

An Exhaustive Meta-analytical Study of the History, Evolution and Development of ‘Saraiki NLP’

JATINDERKUMAR R. SAINI¹

Narmada College of Computer Application,
Zadeshwar, Bharuch, Gujarat – 392 011, India

¹saini_expert@yahoo.com

Abstract. The paper is first formal attempt worldwide to present a 200 years based chronicle of Natural Language Processing (NLP) research activities in resource-scarce Saraiki language, spoken mostly in Pakistan and India. It presents a single major bibliography till-date comprising of nearly 100 research instances of ‘NLP of Saraiki Language’ or ‘Saraiki NLP’ (the terms used first time in research community). The listed Saraiki NLP research instances are classified in one of six proposed categories for Corpus Development, Evolution, Machine Translation System, Phonetics, Script and Social-Science. State-of-art of Saraiki NLP sub-domains along with recommendations for potential sub-domains is presented. As a best attempt exhaustive annotated and classified bibliography of Saraiki NLP, the paper is intended to be a stepping-stone one-stop resource for future Saraiki NLP research activities.

Keywords: language, Multani, Natural Language Processing (NLP), Saraiki, Seraiki, Siraiki

(Received October 28th, 2018 / Accepted December 5th, 2018)

1 Introduction

Saraiki language is spoken by almost 20 million people in Pakistan as the first language while almost 78000 people in India too speak Saraiki language [68]. The Saraiki language is mostly spoken in various parts of Pakistan as well as North India. The people who have migrated from North India to other parts of India are also found to speak Saraiki language. This language is similar to Punjabi and Sindhi and has many dialects like Multani (being the most commonly found dialect), Riasati, Jafri, Hindki and Thali. The Saraiki language belongs to Indo-Aryan sub-family of Indo-Iranian language family. In the year 2007, based on the percentage-fraction of world population of native speakers, Saraiki language was ranked 60th out of 100 languages spoken world-wide [92].

The current research work has been done with intention of creating a bibliography of Saraiki Natural Language Processing (NLP) research tasks as well as finding, listing and directing possible future research avenues for Saraiki NLP. The paper brings home the

current state of art of Saraiki NLP. This paper intends to be a one point stop for various NLP research activities till date for the Saraiki language. The period of almost 200 years, for which the journey of research in Saraiki language has been analyzed, ranges from 1818 to 2018. For this purpose, the related research works were searched from various possible platforms like major journal publishers, various local journals as well as various indexing and abstracting service providers. Such platforms included, in alphabetical order, EBSCO, IEEE, ProQuest, ScienceDirect, Springer and SSRN, to name a few. For the searched records, mere inclusion of Saraiki word in a research paper was not treated as a sufficient criterion for inclusion in the current research work. Direct or indirect contribution to NLP field was treated as a requisite for consideration of a research work for the current work. In this process, many general papers and articles were also found for languages of Pakistan and North India. Each of the references listed in the current work have been compiled based on multiple sources and detailed here enough elaborately to act as a first step for further research in the field. Also, all the non-English

titles have also been provided with best-attempt near-equivalent English titles of the papers. Further, the source language with each paper too has been specified.

Given hardly any available Saraiki NLP research works, the current research work has not differentiated among the various dialects of the Saraiki language. The word ‘Saraiki’ has been found to be spelled in multiple ways including ‘Saraiki’, ‘Seraiki’ and ‘Siraiki’. The current work spells it as ‘Saraiki’ uniformly and without making any difference of the language characteristics based on the way it is spelled. The current work also does not distinguish between ‘Podhohar’ and ‘Pothohari’ dialects of Punjabi. The current work, further, also does not distinguish between differently spelled ‘Pashto’ and ‘Pushto’ language. Further, ‘Saraiki’ may refer to a language, a culture, a community, a dialect, etc. all of name ‘Saraiki’ but the current work refers to language only, unless explicitly stated. The current work also does not elaborate on Saraiki tribes or Saraiki nationalism, which seeks to establish a separate province in Pakistan. Finally, the current work also does not distinguish between ‘Multani’ and ‘Saraiki’ languages.

Rest of the paper is presented in fourteen sections. There are six sections listing the bibliographic references to various Saraiki NLP research works with each section having focus on Saraiki Language Corpus Development, Saraiki Language Evolution, Saraiki Language Machine Translation System (MTS), Saraiki Language Phonetics, Saraiki Language Script and Saraiki Language Social Science, respectively. Each of these six sections is followed by a respective sub-section containing remarks on the preceding section as well as discussions on existing research work done as well as possible future research in that area. The second last section of the paper body elaborates on the findings, in addition to already listed remarks in the foregoing six sub-sections. Concluding remarks are presented in the last section of the paper body.

It is noteworthy to mention that the line demarcating the classification of a research work in one category instead of other has been found to be thin. This is so because there are research works which may contribute to multiple categories of NLP tasks listed here. For instance, the work of Rasoolpuri [62] details about the script as well as the phonetics of the Saraiki language and has therefore been classified in both of these categories. It is also notable to mention that a paper classified in one category with a single word title may also contribute to many sub-fields of that category too. All the Saraiki language work instances presented here under various categories may not necessarily represent true research works from perspective of novelty, but

definitely they provide a first step in that direction. Most of the works of the Saraiki language also do not bear the characteristics of true NLP research because they were proposed before the formal definitions of NLP were finalized and put to implementation widely.

