



COMPARISON OF GURUMUKHI AND LATIN SCRIPTS IN HANDWRITING IDENTIFICATION: A FORENSIC PERSPECTIVE

Amanpreet KAUR¹, R. K. GARG², Meenakshi MAHAJAN³

¹ *Department of Forensic Science and Chemistry, RIMT University, Mandi Gobindgarh, Punjab, India*

² *Department of Forensic Science, Punjabi University, Patiala, Punjab, India*

³ *Regional Forensic Science Laboratory, Dharamshala, HP, India*

Abstract

The present study has been carried out to study the possibility of handwriting comparison in cases where document under question is inscribed in different language and the available standards are in other language. The aspect is prevalent in the field of languages since people around the globe are being attracted to learn new languages, but their native language always has some impact on the next learned languages. Punjabi is the first language taught in the schools of Punjab and other languages afterward. Punjabi is the mother tongue of Punjab area and in contemporary times most of the population of Punjab leaning towards learning English language. For this study, 1200 samples in Punjabi and English from 100 individuals who studied English as well as Punjabi and using these languages on a daily basis were collected and analyzed. Class characteristics like alignment, slant, size, the arrangement of a piece of handwriting, connections, fluency line quality of handwriting along with the characters of letter form and formation were considered and the results were further analyzed statistically using Chi-square test. The results of the present study revealed the handwriting written in distinct scripts is similar and comparable with each other. Additionally, the study will help correlate a person's handwriting with ethnicity, nationality, and region based on the characters present in the handwriting samples.

Keywords

Language; Handwriting; Script; Questioned documents; Punjabi.

Received 23 November 2020; accepted 13 July 2021

Introduction

Linguistic diversity is a well-versed fact among the world's different populations. Over a hundred languages and dialects are used by Indian people and 22 out of those languages/scripts have been recognized by the Indian Constitution. In addition to this, there are numerous other dialects and scripts used by people of India for writing as well as for speaking purposes (Saini, Kaur, 2018). Apart from English, people also use their native language for their official and routine work. In Punjab, the Punjabi language is used as an

official language, although some of the official communication is done in English also, and similarly other states of India follow the same pattern. Very few references related to the present study are available in the literature. Latin and Cyrillic alphabets were compared to find out the possibility of similarities or dissimilarities between the alphabets of two scripts based on handwriting comparison principles (Ledić, Rosandić, 2018). In another study, English handwriting of individuals belonging to three states, Punjab, Jammu and Kashmir, and Andhra Pradesh respectively was analyzed on the basis of various class characteristics like

letter form, punctuation marks, the spacing between letters, etc. and showed that the regional language significantly influences the English language (Saini, Kaur, 2018).

In the present study along with these class characteristics peculiar character, letter formation is included which is dependent upon the abstract representation of the letter known as grapheme as reported by Ellis (1979). Punjabi is originated from the Indo Aryan group of languages (Masica, 1991) and modern Punjabi writings follow Gurumukhi script descendant of Brahmi script (Saxena, Singh, 1992; Jasuja, Komal, Singh, 1996). Gurumukhi script was acquired from the Sharda script and shaped up by Guru Angad Dev Ji in the 16th century (Gill, Gleason, 1963). On the other hand, English is an omniglotus language with 1.2 billion speakers all over the world. It is a West Germanic language related to Scots, Dutch, Frisian, and German that developed its exploration amid the medieval and early modern periods from England to Wales, Scotland, and other parts of the British peninsula, as well as Ireland. Consequently, during the 17th century English disseminated to other parts of the world through trade and colonization and new variants appeared. In addition to the aforementioned researches, other researchers such as Turner Sidhu and Love (2008) carried out a study in the United Kingdom to determine similarity among class characteristics of first and second-generation Punjabis. Further statistical studies were accomplished by Cheng et al. (2005) for the assurance of nationality and ethnic origin of writers from their English writings. Turnbull, Jones and Allen (2010) analyzed the English handwriting of the English and Polish people to ascertain the effect of language on handwriting.

The literature survey revealed that there is very little research available on analyzing and comparing different scripts, especially concerning the Indian

context, where a lot of population variation occurs and distinguished languages are spoken in each state. This depicts that there is no definite conclusion that has been derived by different workers regarding this particular concept. The present study is focused on the comparison of class and peculiar characteristics of two different languages Punjabi and English following two different scripts Gurumukhi and Latin respectively, commonly used by the inhabitants of Punjab and establishes the possibility of inter-script comparison of handwriting samples.

Materials and method

In totality 100 individuals partake in the present study, comprising of 50 males and 50 females. All the writers were skilled writers of the age group 18–50. The chosen group of individuals has acquired their primary education in Punjabi and at later stages learned English. Along with developed handwriting skills, all the participants knew to write in both languages. This study was done in two parts, in the first stage the interrelation of the two languages was established on the basis of class characteristics and later on the basis of letter and stroke formation. Figure 1a depicts the copybook form of Punjabi letters, which is native to the individuals as a primary learned language, while Figure 1b presents the English letters in both uppercase and lowercase form. Before the acquisition of handwriting samples written consent from the subjects had been obtained. No prior training for writing the samples was provided to the subjects before the collection of samples. To overcome the problems and errors in the results due to natural variation in collected samples; six samples in each script were collected on a plain white sheet of A4 size using a blue ball-point pen and fountain pen from each subject. Subjects

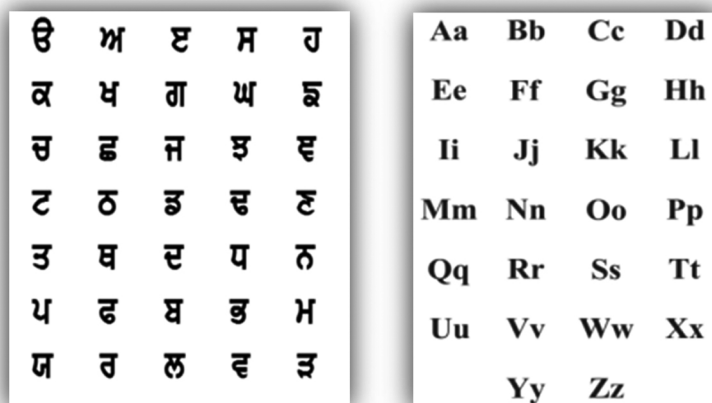


Fig. 1a, 1b. Gurumukhi and Latin scripts (16, 18).

were asked to write the selected paragraphs consisting of almost all letters and diacritics marks in both the languages while sitting in a comfortable writing position. Samples were collected through dictation and copying methods. Two handwriting samples (one with a ballpoint pen and one with a fountain pen) were dictated to the subjects along with that rest of the four handwriting samples were copied by the subjects from printed test paragraphs. Hence, in total 1200 samples were collected from 100 individuals.

