.23	0.2	0.16	0.4	0.24	0.42
0.31	0.41	0.19	0.26	0.26	0.65	0.38	0.73	0.41	0.15	0.12	0.1	0.25	0.11
0.32	0.37	0.57	0.74	0.74	0.31	0.18	0.25	0.37	0.14	0.11	0.09	0.34	0.1
0.52	0.65	0.25	0.33	0.33	0.4	0.17	0.49	0.65	0.2	0.16	0.14	0.29	0.14
0.18	0.24	0.12	0.16	0.16	0.48	0.26	0.54	0.24	0.27	0.23	0.21	0.1	0.19
0.24	0.31	0.13	0.32	0.32	0.52	0.29	0.57	0.31	0.1	0.08	0.06	0.32	0.07
0.2	0.26	0.09	0.17	0.17	0.21	0.34	0.29	0.26	0.46	0.35	0.63	0.28	0.65
0.23	0.3	0.14	0.2	0.2	0.55	0.29	0.63	0.3	0.32	0.27	0.25	0.13	0.22
	0.58	0.28	0.35	0.35	0.36	0.2	0.3	0.58	0.19	0.45	0.14	0.26	0.14
0.58		0.3	0.39	0.39	0.47	0.21	0.39	0.73	0.25	0.2	0.17	0.34	0.18
0.28	0.3		0.55	0.55	0.22	0.29	0.17	0.3	0.31	0.25	0.23	0.31	0.2
0.35	0.39	0.55		0.74	0.3	0.18	0.25	0.39	0.14	0.11	0.09	0.35	0.09
0.35	0.39	0.55	0.74		0.3	0.18	0.25	0.39	0.14	0.11	0.09	0.35	0.09
0.36	0.47	0.22	0.3	0.3		0.42	0.66	0.47	0.2	0.16	0.13	0.29	0.14
0.2	0.21	0.29	0.18	0.18	0.42		0.37	0.21	0.27	0.21	0.48	0.32	0.5
0.3	0.39	0.17	0.25	0.25	0.66	0.37		0.39	0.16	0.12	0.1	0.24	0.12
0.58	0.73	0.3	0.39	0.39	0.47	0.21	0.39		0.25	0.2	0.17	0.34	0.18
0.19	0.25	0.31	0.14	0.14	0.2	0.27	0.16	0.25		0.72	0.71	0.58	0.66
0.45	0.2	0.25	0.11	0.11	0.16	0.21	0.12	0.2	0.72		0.57	0.44	0.52
0.14	0.17	0.23	0.09	0.09	0.13	0.48	0.1	0.17	0.71	0.57		0.45	0.85
0.26	0.34	0.31	0.35	0.35	0.29	0.32	0.24	0.34	0.58	0.44	0.45		0.43
0.14	0.18	0.2	0.09	0.09	0.14	0.5	0.12	0.18	0.66	0.52	0.85	0.43	