2 Saraiki Language Corpus Development

The first available resource for the Saraiki language is by Adward [2] and possibly it is the oldest of its kind. This is available in the form of a glossary of Multani language. Monnier [46], who actually is the printer of the book and not perhaps the author of the book on comparison of Sindhi and Punjabi languages with Multani, presents a detailed discussion on Saraiki proverbs and grammatical usage. The book intends to provide an almost exhaustive coverage of the usage of Saraiki language in day to day routine activities specifically carried out by Saraiki people. While going through the book, though it is clearly evident that most of the activities mentioned in the book have become a thing of past when compared with current times, the reader encounters an indeed candid treatise of the Saraiki language used in olden times. Brien [13, 14] too has presented a glossary of the Multani language along with its revised version. Brien and Mughal [15] have presented a Saraiki Urdu dictionary. Shackle [69] has presented an encyclopaedia of the Saraiki language.

The translation of holy book Kuran is provided by Abbasi [1] in the form of handwritten manuscript of Saraiki language. A similar Saraiki translation is also provided by Molana [45]. The translation of the holy bible in Saraiki is also available [82]. A corpus of Saraiki poetry is provided by Jampuri [32]. Skemp [83] have collected and presented a translation of various Multani stories.

Parveen and Anjum [55] have proposed the development of a bank of words translated from English language into six other languages namely Punjabi, Saraiki, Balochi, Urdu, Sindhi and Pashto. They have aimed to develop it as a language resource through which learners can better understand language. They propose to implement their developed vocabulary word-bank for school going children for their better education and betterment of language too. The Russian and English language works of Samarnov [87] and Shackle [88] respectively have been translated into Saraiki and in a way this too has provided the research community with parallel corpus for the Saraiki language.

Rauf et. al. [63] have presented a speech corpus developed by them for the purpose of Automatic Speech Recognition system for Urdu. They have recorded district names in six accents namely Punjabi, Urdu,

Saraiki, Pashto, Sindhi and Balochi. They have also taken care of various challenging factors for recording, including the disturbances that creep into while recording.

Amritsari [84] also has created a Saraiki language corpus consisting of the proverbs, sayings and verses of the Saraiki language. Akhter [3] in her thesis has presented an exhaustive study of idioms and proverbs of the Saraiki language. The specific noteworthy mention about her thesis is the clear differentiation presented by her among the idioms and proverbs, unlike many other authors who fail to bring out this differentiation and present an ambiguous or mingled picture of the similar treatment to idioms as well as proverbs.

2.1 Remarks on Saraiki Language Corpus Development

The researcher in current work has classified all attempts of linguists and researchers dealing with presentation of encyclopaedia, dictionary, book translations, various compilations and collections (specifically for poetry, proverbs and idioms) as well as the presentation of language glossary in the category of 'Saraiki Language Corpus Development'. This is done because from the perspective of NLP, specifically the statistical school of thought, corpus is the first step of NLP and any focused collection of language data is a corpus of that language. The corpora listed here, hence, provide a scope of further statistical NLP research in the Saraiki language. A single corpus, a combination of corpora or all corpora listed here together could be used for further research. Specifically machine learning based research work as well as a formal attempt for corpus building for the Saraiki language could be undertaken on these lines. Also, there is a definite scope for development of more parallel corpora with corpus consisting of paragraphs and documents instead of just words. This task in turn could be aid the development of Saraiki MTS too.

The researcher in current work believes that the works of various litterateurs as well as researchers listed in above section are actually pseudo NLP contributions because the intentions of authors were actually not inclined to contribute towards NLP. The development of corpus for a language is evidently a full-fledged research project in its own right. Hence, the current research work terms the various works listed above as 'pseudo NLP works'. However, it is also found that the research works of Parveen and Anjum [55] as well as Rauf et. al. [63] are both the true NLP research works and only exact NLP contributions as far as 'Saraiki Language Corpus Development' is concerned. This is so because from NLP perspective, Parveen and Anjum

[55] have actually proposed and created a parallel corpus with L_1 language being English and L_2 to L_n languages being Punjabi, Saraiki, Balochi, Urdu, Sindhi and Pashto, corresponding to value of n equal to 6. Similarly, Rauf et. al. [63] have also presented a parallel corpus with L_1 language being Urdu and L_2 to L_n languages being Punjabi, Saraiki, Pashto, Sindhi and Balochi, corresponding to value of n equal to 5. Though both these research instances are similar, there are two major differences. Firstly, the value of n is 6 in former case whereas 5 in later case and secondly, the former relates to parallel corpus for text while the later relates to parallel corpus for speech. The presentation of various translations by authors for religious holy books like Quran and Bible as well as those for Saraiki stories are all instances of creation of parallel corpus from the perspective of Saraiki NLP.

3 Saraiki Language Evolution

This category presents the Saraiki language works which directly indicate the authors' intention of presenting the evolutionary stages of the Saraiki language. The word 'directly' has been used here emphatically because most of the works themselves have used the word 'irteqa', meaning 'evolution' for the Saraiki language. For instance, the works of Bahawalpuri [9], Nasir [50] and Zami [79] fulfil this criterion. In a broader sense, the other similar works present the historical growth and development of the language as well as insights into Saraiki linguistics. The works of Bakhsh [11], Gilani [18], Hussain [26], Kalanchvi [38], Mughal [49], Pervaiz [56], Qais [57], Rasoolpuri [59], Sajjad [67] and Wagha [78] as well as Haq [21], Kalanchvi [36] and Rasoolpuri [61] respectively fulfil these criteria.