The handwriting samples were analyzed to identify the similarities and dissimilarities among the class characteristics and peculiarities in the formation of letters. When examining the samples, the following features were taken into account:

- Alignment of handwriting with respect to the imaginary baseline,
- Inter-linear and intra-linear spacing,
- Style of handwriting,
- Size of handwriting,
- Slant of handwriting,
- Arrangement of handwriting on the paper,
- The angularity of the strokes,
- Overall line quality of the handwriting,
- Connections,
- Certain specific letter forms which possessed similarity between them.

The data collected were statistically analyzed by using the Chi-square test, to find out the significant similarities between the handwriting of the same individuals written in different scripts. The data was coded in binary values 0 and 1. The likeness between the characteristic was assigned as ‘yes’ and coded by 1, while the difference in character is considered as ‘no’ and denoted by 0.

Results

The collected handwriting samples were examined for the similarity between the class and peculiar characteristics in both the scripts inscribed by a single individual. The results are tabulated as given in Table 1 and Pearson Chi-square test was applied to find out the effect of language on the handwriting characteristics (Table 2). The analysis revealed that most of the individuals acquire similar handwriting characteristics irrespective of the language. Primarily, characteristics including alignment, spacing, style, size, slant, angularity, overall line quality and connections were analyzed and compared in the present investigation. Following this, the letter formation and overall arrangement of text on the paper were examined and compared in the handwriting of the same individual in Punjabi as well as in English.

After the analysis of samples, the component table was established to calculate the Chi-square value at the significance level 0.05, degree of freedom ($c = 1$), and the critical value 3.84 for each characteristic. If the calculated value will exceed the critical value, then the null hypothesis would be rejected, thus, an alternative hypothesis would be accepted because the calculated values are statistically significant; that is, both languages have similarities in class characteristics. Chi-square represented the frequency of samples written in two different alphabets having similarities in their class characteristics. The calculated values for each characteristic was found to be significant as given in Table 2, thence, the alternate hypothesis was accepted that depicts the similarities between the two different scripts.

Table 1

Representing the percentage of similarity between the characteristics of Punjabi and English handwriting

S. no.	Characteristic of handwriting	Percentage of likeness (%)	Percentage of dissimilarity (%)
1.	Alignment	82.6	17.4
2.	Inter-linear and intra-linear spacing	72.5	28.5
3.	Style	43.85	56.15
4.	Size	75.6	24.4
5.	Slant	76.8	23.2
6.	Angularity	87.8	12.2
7.	Overall line quality	76.9	23.1
8.	Connections	40.5	59.5
9.	Arrangement	75.3	24.7

Table 2

Representing statistical consideration Chi-square, sum of Chi-square and p-values for the handwriting samples examined

Handwriting characteristics	Type	Values	English	Punjabi	Sum of Chi-squares	p-value
1. Alignment of handwriting (a) df = 3 (b) expected value = 7.82	Uphill	Observed Expected Chi-square	10 12.50 0.50	15 12.50 0.50	8.90	0.03
	Downhill	Observed Expected Chi-square	39 32.50 1.30	26 32.50 1.30		
	Straight	Observed Expected Chi-square	12 8.50 1.44	05 8.50 1.44		
	Mixed	Observed Expected Chi-square	39 46.50 1.21	54 46.50 1.21		
2. Inter-linear and intra-linear spacing (a) df = 2 (b) expected value = 5.99	Narrow (0–2 mm)	Observed Expected Chi-square	14 10.50 1.17	7 10.50 1.17	6.50	0.03
	Medium (3–5 mm)	Observed Expected Chi-square	74 71.50 0.09	69 71.50 0.09		
	Wide (6mm-above)	Observed Expected Chi-square	12 18.00 2.00	24 18.00 2.00		
3. Style (a) df = 3 (b) expected value = 7.82	Cursive	Observed Expected Chi-square	42 27.00 8.33	12 27.00 8.33	87.17	0.00
	Manuscript	Observed Expected Chi-square	12 6.00 6.00	00 6.00 6.00		
	Handprinting	Observed Expected Chi-square	00 28.00 28.00	56 28.00 28.00		
	Composite	Observed Expected Chi-square	46 39.00 1.26	32 39.00 1.26		
4. Size (a) df = 2 (b) expected value = 5.99	Small (0–2 mm)	Observed Expected Chi-square	36 23.00 7.35	10 23.00 7.35	20.42	0.00
	Medium (3–5 mm)	Observed Expected Chi-square	54 61.50 0.91	69 61.50 0.91		
	Large (6 mm-above)	Observed Expected Chi-square	10 15.50 1.95	21 15.50 1.95		

Handwriting characteristics	Type	Values	English	Punjabi	Sum of Chi-squares	p-value
5. Slant (a) df = 2 (b) expected value = 5.99	Forward (0°–89°)	Observed Expected Chi-square	68 62.50 0.48	57 62.50 0.48	8.30	0.01
	Backward (91°–180°)	Observed Expected Chi-square	26 34.00 1.88	42 34.00 1.88		
	Straight (90°)	Observed Expected Chi-square	6 3.50 1.79	1 3.50 1.79		
6. Angularity of the strokes (a) df = 2 (b) expected value = 5.99	Angular	Observed Expected Chi-square	13 10.50 0.60	08 10.50 0.60	6.28	0.04
	Rounded	Observed Expected Chi-square	72 79.00 0.62	86 79.00 0.62		
	Oblong	Observed Expected Chi-square	15 10.50 1.93	06 10.50 1.93		
7. Overall line quality (a) df = 2 (b) expected value = 5.99	Good	Observed Expected Chi-square	21 28.50 1.97	36 28.50 1.97	6.09	0.04
	Moderate	Observed Expected Chi-square	60 56.00 0.29	52 56.00 0.29		
	Poor	Observed Expected Chi-square	19 15.50 0.79	12 15.50 0.79		
8. Connections (a) df = 1 (b) expected value = 3.84 (a) Intra-word	Present	Observed Expected Chi-square	84 60.0 9.60	36 60.0 9.60	48.0	0.00
	Absent	Observed Expected Chi-square	16 40.00 14.40	64 40.00 14.40		
(b) Inter-word	Present	Observed Expected Chi-square	16 13.00 0.69	10 13.00 0.69	1.59	0.20
	Absent	Observed Expected Chi-square	84 87.00 0.10	90 87.00 0.10		

1. Alignment of handwriting: The variations and likeness in the alignment of handwriting are depicted in Table 1. The character is exemplified in Figures 2a and 2b, where the given piece of handwriting in both scripts has mixed type of alignment, at certain instances, it is curved and at other, it is either downhill or uphill. It was expected that the alignment will change with the script because Punjabi letters are usually placed with a headline while English letters are placed on an actual or imaginary

baseline. However, on the contrary, it was observed that alignment in the case of Punjabi with respect to headline and English with respect to baseline was found to be similar in 82.6% of the samples. The sum of Chi-square for this particular feature is 8.90 that is higher than the critical value (Table 2), which ultimately represents the similarity of this character in writing of single individuals inscribed in different languages.

2. **Interlinear and intra-linear spacing:** Writing, inscribed in both Punjabi and English by an individual was examined and compared for interlinear and intra-linear spacing. It was observed that most of the individuals use medium spacing ranging from 3–5 mm and 72.5% of samples possesses similar spacing in Punjabi and English writings. The sum of Chi-square for this character is 6.50, which is higher than the critical value and represents the likeness of this feature in both languages.
3. **Style of handwriting:** To analyze this characteristic, four different styles of handwriting were considered, i.e., cursive, manuscript, handprinting, and composite (Huber, Headrick, 1999). In India people usually follow the cursive style in which uppercase and lowercase letters are connected through connections while writing English but Punjabi follows the different styles of handwriting that is handprinting (Jasuja, Komal, Singh, 1996; Kaur, Singh, Jasuja, 2017) where letters do not possess connections between them. It was found that 12% of individuals follow the cursive style and 32% of individuals follow composite style even in Punjabi handwriting, which may be due to the regular practice and their penmanship in the language. Overall, in 43.85% samples, this character is similar and *p*-value for this character is also significant.
4. **Size:** For the measurement of size measuring scale, having horizontal lines, each line is 1 mm apart from the other, was used. It was noticeable that 54% of individuals have written in the medium

range (3–5 mm) in English and 69% of individuals have also written Punjabi in the same range. Overall, 75.6% of samples showed a similarity in this characteristic. The *p*-value is highly significant for this characteristic and depicts the relation of both the languages with each other.

5. **Slant:** In the present study, the slant of vertical as well as rounded strokes has been measured with the aid of a template referred by Osborn (1929). Though, the method used to measure slant in English (from baseline) is different from that used in Punjabi (from headline), but the type of strokes (vertical and oval) chosen is similar in both languages. The results revealed that 76.8% of samples have a similar slant. Earlier studies revealed that the Punjabi population mostly use forward slant (Jasuja et al., 1996; Turner et al., 2008) in the present study it was found that 68% of English writings and 57% Punjabi writings have a forward slant. Figures 3a and 3b depict the measurement of slant in Punjabi and English handwriting of a person, although, way of measuring slant in both the languages is different but still slant in both the languages is similar that is 110.
6. **The angularity of the strokes:** In actuality, the angularity of a particular letter is defined in the copybook form, but writers usually deviate from the copybook and choose to write according to their will. The angularity can be categorized into three categories, i.e., angular, rounded, and oblong (Huber, Headrick, 1999). In the current study, 87.8% of

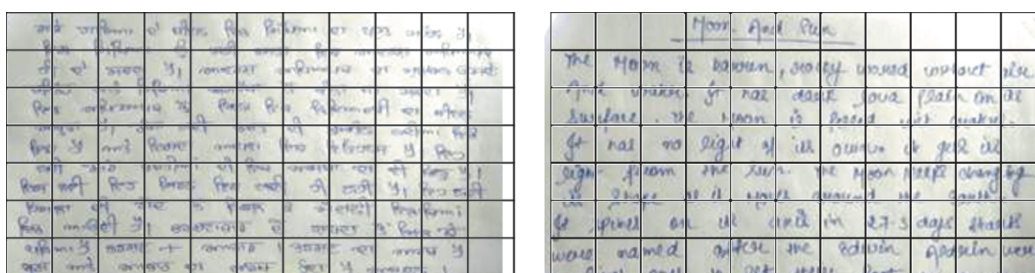


Fig. 2a, 2b. The alignment of handwriting of an individual in Punjabi and English.

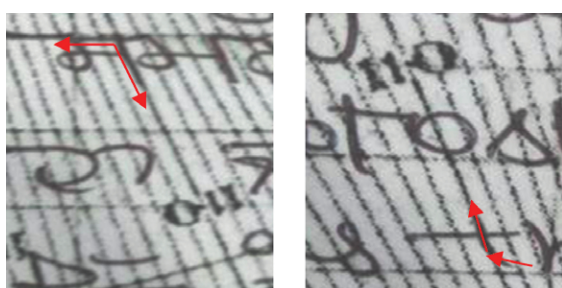


Fig. 3a, 3b. The measurement of slant of an individual’s writing in Punjabi and English scripts.

the samples depict similar angularity of the strokes. In English writings of the individuals, 72% handwriting samples possessed rounded strokes while 13% angular and 15% oblong strokes. On the other hand, 86% of the Punjabi handwritings possessed rounded strokes while 6% oblong and 8% angular strokes. The Chi-square value for this characteristic is significant representing the effect of one language variable on the other.

7. **Overall line quality:** The line quality of the samples written in Punjabi as well as in English was compared. All the participants were well known with the basic principles of handwriting using both languages. In 21% of samples of English handwriting, good line quality was observed but only 36% Punjabi handwriting samples possess this. The remaining 60% and 19% English handwriting and 52% and 12% of the Punjabi handwriting samples have moderate and bad line quality, respectively. 76.9% of samples revealed similarity in the overall line quality in the Punjabi as well as in the English handwriting. The Chi-square value for this characteristic is 6.09, which is significant and reveals a connection between the two languages.
8. **Connections:** In the case of English handwriting it is evident that connections should be present between letters and/or words, but in Punjabi in accordance to copybook, connections are neither placed between letters nor words. The present study revealed a 40.5% similarity in this characteristic between the two handwritings.
9. **Arrangement of handwriting:** To analyze this feature margins (left and right), placement of heading, numerals, and punctuation marks was included. After comparison, it was found that 75.3% manner of placement of text on paper is similar. Figures 4a and 4b represent the similarities between the patterns of arrangement of a piece of handwriting over the paper. In both the handwritings, the heading is

placed in the center without any kind of embellishment and underline. The left and right margins are present, but irregular, numerals, and punctuation marks were frequently placed on the baseline by the writer.