Gilani [18] has provided very interesting discussion on origin and evolution of word India as well as the evolution of the Saraiki language itself. He has strongly advocated the influence of Persian and other rulers as well as their languages over the evolution of the Saraiki language. Wagha [78] has discussed about the growth and development of the Saraiki language with specific context of Pakistan. Haq [20] has presented a comparison of Multani language with Urdu and depicted their relationships. Haq [23] has presented and discussed a comparative view of Saraiki language and other languages spoken in the areas where Saraiki is spoken. Hussain and Farooq [24] have lamented the issues on Punjabi and Saraiki language where the people speaking one of these languages have termed the other language as the dialect of their language. Their paper argues that both the languages have regional identity and have significant differences in context of historical as well as linguistic perspectives. They have also

focused on the separate identity and existence of Saraiki language presenting reasons like its inclusion in the 'Paishacha' and 'Dardic' group of languages as well as similarity and impact of other languages like Pashto, Punjabi, Pothohari, Sindhi and Persian on vocabulary of Saraiki language.

Hussain and Khan [25] have proved the close linkage and relationship of Pushto and Saraiki languages. They have also discussed the morphological and syntactic similarity of both of these languages. Hussain [26] has proved that Saraiki language has links with 'Pashachi' and 'Dardik Alsana'. He has further added that there is a lot of Islamic influence on the Saraiki language and that the later has hence largely benefited from the Arabic and Persian languages which have contributed many words and proverbs too to the Saraiki language. Singh [90] has presented a detailed discussion on the comparison of Majhi and Multani variants of the Punjabi language. He has also claimed that the Multani language has a large percentage of Perso-Arabic words as well as it has many diminutives.

An ethnologue of the Multani language has been presented by Lewis et. al. [42]. Khan [89] has provided a detailed discussion focusing on the genealogy of the word 'Saraiki' as well as the proliferation of the Saraiki language. Raza [64] has provided an etymological evidence based on the linguistic and phonological discussion about the derivation of the word Saraiki. He advocates that neither *Siraikī* nor *Seraikī* are the valid forms of the usage of the word. He has also proved the origin of the word Saraiki to the word *Sauvīra* and claimed that the word Saraiki is not derived from the word *Siro*.

Wikipedia [68] at its page on Saraiki dialect has mentioned that Saraiki language shares a large amount of morphology and vocabulary with Punjabi language. It also mentions that Saraiki language also shares many grammatical features with Sindhi language. It adds further that even though there are similarities of Saraiki language with Punjabi and Sindhi languages the phonology and accent usage of Saraiki language is quite distinct.

Singh and Kaur [74] have reviewed the dialects of Punjabi language. During their review, they have found that the main dialects of Punjabi language are Majhi, Malwai, Doabi, Powadhi, Multani and Podhohar. They have further advocated that Saraiki is same as Multani and that it is the dialect of Punjabi language which has originated from Pakistan.

3.1 Remarks on Saraiki Language Evolution

Apart from depicting the Saraiki language's roots from the cultural, geographical and linguistic angles, the litterateurs and researchers have attempted various works to trace its origin and have presented elaborate discussions on the history, growth, development, decline and issues of the Saraiki language. There is clear evidence from all of these instances, as listed above, that at times the works have focused on Saraiki community and the Saraiki language has been studied on the sidelines and at times, the Saraiki language has been studied with specific concentration and the cultural dimension has been on the side-lines. Irrespective of either case, the Saraiki language's evolutionary study has been treated as a major element to qualify a work in the category of 'Saraiki Language Evolution' by the researcher in the current work.

The works presented in this category have depicted the evolution and derivation of Saraiki language as well as also presented the linkage, relationship and influence of other languages on the Saraiki language. To summarize the contributions of various authors, the current research work claims that the Saraiki language has been linked with various languages, namely, Arabic, Doabi, Majhi, Malwai, Multani, Pashto, Persian, Pothohari, Powadhi, Punjabi, Sindhi and Urdu.

4 Saraiki Language MTS

Bhaiya [12] has presented the description of the rules of usage of the Saraiki language. The grammatical rules for the usage of the Saraiki language have also been presented by Haq [22]. Hussaini [29, 30] have published the knowledge books for the usage of Saraiki language. Shackle [72] has presented a reference grammar of the Saraiki language. Khawaja [85] has deployed an online blog for the Saraiki language. Though his work is not a formal research work, but the presented concepts about the verbs, tenses and grammatical usage of the Saraiki language could act as a key contribution to the development of Saraiki Part-of-Speech (POS) tagger and in turn, Saraiki MTS.

Raza [65] has presented the notion of development of transfer-based Machine Translation System (MTS) for language pair consisting of languages Saraiki and Urdu. He has also lamented the lack of availability of resources required for implementation of otherwise sophisticated techniques of artificial intelligence for designing MTS. The paper also discusses about the problems and opportunities for MTS for such languages. In another paper, Raza [66] has presented an interesting comparison among Persian, Urdu and Saraiki languages.

The basis used for comparison is adposition usage in the respective language. On the sidelines of the paper the author has also provided a detailed description of basic adpositions and compound adpositions.

4.1 Remarks on Saraiki Language MTS

As such, from the perspective of NLP, only one true MTS application has been proposed by Raza [65]. Even he too has lamented the lack of resources for the development of a full functional MTS for the Saraiki language. It is notable that another work of Raza [66] dealing with adpositions is actually a work on Part-of-Speech (POS) tagging. Through the current work, the researcher has classified, compiled and presented various works dealing with the Saraiki language and its grammar at a single location. It is noteworthy that for the development of an accurate MTS, it is very important to have clear comprehension of the rules of the language used in MTS either as a source or destination language. Hence, through the current work, the researcher believes that a lot of scope is evident for the development of a MTS which makes use of the Saraiki language. The, claimed to be exhaustive, listing of works presented here and providing rules for the usage of the Saraiki language is definitely going to act as an extremely useful resource for the development of Saraiki MTS, specifically the rule-based one, in future.

5 Saraiki Language Phonetics

Awan et. al. [8] have provided a detailed discussion of Saraiki language. They have discussed about the varieties of Saraiki language as well as lamented the fact that most of work done for Saraiki language was from non-native speakers of the language. The key contributions of their paper include a definite and thorough comparison of vowels and consonants of Saraiki language and English language. The variety of Saraiki used for comparison purpose by them is Multani. A detailed study of injective consonants usage in Saraiki language has been presented by Bahl [10].

Farooq [16] has provided a phonetic based comparison of accent of Urdu language spoken in Pakistan. The six different languages used for comparison by the author include Sindhi, Punjabi, Saraiki, Urdu, Balochi and Pashto. In a similar work, Afsheen [93] et al. classified accents of Urdu language among Punjabi, Urdu, Pashto, Saraiki and Sindhi languages using Mel Frequency Cepstral Coefficient (MFCCs) and feature formants. Farooq and Hussain [94] have presented an analytic study of phonetic influence of six languages namely Urdu, Punjabi, Sindhi, Balochi, Pashto and Saraiki on the pronunciations of vowels of true Urdu language.

Gardezi and Hassan [17] have discussed about the Saraiki language and presented an introductory note on the poetics of the Saraiki language. Jumani et. al. [34] have presented a detailed discussion on the influence of Saraiki language on the pronunciation of English words during the usage of English language by speakers whose native language is Saraiki.

Latif [41] has worked extensively on the phonemic inventory for the Saraiki language and also proposed the analysis of voiced implosives of the language based on acoustics. He laments the lack of availability of corpora for research purpose for the Saraiki language and experimented on Multani dialect of the Saraiki language for his research purpose. In a similar work, Moran et. al. [47] have presented a sound inventory for Saraiki language. Rasoolpuri [62] also has presented the description about the sounds of Saraiki language words. Mughal [48] has provided a detailed description on the peculiar sounds used in the Saraiki language. Malik et. al. [44] have proposed the use of Finite State Machines with an aim of finding solution to the problem of scriptural translation. They claim that their model, based on finite state transducers and Universal Intermediate Transcription, is reported on Saraiki, Punjabi, Hindi and Urdu languages but can also be used with extension for other languages. They have emphasized the significance of subjective evaluation of such models before implementation and rolling them out for usage in the real world scenarios.

Nasir [51] has provided a detailed description on nasality of the Saraiki language. He has explained about nasality level in language segments, phonotactics of the language, the nature of word-medial nasals as well as the analysis of the relation between contextual and independent nasalization. Nasir [52] has presented a phonetic discussion of Saraiki language in context of comparison with English language. An in-depth study and analysis of nasalization in Saraiki language with context of the phonetic properties of the language has been provided by Syed [75]. The research work of Syed [76] aimed to notice the Saraiki language's sensitivity towards the stress required for heaviest syllable in a word. Similarly left-headedness posing another scenario for stress requirement too was studied by him. He concluded that many seemingly contradictory modifications take place, specifically for the Saraiki language words which are borrowed from Arabic language.

5.1 Remarks on Saraiki Language Phonetics

There are a good number of instances that could be truly called 'research works' and are available for Phonetic

work on the Saraiki language. The current work has classified all such instances under the common class of 'Saraiki Language Phonetics' and includes research works which deal with Saraiki language's comparison of vowels and consonants, study of injective consonants, comparison of accent with Urdu, scriptural translation, phonotactics, poetics, varieties, nasality and phonetics of loan-words from Arabic language. The influence of Saraiki language on the speakers of English has been discussed for speakers whose native language is Saraiki. The nasality comparison of English and Saraiki language too has been presented.

The research works also include lamentation on many Saraiki language works being done by non-native speakers of the Saraiki language as well as lack of formal corpus for Speech and Phonetic research work in the Saraiki language. But at the same time, it has been found through the current work that a good number of speech corpora have been developed by researchers for the Saraiki language. There is a definite need, now, to standardize these corpora and create uniform common corpus for the Speech and Phonetics related NLP tasks in the Saraiki language.

Saraiki language is very typical as far as its characteristic dimensions like pronunciations, nasality and phonetics are concerned. In fact it differentiates from many other dialects based on these features. Hence, the current work advocates that there is a lot of research scope in Saraiki language from the said dimensions.