After calculating the Chi-square value for selected features, letter formation was analyzed and photographed as shown in figures 5–8. It was found that all the selected letters written by the single individual in two different languages possess similarities, which represents the effect of former vernacular on the latter one.

Figures 5–8 represent the letters of both English and Punjabi language presenting similar formation and structure. In figure 5, the similar letter formation of English letter ‘E’ and Punjabi letter ‘ੲ’ has been depicted. In both letters the center portion of the letters where two curves meet with retracing of the strokes. The movement of the strokes follows the same line in both letters, along with that writer initiated the letter with a small hook and connected it with the forthcoming letter. Figures 6 and 7 represent the same kind of formation in diacritics ‘ੜ’ and ‘ਫ਼’ of Punjabi and terminating stroke of ‘s’ and the letter ‘T’ of English. Figure 6 shows that the individual terminates the letter ‘s’ with an eyelet similar to the diacritic mark ‘ੜ’; both the letters are in continuation with the preceded letter. Similar letter formation of letter ‘T’ and Punjabi diacritic mark ‘ਫ਼’ is depicted in Figure 7. Figure 8 shows a similar letter formation of English letter ‘n’ and Punjabi letter ‘ਨ’ An eyelet is present at the meeting point of two shoulders; in both cases, the shoulders are rounded. In the case of Punjabi letter ‘ਨ’, the horizontal stroke is straight but while writing ‘ਨ’ and ‘n’ individual drew it in a bit slanting form. Therefore, the above studies and comparisons between the two vernaculars revealed that forensic document examiners can compare the documents written in distinctive scripts and the comparison of two distinct scripts and

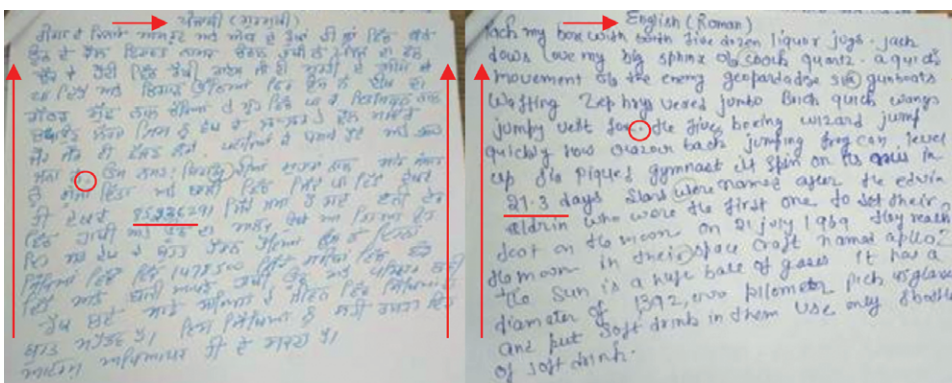


Fig. 4a, 4b. The arrangement of piece of handwriting.

languages can lead to a definite conclusion related to the authorship and authenticity.

Discussions

The samples written in two different scripts (Gurmukhi and Latin) were analyzed to unfurl the similar characteristics between two unrelated scripts and languages by examining various handwriting characteristics (alignment, inter-linear and intra-linear spacing, style, size and proportions, slant, arrangement, angularity of the strokes, overall line quality, connections and letter formation). Since both, the languages follow the different script and vast difference between the letters of both the scripts exists, but the present study revealed that still, numerous recurring identifying features are available, which can assist in the comparison of totally different scripts in an individual's writing and aid in the formation of opinions by the experts.

The prominent reason of the similarity between the writing of a person in two different scripts is that the individuals follow a peculiar method to make the strokes irrespective of the language or the script they intend to write. In India individuals usually learn more than one language due to many variations, as numerous different languages are constitutionally recognized for each Indian state. There are still few characteristics in their second language writing which are similar to the characteristics of primary language handwriting that are imparted. Saini and Kaur (2018) studied this effect by analyzing English handwriting of population from three different states of India: Punjab, Andhra Pradesh, Jammu and Kashmir, and their studies revealed that ethnicity and nationality can be determined with the assistance of handwriting due to the availability of certain characteristics in the writings of individuals that represents the influence of their regional language. In the present study the similar type of relationship between the regional language Punjabi and English also has been observed. Kaur et al. (2017) also presented the possibility of interscript comparison of three

different scripts and results depicted that two different languages can be compared in the absence of standards written in the same language. Singh et al. (1994) explained the effect of idiosyncratic features of handwriting on languages especially in cases of forgeries.

Ledić and Rosandić (2018) compared Latin and Cyrillic in handwriting and revealed that the comparison of two different languages can be made and if one person acquires skill in more than one language than the possibility of an effect of one language on another is there. In all the researches made so far; the possibility of comparison was introduced, but the exact reason is not explained. Yet this study tried to explain and concluded that the most probable reason behind this phenomenon can be the abstract form of the letter known as grapheme, as reported by Ellis (1979). The formation of strokes is similar if they are drawn by a single individual irrespective of the script. Therefore, the study revealed that the handwriting written in distinct languages are comparable with one another which can further help to find out the ethnicity, nationality and may be the region of the individual.

This study is still at the preliminary stage and with the advancement in the OCR technology these types of comparisons would be possible with more precision and repeatability.

Conclusion

It is a well-known fact that there are numerous evident distinguishable features of each script and language, but still the similar handwriting principles are applied to each language. This raises the possibility of a comparison of two different scripts that can assist the document examiners to compare different languages. Moreover, the possibility of presenting an exact opinion is vested with the capabilities and expertise of the expert, irrespective of the complexity induces by the language. It is expected that this preliminary study will assist forensic document examiners in cases of unfamiliar scripts and under complex situations,



Fig. 5. The similar letter formation of English letter 'E' and Punjabi letter 'ੲ'.



Fig. 6. The similarity in the terminating stroke of letter 's' and diacritic mark 'ੜ' of Punjabi.

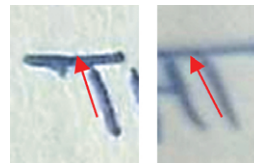


Fig. 7. The similar letter formation of English letter 'T' and Punjabi diacritic mark 'ੜ'.



Fig. 8. The similar letter formation of English letter 'n' and Punjabi letter 'ਮ'.

where standards in the same script may not be available. This study can assist in further authentication of the questioned documents especially when the standards are not available in the same script. This study needs to be explored geographically, further with other languages and populations, especially with respect to the Indian context.