6 Saraiki Language Script

Pandey [53] submitted a formal preliminary proposal to encode the Multani script. He [54] also submitted a formally finalized proposal to encode the Multani script. He has also presented a comparison of Multani, Khudawadi, Khojki, Gurmukhi, and Devanagari letters with the Multani script of the Saraiki language. Rasoolpuri [62] has presented a description on the written letters of the Saraiki language. Jan and Saleem [33] have deployed Neural Network (NN) to recognize written characters of Saraiki script. They have used MATLAB with segmentation of character set of the language and achieved 85% accuracy in their Optical Character Recognizer (OCR).

Anonymous researchers, programmers and website developers have contributed informally to the development of Saraiki language script-based research. This includes presentation of Saraiki font and keyboards for download for Windows and Android operating systems [5], Saraiki alphabets with its corresponding

Gurmukhi equivalents [6] and Saraiki online transliteration [7].

6.1 Remarks on Saraiki Language Script

The researcher through the current work believes that the area of Saraiki language script-based NLP research is the only area where the extremes from completely formal NLP research to completely non-formal research exist. The Unicode encoding of Saraiki script and the usage of NN and MATLAB for Saraiki script OCR are true script-based NLP researches for the Saraiki language. Given this presentation and listing of single digit research instances of the script-based NLP research in Saraiki language, there is a lot of scope for Saraiki script OCR development and improvement.

7 Saraiki Language Social Science

Alam and Gill [4] have presented a study of a small group of subjects to advocate that there is a definite influence of mother tongue on the learners of other languages. They have also highlighted through the case study of students of higher education of a university that students try to grasp the concepts in any language but mentally they process it in their mother language only. The languages used for their research work include Pashto, Saraiki and English. Gilani [19] has discussed about the different varieties of translations of The Holy Quran. The language focused by the author is Saraiki and has been studied on different basis like lexis, poetic and semi-idiomatic.

Hussain et. al. [27] have highlighted the importance of oral traditions and specifically focused on the way by which traditional values are transferred from one generation to the other through the use of folktales. They have deployed the case of three different folktales for case study of Saraiki farmers. They have also proposed recommendations for the educational policy for the Saraiki speaking region of Pakistan. Hussain et. al. [28] have studied the folktales of Saraiki language and tried to find out the gender usage in such folktales. They have found that such folktales are largely dominated by male gender. Khan and Shami [80] have elaborated on the benefits that Europeans had owing to the knowledge they acquired from the Saraiki language. Perveen [81] has discussed about the communication about HIV using the Pakistani languages including Saraiki.

Khokhlova [40] has lamented the death of a language which is spoken by majority of people in the country. The language lamented by the author is Punjabi language wherein Saraiki language has been included as a dialect of the Punjabi language. The

author has used the term 'intellectual death' of the language for Punjabi language's decreasing usage in the country. Mahar et. al. [43] have depicted the crane bird as the symbol of affection and love through the analysis of poetry written in Saraiki language. They have also cited examples from the Saraiki language's Sufi and contemporary poets namely Baba Fariduddin Ganj Shakar, Ashoo Lal and others. They have also discussed the change in symbolism of the bird with time. Shackle [86] has presented a discussion on the styles and themes in the mystical poetry of the Saraiki language of the Sind region of Pakistan. Shackle [91] has presented a thesis based upon the study of the Saraiki poetry of the writings during the period of 1750 to 1900 by the seven poets. He has further narrowed the scope of research by considering the seven poets only from the Upper Sind and South-West Punjab regions of the Indian sub-continent.

Shackle [70] has presented the discussion on Saraiki language in the context of the language movement in Pakistan with specific focus on Saraiki language. Shackle [71], way back in the late 70s, lamented the research focus required to be given to languages like Saraiki. He has specifically focused on the unique style of writing poetries in the Saraiki language. In the context of planning of a language for a nation and the problems that a language can pose, Rahman [58] has discussed about the Saraiki language movement in Pakistan.

Syed et. al. [77] have conducted an experiment to show the attrition of Saraiki language by the other languages spoken by people in the areas inhabited by Saraiki-speaking population. It is noteworthy that the other language focused by researchers was Hindi and its influence as a dominating language has been studied in detail by the researchers in the context of reduction in use of Saraiki language by even native speakers of Saraiki language against the usage of Hindi by them.

7.1 Remarks on Saraiki Language Social Science

The researcher in current work has categorized those research works in category of 'Saraiki Language Social Science' which deal with Saraiki language but their NLP focus is negligible and they have typical social science relevance. It is noteworthy that the said works do have NLP focus and relevance but it is either quite less or in-direct. Hence a research work which merely makes use of the word Saraiki language is not considered for classification in any category in the current work. In fact, numerous such works exist but have not been counted for in the current work. The research works of Imtiaz [31], Kalanchavi [35],

Kalanchvi [37], Khan [39], Rasoolpuri [60] and Shoukat [73] are a few examples of the same. It is notable that this does not indicate these research works are insignificant in totality but it indicates that they have not been treated with relevance for the current work.

The research works in current category mainly include cases of influence of mother tongue on learners' mind, attrition of Saraiki language, transfer of oral traditions through Saraiki language, Saraiki poetry symbolism and language movement in Pakistan with specific focus on Saraiki language.

8 Additional Remarks and Findings

This section presents findings in addition to the remarks already listed above corresponding to respective sections on Saraiki NLP. It has been found that Saraiki literature, and hence its language unlike many other languages of the world, is very rich in agricultural terms, proverbs, folktales and poetry ('shaayari' as well as 'dohrey' forms). The author believes that it is the language of common man and mostly used in the remote interior parts of the villages in its purest form. Though influenced by many other languages, like any other language in the world, has still remained a lot aloof from possible alien-language attacks. The context of this statement becomes more evident from the following scenario. There are natural languages like Hindi, Punjabi, etc. for which the speakers use words of other languages (mostly English) freely during conversation. The author believes that this is the process of slow-death of the language. This is so because each substitution of another language word in a conversation going in a language attempts to kill the original word, which perhaps for the next generation completely ceases to exist. In this context, Saraiki has still the continuing legacy which is dwindling with time though. Obviously, languages do not remain constant, new words keep on adding and old words keep on being removed from its usage. But there is difference in both these cases. The first case corresponds to language wearing away due to substitution of other words and decreasing vocabulary of the speaker for the native language. In second case, the old words get removed because of time. For instance, a word like 'dhonkni' meaning 'Blowpipe or Aerator, used in kitchens' will not be heard of in times to come as people have stopped using it owing to advent of new technologies like gas-stoves and electric-stoves. In either case, it is strongly felt that unless there is a dictionary of a language, the word will be completely wiped out from world. It is assumed that the dictionary contained the word and new editions of dictionary too retained that word.

It is notable here that though resources like magazines, news-papers, TV channels and YouTube

videos exist for Saraiki language, there is no formal attempt by any researcher till date to standardize them uniformly and create a specifically focused corpus. This is also the reason of their non-inclusion in ‘Saraiki Language Corpus Development’ category. A lot of work on language movement in Pakistan has been found but not included for the current research work as it does not contribute to core Saraiki NLP.

Maximum work on Saraiki language has been done by Muslim authors whereas Christian authors rank second. Hindu and Sikh authors, both rank third as far as the community-wise number of research instances are concerned. From geographical perspective, maximum number for research works in Saraiki language hailed from Pakistan followed by United Kingdom (UK), United States of America (USA), India and France, in the listed ascending order of number of research works. These findings have a logical justification because the Saraiki language is spoken mostly in Pakistan which is a Muslim-majority nation. After Pakistan, maximum number of Saraiki language speakers are present in India. Also, the Saraiki language is similar to Punjabi language and hence the Sikh research workers for the Saraiki language have been found too. It is notable that irrespective of community viz. Muslim, Hindu, etc., the native language of a person could be Punjabi. The Saraiki language research instances by Christian authors and those from UK could be understood from the fact that both India and Pakistan were under the British rule pre-1947. The country of origin and community for authors of three research works could not be established as the works were presented online in the form of websites and not through the formal way of publication of research findings through thesis, books or research papers.

Out of the 94 NLP research works found corresponding to Saraiki language, as many as 32 (nearly 35%) were found to be scripted in Urdu, Shahmukhi Punjabi, Gurmukhi Punjabi or Saraiki language itself, one was found to be scripted in French whereas all other were found to be scripted in Roman English script. It is notable here that the author of current research work himself is a native speaker of Punjabi and Saraiki. A few words from Urdu were understood with the example of Translator facility of Google Inc. Examples of such words include ‘irteqa’ meaning ‘evolution’, ‘lisaaniyaat’ meaning ‘linguistics’ and ‘saqaafat’ meaning ‘culture’. The author of current research work strongly believes that the Saraiki language NLP research work being done by non-Saraiki speakers as well as non-availability of many important Saraiki language works in international languages like English are two main reasons responsible for non-inclusion of Saraiki language NLP research in the

otherwise main-stream of the field. To the best of author’s knowledge, the current work is the first formal research work worldwide which actually has used the coined terms ‘Saraiki Language NLP’ and ‘Saraiki NLP’.

It has been found from the analysis of various research works presented in the current paper that comparison of Punjabi and Sindhi languages with Saraiki language contributes to more than 50% of the total research works based on comparison of some language with the Saraiki language. Table 1 presents the statistics on such comparisons of various languages by the researchers with the Saraiki language. Table 1, being sorted on the number of research works in descending order, also indicates the proximity of languages like Punjabi, Sindhi, Pashto and Urdu with the Saraiki language. It also paves further future research path for exploring the comparison and linkage of other languages which are less-researched and less-studied but otherwise known to be closely related with Saraiki language. Such languages include Persian, Arabic, Balochi, Majhi, Pothohari and Dogri. It is notable to mention that the statistics on comparison of various languages with the Saraiki language presented in Table 1 deal with the language comparisons which focus on nasality, phonetics, accent, vocabulary and grammar of the language. Only one research instance by Pandey [53] has been found which formally presents detailed comparison of script of Saraiki language with that of scripts of Multani, Khudawadi, Khojki, Gurmukhi, and Devanagari letters.