References

1. Cheng, N., Lee, G. K., Yap, B. S., Lee, L. T., Tan, S. C., Tan, K. P. (2005). Investigation of class characteristics in English handwriting of the three main racial groups: Chinese, Malay and Indian in Singapore. *Journal of Forensic Science*, 50(1), 1–8.
2. Ellis, A. W. (1979). Slips of the pen. *Visible Language*, 13(3), 273.
3. English language. Retrieved September 20, 2020 from: <https://omniglot.com/writing/english.htm>.
4. Gill, H., Gleason, H. A. (1963). *A reference grammar of Punjabi*. Patiala: Punjabi University Publication.
5. Gurumukhi. Retrieved October 13, 2020 from <https://en.wikipedia.org/wiki/Gurmukhi>.
6. Huber, R. A., Headrick, A. M. (1999). *Handwriting identification: Facts and fundamentals*. Boca Raton: CRC Press.
7. Jasuja, O. P., Komal, Singh, S. (1996). Examination of Gurumukhi script: a preliminary report. *Science and Justice*, 36(1), 9–13.
8. Kaur, A., Singh, M., Jasuja, O. P. (2017). Interscript comparison of handwriting features leading to their identification and authorship. *Nowa Kodyfikacja Prawa Karne-go*, 45, 15–36.
9. Kelly, J. S., Lindblom, B. S. (2006). *Scientific examination of questioned documents*. Boca Raton: CRC Press.
10. Ledić, A., Rosandić, G. (2018). Comparing Latin with Cyrillic script in handwriting identification? *Policija i sigurnost*, 28(1), 27–43.
11. Masica, C. P. (1991). *Indo-Aryan languages*. Cambridge: Cambridge University Press.
12. Osborn, A. S. (1929). *Questioned documents*. Albany: Boyd Printing Co.
13. Saini, K., Kaur, N. (2018). Forensic examination of class characteristics in English handwriting of the three states: Andhra Pradesh, Punjab and Kashmir in India. *Arab Journal of Forensic Sciences and Forensic Medicine*, 1(8), 1034–1047.
14. Saxena, H. M., Singh, M. (1992). Classification of the writing elements in Devanagari scripts. *Journal of Forensic Science Society*, 32(2), 143–150.
15. Singh, A., Gupta, S. C., Saxena, H. M. (1994). Influence of the primary language and idiosyncratic features in simple forgeries. *Journal of Forensic Science Society*, 34(2), 83–87.
16. Turnbull, S. J., Jones, A. E., Allen, M. (2010). Identification of class characteristics of Polish people writing in English. *Journal of Forensic Science*, 55(5), 1296–1303.
17. Turner, I. J., Sidhu, R. K., Love, J. M. (2008). A preliminary study investigating class characteristics in Gurumukhi handwriting of 1st and 2nd generation Punjabis. *Science and Justice*, 48, 126–132.

Corresponding author

Amanpreet Kaur
 Department of Forensic Science and Chemistry
 RIMT University
 Mandi Gobindgarh, India, 147301
 e-mail: kaur.amanpreet74@gmail.com

PORÓWNANIE RĘKOPISÓW WYKONANYCH PISMEM *GURMUKHI* I ALFABETEM ŁACIŃSKIM POD KĄTEM MOŻLIWOŚCI ICH ZESTAWIANIA W BADANIACH IDENTYFIKACYJNYCH

Wstęp

Mówiąc o różnorodności językowej jako powszechnym zjawisku na świecie, warto zaznaczyć, że mieszkańcy Indii używają ponad stu języków i dialektów, a 22 z nich zostały wymienione jako urzędowe w indyjskiej konstytucji. Oprócz tego istnieje wiele innych dialektów oraz alfabetów używanych przez mieszkańców Indii, zarówno w mowie, jak w piśmie (Saini, Kaur, 2018). Oprócz angielskiego używają oni również swojego ojczystego języka w pracy zawodowej i życiu codziennym. W Pendżabie język pendżabski ma status języka urzędowego, chociaż część oficjalnej komunikacji odbywa się także w języku angielskim. Inne stany Indii powielają ten sam wzorec. W literaturze przedmiotu można znaleźć bardzo niewiele odniesień do zagadnienia, o którym mowa w niniejszej pracy. Metodyka badania pisma ręcznego do porównania rękopisów wykonanych za pomocą różnych alfabetów: łacińskiego i cyrylicy została zastosowana przez Ledić i Rosandić (2018).

W innym badaniu poddano analizie rękopisy sporządzone w języku angielskim przez osoby pochodzące z trzech stanów – odpowiednio Pendżabu, Jammu i Kaszmiru oraz Andhra Pradesh, biorąc pod uwagę różne cechy pisma, takie jak kształt liter, znaki interpunkcyjne, odstępy między literami itp. W badaniu tym wykazano znaczący wpływ języka regionalnego na język angielski (Saini, Kaur, 2018).

W niniejszej pracy, obok innych cech, uwzględniono budowę znaków, która jest pochodną abstrakcyjnej postaci litery zwanej przez Ellisa (1979) „grapheme”.

Pendżabski należy do grupy języków indoaryjskich (Masica, 1991), a współczesne zapisy w tym języku stosują pismo *gurmukhi* pochodzące od pisma *brahmi* (Saxena, Singh, 1992; Jasuja, Komal, Singh, 1996). Pismo *gurmukhi* zostało przejęte od pisma *shardha*, a do jego ukształtowania i rozwoju znacząco przyczynił się guru Angad Dev Ji w XVI wieku (Gill, Gleason, 1963). Językiem angielskim posługuje się 1,2 miliarda ludzi na całym świecie, przez co zyskuje on miano języka globalnego. Jest to język zachodniogermański, spokrewniony ze szkockim, holenderskim, fryzyjskim i niemieckim, który w okresie średniowiecza i wczesnej nowożytności rozprzestrzenił się z Anglii do Walii, Szkocji i innych części Wysp Brytyjskich, a także Irlandii. W konsekwencji w XVII wieku język angielski za sprawą handlu i kolonizacji stawał się popularny także w innych częściach świata, pojawiały się również jego nowe warianty.

W Wielkiej Brytanii Turner, Sidhu i Love (2008) prowadzili badania mające na celu określenie podobieństwa cech pisma używanego przez pierwsze i drugie pokolenie Pendżabów. Kolejne badania statystyczne, prowadzone przez Chenga i in. (2005), dowiodły, że na podstawie pisemnych wypowiedzi w języku angielskim udaje się potwierdzić narodowość i pochodzenie etniczne osób piszących – ich autorów. Turnbull, Jones i Allen (2010) analizowali anglojęzyczne rękopisy Anglików i Polaków, aby ustalić wpływ języka rodzimego na ich pismo.