Table 1: Number of Research Works based on Comparison of Saraiki Language

Sr. No.	Compared Language	Number of Research Works
1	Punjabi	7
2	Sindhi	7
3	Pashto	5
4	Urdu	5
5	Persian	3
6	Arabic	2
7	Balochi	2
8	English	2
9	Majhi	1
10	Pothohari	1
Total		35

Out of a total of 94 Saraiki NLP research papers listed in the current work, Figure 1 provides a time period based distribution of number of such research

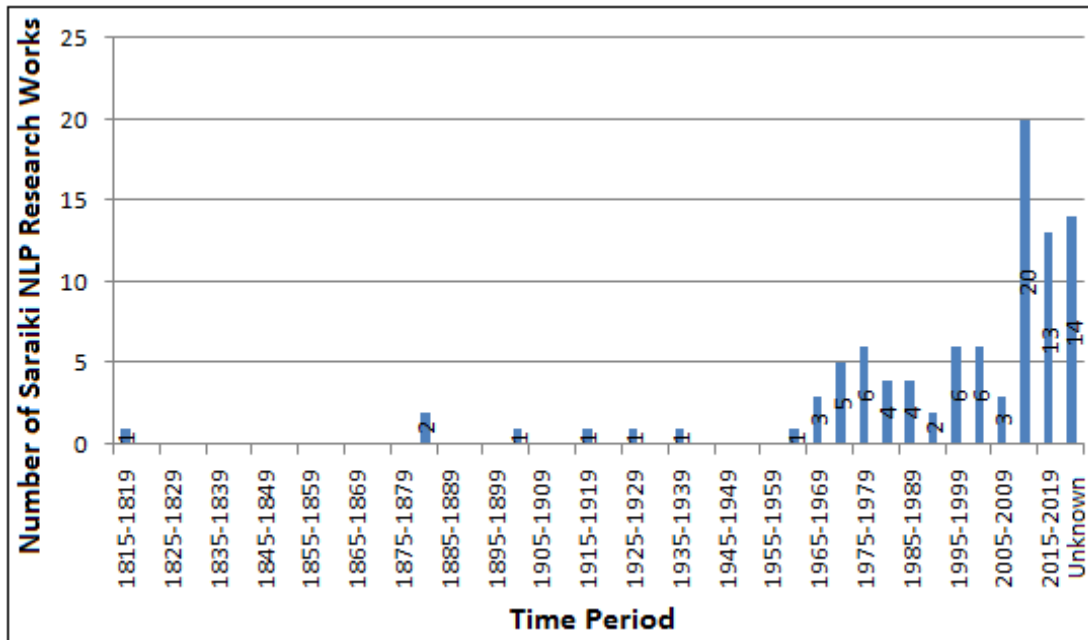


Figure 1: Distribution of Number of Saraiki NLP Research Works Over Time Period

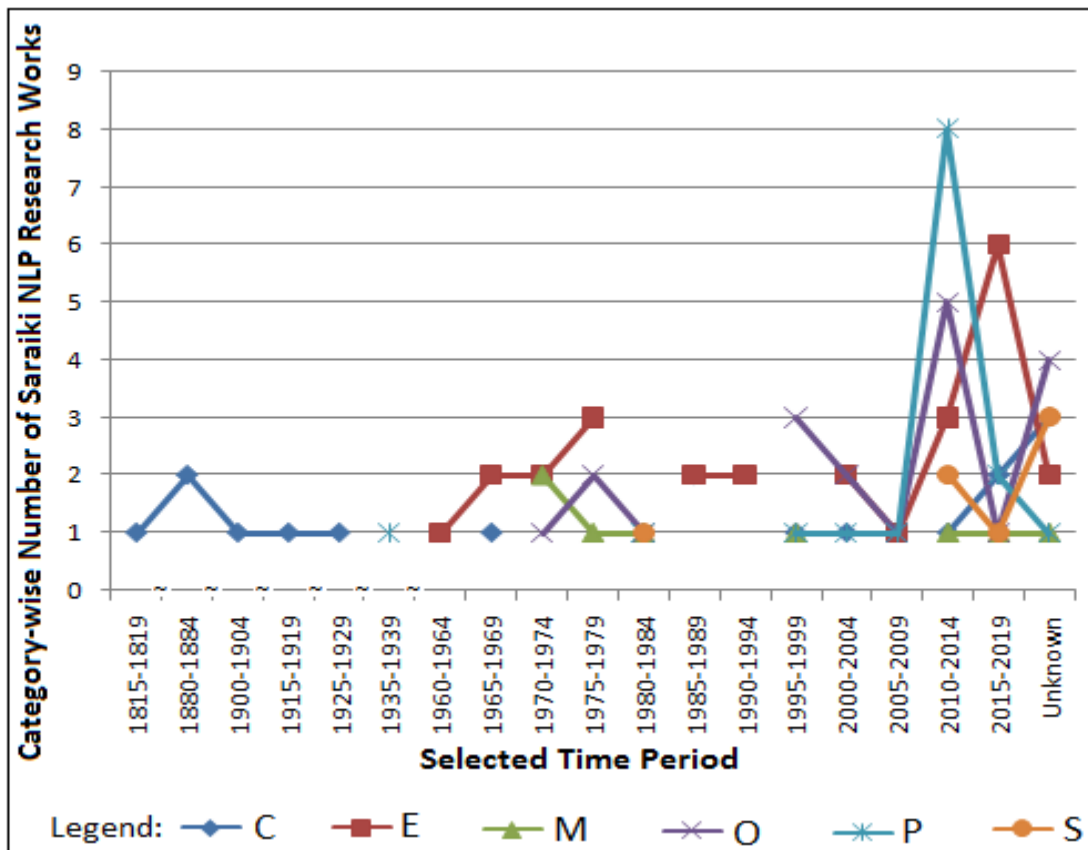


Figure 2: Distribution of Category-wise Number of Saraiki NLP Research Works Over Selected Time Period

works for a period of nearly 200 years. It is remarkable that 1960s triggered the active Saraiki NLP research activities, the trend that has more or less continued till the first decade of 21st century. The last decade from 2010 till the date of final compilation of this paper in November 2018, has seen a sharp increase in the number of such activities. As the number of non-zero research works for each year has been included in the graphical representation itself, a separate table has not been provided for the same redundantly. The graphical representation also shows 14 research works for which information about year of publication is not available.