Przegląd literatury ujawnił, że niewiele jest dostępnych badań dotyczących analizowania i porównywania rękopisów sporządzonych za pomocą różnych alfabetów, zwłaszcza w odniesieniu do języka indyjskiego, w którym występuje duże zróżnicowanie populacji i w każdym stanie używane są inne języki. Brak również jednoznacznych wniosków, które mogłyby zostać sformułowane przez różnych badaczy zajmujących się tym konkretnym zagadnieniem. Niniejsze badanie koncentruje się na porównaniu cech rękopisów sporządzonych w dwóch różnych językach – pendżabskim i angielskim, w oparciu o dwa różne alfabety – odpowiednio *gurmukhi* i łaciński – powszechnie używane przez mieszkańców Pendżabu. Ustala także sposób porównywania próbek pisma ręcznego.

Materiały i metody

W badaniu wzięło udział łącznie 100 osób sprawnych posługujących się pismem: 50 mężczyzn i 50 kobiet w wieku 18–50 lat. Uczestnicy badania odebrali podstawową edukację w języku pendżabskim, a na późniejszych jej etapach nauczyli się angielskiego. Wraz z rozwijaniem umiejętności pisania ręcznego wszyscy nauczyli się pisać w obu językach. Omawiane badanie zostało przeprowadzone w dwóch częściach: w pierwszym etapie wzajemne powiązanie dwóch języków ustalono na podstawie cech ogólnych pisma, a następnie – na podstawie budowy liter i ich elementów. Rycina 1a przedstawia zapisane w zeszycie do kaligrafii litery w języku pendżabskim, który jest rodzimym, podstawowym wyuczonym językiem dla tych osób, a rycina 1b przedstawia litery angielskie zarówno w formie majuskuł, jak i minuskuł. Przed uzyskaniem próbek pisma ręcznego otrzymano pisemną zgodę od badanych. Pobranie próbek nie było poprzedzone żadnym wcześniejszym szkoleniem w zakresie ich wykonywania. W celu uniknięcia trudności

i błędów w wynikach, które mogłyby być spowodowane naturalnymi różnicami w zgromadzonych próbkach, zebrano od każdego uczestnika po sześć próbek wykonanych w obu alfabetach za pomocą długopisu piszącego niebieską pastą i pióra wiecznego, na zwykłej białej kartce formatu A4. Badani zostali poproszeni, by, siedząc w wygodnej pozycji, zapisali wybrane akapity składające się z prawie wszystkich liter i znaków diakrytycznych w obu językach. Dwie próbki (jedną sporządzoną piórem kulkowym i jedną wiecznym) uzyskano, dyktując tekst, pozostałe cztery powstały podczas przepisywania tekstów drukowanych. W ten sposób od 100 osób pobrano w sumie 1200 próbek.

Zgromadzone w ten sposób wzory pisma ręcznego przeanalizowano w celu określenia podobieństw i różnic między cechami ogólnymi i budową znaków. Badając próbki, wzięto pod uwagę następujące cechy:

- sposób wyrównania znaków,
- odstępy międzywierszowe i wewnątrzwierszowe,
- typ pisma,
- wielkość pisma,
- nachylenie osi znaków,
- rozmieszczenie pisma na podłożu,
- ogólny obraz pisma,
- jakość linii graficznych,
- wiązania międzyliterowe,
- podobieństwa konstrukcyjne znaków.

Zebrane dane zostały poddane analizie statystycznej z użyciem testu chi-kwadrat, aby znaleźć znaczące podobieństwa między pismem ręcznym tych samych osób, lecz sporządzanym w różnych językach. Dane zostały zakodowane w wartościach binarnych 0 i 1. Podobieństwo między cechami zostało przypisane jako „tak” i zakodowane cyfrą 1, podczas gdy różnica jako „nie” i oznaczona cyfrą 0.

Wyniki

Zebrane próbki pisma zbadano pod kątem podobieństw cech w rękopisach jednej osoby wykonanych w obu językach. Wyniki zostały zestawione w tabeli 1, a test chi-kwadrat Pearsona został zastosowany w celu określenia wpływu języka na cechy pisma ręcznego (Tabela 2). Analiza wykazała, że większość osób nabywa podobne cechy pisma niezależnie od języka. W niniejszym badaniu zostały przeanalizowane i porównane przede wszystkim cechy takie, jak: sposób wyrównania znaków, szerokość odstępów między wierszami i w ich obrębie, typ pisma, nachylenie, ogólny obraz, jakość linii graficznych oraz wiązania międzyliterowe. Następnie zbadano i zestawiono budowę liter i ogólny układ tekstu na papierze.

Po przeanalizowaniu próbek została utworzona tabela składowa w celu obliczenia wartości chi-kwadrat na

poziomie istotności 0,05, stopnia swobody ($c = 1$) i wartości krytycznej 3,84 dla każdej cechy. Jeśli obliczona wartość przekracza wartość krytyczną, wówczas hipoteza zerowa zostaje odrzucona, a tym samym hipoteza alternatywna zostanie przyjęta, ponieważ obliczone wartości są statystycznie istotne. Oznacza to, że oba języki mają podobieństwa w cechach odnoszących się do klasy języka. Rozkład chi-kwadrat umożliwił przedstawienie częstości występowania próbek napisanych w dwóch różnych językach, które mają podobieństwa w swoich cechach klasowych. Obliczone wartości dla każdej cechy okazały się znaczące, jak podano w tabeli 2, a zatem przyjęto hipotezę alternatywną, która zakłada podobieństwa między zapisami w dwóch różnych językach.

1. Sposób wyrównania znaków: różnice i podobieństwa w wyrównaniu pisma ręcznego są przedstawione w tabeli 1. Przykłady zilustrowano na rycinach 2a i 2b – dany fragment pisma w obu zapisach ma mieszany sposób wyrównania, w niektórych przypadkach jest zakrzywiony, a w innych jest skierowany w dół lub ku górze. Oczekiwano, że wyrównanie będzie się zmieniać zależnie od języka, ponieważ litery pendzabskie są zwykle umieszczane z nagłówkiem, podczas gdy litery angielskie są umieszczane na rzeczywistej lub wymagowanej linii bazowej. Jednakże zaobserwowano, że wyrównanie w przypadku pisma pendzabskiego w odniesieniu do nagłówka i angielskiego w odniesieniu do linii bazowej było podobne w 82,6% próbek. Suma chi-kwadrat dla tej szczególnej cechy wynosi 8,90, co przewyższa wartość krytyczną (Tabela 2), a to ostatecznie oznacza podobieństwo tego znaku w piśmie pojedynczych osób zapisanym w różnych językach.

2. Odstępy międzywierszowe i wewnątrzwierszowe: pismo każdej osoby, sporządzone zarówno w języku pendzabskim, jak i angielskim, zostało zbadane i porównane pod kątem odstępów międzywierszowych i wewnątrzwierszowych. Zaobserwowano, że większość osób używa średnich odstępów w zakresie 3–5 mm i 72,5% próbek posiada podobne odstępy w pismach pendzabskich i angielskich. Suma chi-kwadrat dla tej cechy wynosi 6,50, co stanowi wartość wyższą niż wartość krytyczna i oznacza jej podobieństwo w obu językach.

3. Typ pisma: aby przeanalizować tę cechę, uwzględniono cztery różne typy pisma ręcznego, tj. zwykłe, blokowe, na wzór druku i mieszane (Huber, Headrick, 1999). W Indiach ludzie używają zazwyczaj pisma zwykłego, w którym małe i duże litery są łączone podczas pisania po angielsku, ale Pendzabcy stosują pismo na wzór druku (Jasuja, Komal, Singh, 1996; Kaur, Singh, Jasuja, 2017), w którym litery nie posiadają połączeń między sobą. Stwierdzono, że 12% osób stosuje pismo zwykłe, a 32% osób – pismo mieszane

nawet w pendzabskim piśmie ręcznym, co może wynikać z regularnej praktyki i wyrobionego charakteru pisma osób piszących w tym języku. Ogólnie rzecz biorąc, w 43,85% próbek ta cecha jest podobna i wartość p dla niej jest również istotna.

4. **Wielkość:** do pomiaru wielkości użyto skali pomiarowej posiadającej poziome linie, a każda linia jest oddalona od siebie o 1 mm. Zauważono, że 54% osób pisało w średnim zakresie (3–5 mm) w języku angielskim, w języku pendzabskim 69% osób pisało również w tym samym zakresie. W sumie 75,6% próbek wykazało podobieństwo tej cechy. Wartość p jest wysoce znacząca dla omawianej cechy i obrazuje związek obu języków ze sobą.
5. **Nachylenie:** w omawianym tutaj badaniu nachylenie trzonów i owali zostało zmierzone przy pomocy kątomierza Osborna (1929). Chociaż metoda zastosowana do pomiaru pochylenia w języku angielskim (od linii bazowej) różni się od tej zastosowanej w języku pendzabskim (od nagłówka), to rodzaj wybranych znaków (pionowe i owalne) jest podobny w obu językach. Wyniki wskazały, że 76,8% próbek ma podobne nachylenie. Wcześniejsze badania wykazały, że ludność pendzabska w większości używa pochylenia prawostronnego (Jasuja i in., 1996; Turner i in., 2008), natomiast w niniejszym badaniu stwierdzono, że 68% próbek pism angielskich i 57% próbek pism pendzabskich ma nachylenie prawostronne. Ryciny 3a i 3b przedstawiają pomiar pochylenia w pendzabskim i angielskim piśmie jednej osoby. Pomimo że sposób pomiaru pochylenia w obu językach jest inny, nachylenie w obu językach nadal jest podobne i wynosi 110.
6. **Ogólny obraz pisma:** formalnie kątowość danej litery jest określona podczas szkolnej nauki pisania, jednak piszący zazwyczaj odchodzą od tego i sami wybierają sposób jej zapisywania. W ogólnym obrazie wyróżniane są trzy kategorie: kątowe, okrągłe i owalne (Huber, Headrick, 1999). W obecnym badaniu 87,8% próbek przedstawia podobny obraz ogólny. W angielskich próbkach pisma badanych osób 72% z nich posiadało obraz okrągły, 13% kątowy, a 15% owalny. Jeśli chodzi o próbki pisma pendzabskiego, to 86% z nich posiadało obraz okrągły, podczas gdy 6% pociągnięć było owalnych i 8% kątowych. Wartość chi-kwadrat dla tej cechy jest znacząca i wskazuje na wpływ jednej zmiennej odnoszącej się do jednego języka na drugą, która odnosi się do drugiego z nich.
7. **Jakość linii graficznych:** porównano jakość linii próbek napisanych w języku pendzabskim, jak również w języku angielskim. Wszyscy uczestnicy badania byli dobrze zaznajomieni z podstawowymi zasadami pisma ręcznego w zastosowaniu do obu języków. Dobra jakość linii została zaobserwowana w 21% próbek angielskiego pisma ręcznego, ale tylko w 36% próbek

pisma pendzabskiego. Pozostałe 60% i 19% próbek angielskiego pisma ręcznego oraz 52% i 12% próbek pisma pendzabskiego ma odpowiednio umiarkowaną i złą jakość linii. 76,9% próbek wykazało podobieństwo w ogólnej jakości linii zarówno w pendzabskim, jak i angielskim piśmie ręcznym. Wartość chi-kwadrat dla tej cechy wynosi 6,09, co jest znaczące i ujawnia związek między omawianymi dwoma językami.

8. **Łączenia:** w przypadku angielskiego pisma ręcznego jest oczywiste, że łączenia powinny być obecne między literami i/lub słowami, ale w języku pendzabskim połączenia nie są umieszczone ani między literami, ani słowami. Obecne badanie wykazało 40,5% podobieństwa tej cechy.
9. **Rozmieszczenie pisma na podłożu:** do analizy tej cechy uwzględniono marginesy (lewy i prawy), rozmieszczenie nagłówków, cyfr i znaków interpunkcyjnych. Po porównaniu okazało się, że w 75,3% sposób rozmieszczenia tekstu na podłożu jest podobny. Ryciny 4a i 4b przedstawiają podobieństwa pomiędzy wzorami rozmieszczenia fragmentu pisma na podłożu. W obu językach nagłówek umieszczony jest na środku, bez żadnych ozdóbek i podkreśleń. Marginesy lewy i prawy były obecne, lecz nieregularne, natomiast cyfry i znaki interpunkcyjne były często umieszczane przez piszącego na linii podstawowej. Po obliczeniu wartości chi-kwadrat dla wybranych cech przeanalizowano kształt liter i sfotografowano je w sposób przedstawiony na rycinach 5–8. Stwierdzono, że wszystkie wybrane litery pisane przez jedną osobę w dwóch różnych językach wykazują podobieństwa, co stanowi wpływ pierwszego języka na drugi. Ryciny 5–8 przedstawiają litery języka angielskiego i pendzabskiego o podobnym kształcie i budowie. Na rycinie 5 przedstawiono podobną konstrukcję angielskiej litery „E” i pendzabskiej „ᱡ”. W obu przypadkach widoczna jest środkowa część liter, gdzie dwa półwałe połączone są nakrywkowo. Ruch narzędzia pisarskiego następuje po tej samej linii w obu literach, ponadto zaobserwować można, że piszący rozpoczął literę adjustacją i połączył ją z następną literą. Ryciny 6 i 7 przedstawiają ten sam kształt w diakrytykach „ᱠ” i „ᱡ” z pendzabskiego i odpowiednio w finalnej kresce minuskuły „s” oraz w budowie majuskuły „T” z angielskiego. Rycina 6 pokazuje, że osoba kończy literę „s” pętlącą podobną do znaku diakrytycznego „ᱠ”; obie litery są kontynuacją litery poprzedzającej. Na rycinie 7 przedstawiono podobny kształt litery „T” i pendzabski znak diakrytyczny „ᱡ”. Rysunek 8 przedstawia podobne uformowanie angielskiej litery „n” i pendzabskiej litery „ᱢ”. Element nakrywkowy znajduje się na styku dwóch ramion; w obu przypadkach ramiona są zaokrąglone. W przypadku pendzabskiej litery „ᱢ” pozioma kreska jest prosta, ale podczas pisania „ᱢ” i „n” poszczególne osoby nakreśliły ją w nieco pochyłej formie. Dlatego powyższe badania

i porównania między dwoma językami wernakularnymi wykazały, że biegły zajmujący się badaniem pisma może porównać rękopisy sporządzone w różnych alfabetach, a ich zestawienie może pozwolić na jednoznaczną identyfikację wykonawcy kwestionowanego rękopisu.

Dyskusja

Próbki napisane w dwóch różnych alfabetach (*gurmukhi* i łaciński) zostały przeanalizowane w celu znalezienia podobnych cech między dwoma niepowiązаныmi zapisami i językami poprzez zbadanie różnych cech pisma ręcznego (sposób wyrównania znaków, odstęp międzywierszowe i wewnątrzwierszowe, typ pisma, wielkość, nachylenie, ogólny obraz pisma, jakość linii graficznych, łączenia, rozmieszczenie pisma na podłożu). Chociaż oba języki posługują się innym alfabetem i istnieje ogromna różnica między literami, niniejsze badania ujawniły, że widoczne są liczne powtarzające się cechy identyfikacyjne, które mogą pomóc w porównaniu rękopisów sporządzonych za pomocą różnych alfabetów przez jednego wykonawcę i dzięki temu pomóc w formułowaniu opinii przez ekspertów.

Główną przyczyną podobieństwa rękopisów danej osoby w dwóch różnych alfabetach jest to, że zachowują one indywidualny charakter niezależnie od języka lub alfabetu, w którym zostały napisane. W Indiach ludzie zazwyczaj uczą się więcej niż jednego języka, ponieważ wiele różnych języków jest uznanych jako urzędowe w każdym stanie Indii, a obowiązujące języki występują także w różnych odmianach. W piśmie osób wykonanym w drugim nabytym języku występują pewne cechy, które wykazują podobieństwo z cechami pisma ręcznego utrwalonymi przez nie w języku podstawowym. Saini i Kaur (2018) badali ten efekt, analizując rękopisy anglojęzyczne ludności z trzech różnych stanów Indii: Pendżabu, Andhra Pradesh, Jammu i Kaszmiru, a ich badania ujawniły, że etniczność i narodowość można określić z pomocą pisma ręcznego ze względu na istnienie pewnych cech w pismach poszczególnych osób, które odzwierciedlają wpływ ich języka regionalnego. W niniejszym badaniu podobny rodzaj relacji między językiem regionalnym *pendżabi* i angielskim również został zaobserwowany. Kaur i in. (2017) również przedstawiły możliwość porównania interskrypcji zapisów trzech różnych języków, a wyniki pokazały, że dwa odmienne języki mogą być porównywane w przypadku braku próbek porównawczych napisanych w tym samym języku. Singh, Gupta i Saxena (1994) wyjaśnili wpływ idiosynkratycznych cech pisma ręcznego na języki, szczególnie w przypadkach fałszerstw.

Ledić i Rosandić (2018) porównały łacinę i cyrylicę w piśmie ręcznym, wykazując, że można dokonać porównania dwóch różnych języków, a jeśli jedna osoba

nabywa umiejętności w więcej niż jednym języku, istnieje prawdopodobieństwo wpływu pisma w jednym języku na pismo w drugim. We wszystkich badaniach przeprowadzonych do tej pory taka możliwość porównania została zauważona, ale dokładny mechanizm nie został wyjaśniony. W niniejszym opracowaniu podjęto próbę wyjaśnienia i stwierdzono, że najbardziej prawdopodobną przyczyną tego zjawiska może być abstrakcyjna forma litery znana jako „grapheme” (Ellis, 1979). Formowanie kresek jest podobne, jeśli są one stawiane przez jedną osobę niezależnie od pisma, którego używa. Dlatego badanie wykazało, że pisma ręczne napisane przez tę samą osobę w różnych językach są porównywalne ze sobą, co z kolei może pomóc w ustaleniu pochodzenia etnicznego, narodowości, a nawet regionu zamieszkiwanego przez tę osobę.

Badanie jest nadal na etapie wstępnym, a wraz z postępem w technologii OCR (*Optical Character Recognition*) opisywane rodzaje porównań będą z pewnością możliwe do przeprowadzenia z większą precyzją i powtarzalnością.

Podsumowanie

Powszechnie wiadomo, że istnieje wiele wyraźnie odróżniających się cech każdego alfabetu i języka, ale mimo to podobne zasady dotyczące pisma ręcznego mają zastosowanie w każdym z nich. Stwarza to możliwość porównania rękopisów sporządzonych za pomocą różnych alfabetów, co może pomóc badaczom dokumentów w porównywaniu próbek sporządzonych w różnych językach. Co więcej, możliwość przedstawienia dokładnej opinii zależy od możliwości i wiedzy eksperta, niezależnie od złożoności języka. Oczekuje się, że to wstępne studium okaże się użyteczne dla ekspertów z zakresu badania pisma ręcznego w przypadkach rękopisów sporządzonych za pomocą nieznanymi im alfabetów i w sytuacjach, w których próbki porównawcze sporządzone tym samym alfabetem co dowodowe są niedostępne. Powinno ono być rozszerzone pod względem geograficznym z uwzględnieniem innych języków i populacji, szczególnie w odniesieniu do kontekstu indyjskiego.