Table 2: Period-wise Category-wise Number of Saraiki Language NLP Research Works

Year	C	E	M	O	P	S	Total
1815-1819	1						1
1880-1884	2						2
1900-1904	1						1
1915-1919	1						1
1925-1929	1						1
1935-1939					1		1
1960-1964		1					1
1965-1969	1	2					3
1970-1974		2	2	1			5
1975-1979		3	1	2			6
1980-1984			1	1	1	1	4
1985-1989	2	2					4
1990-1994		2					2
1995-1999	1		1	3	1		6
2000-2004	1	2		2	1		6
2005-2009		1		1	1		3
2010-2014	1	3	1	5	8	2	20
2015-2019	2	6	1	1	2	1	13
Unknown	3	2	1	4	1	3	14
Total	17	26	8	20	16	7	94

This leads us to another important question of analyzing the type of activity that was undertaken during the period of nearly 200 years. Figure 2 presents a graphical representation of the distribution of type of Saraiki activity over the period of time. Figure 2 should be viewed in color for better comprehension. Even though the data represented in Figure 2 coincides with the data represented in Figure 1, for simplicity, the years having zero values corresponding to number of activities have been removed in Figure 2. This was not

done for Figure 1 as its main purpose was to find, understand and analyze the distribution of Saraiki NLP research activities over a period of time and hence to analyze a clear time-based pattern of trend. The graphical representation in Figure 2, similar to Figure 1, also shows 14 research works for which information about year of publication is not available. Also, in order to make the figure more readable, the skipped period has been marked with the tilde sign (i.e. ~) in the figure. From Figure 2, it is clear that during the initial periods of early nineteenth century, the Saraiki language activities began with what we can today term as ‘Saraiki Language Corpus Development’. This was followed by developments in the field of tracing origins of the language called ‘Saraiki Language Evolution’. The remaining period was then dominated by activities comprising ‘Saraiki Language Social Science’ with sporadically scattered instances of ‘Saraiki Language MTS’ as well as ‘Saraiki Language Phonetics’ and ‘Saraiki Language Script’. It is also clear from Figure 2 that the latest period has seen parallel research instances in various categories dealing with Saraiki NLP.

It is also evident from Figure 2 that for the entire period under analysis, the maximum number of research instances have been found for the ‘Saraiki Language Evolution’ category followed respectively by ‘Saraiki Language Social Science’, ‘Saraiki Language Corpus Development’, ‘Saraiki Language Phonetics’, ‘Saraiki Language MTS’ and ‘Saraiki Language Script’. It is also notable that maximum number of core NLP activities have been found for the category of ‘Saraiki Language Phonetics’ which when compared with activities in the category of ‘Saraiki Language Script’, has been found to be almost three times of the later. For further clarity about the statistical values represented graphically in Figure 2, the data is also presented in tabular form in Table 2. The common legend of abbreviations used for Figure 2 and Table 2 includes ‘Saraiki Language Corpus Development’ represented by ‘C’, ‘Saraiki Language Evolution’ represented by ‘E’, ‘Saraiki Language MTS’ represented by ‘M’, ‘Saraiki Language Social Science’ represented by ‘O’, ‘Saraiki Language Phonetics’ represented by ‘P’ and ‘Saraiki Language Script’ represented by ‘S’, respectively.

9 Conclusion

To the best of author’s knowledge and study of the field, this is the first formal attempt to present a classified and annotated bibliography of the NLP tasks of the Saraiki language of Pakistan and India. This is also the first research activity comprising a period of 200 years and presenting a bibliography of more than 90 research works. Through the current work, the usage of terms ‘Saraiki Language NLP’ as well as ‘Saraiki NLP’

has also been made for the first time in research community. The resource-scarceness of Saraiki language is evident from the fact that the popular translator facility by Google provides translation to and from 103 languages.

These also include close relatives of Saraiki language like Pashto, Sindhi, Punjabi and Urdu, though Saraiki itself is not included. Even though not listed in any research paper so far, the current work advocates the possible comparison of dialects of Saraiki language too. These dialects can include Multani, Riasti, Jafri, Hindki, Thali and Bahawalpuri. There is also a possibility, which the research world has not seen yet, of comparison of Saraiki language with other closely related languages like Persian, Arabic, Balochi, Majhi, Pothohari and Dogri.

A collection of nearly 100 NLP based works of Saraiki language was done for a period of nearly 200 years. The collected data was subjected to thorough analysis and it was found that though not much NLP research work has been done for the Saraiki language, the research instances found in the process could be divided into six different categories corresponding to the Saraiki language's Corpus Development, Evolution, MTS, Phonetics, Script and Social Science. It is concluded based on the analysis of the current work that good number of attempts have been made to trace the origins of Saraiki language, study influence of other languages on Saraiki language, study comparison of Saraiki language with other languages and analyze the social impact of Saraiki language usage. Urdu, English, Punjabi, Sindhi, Russian and French languages have been found to have parallel corpus with the Saraiki language. Hence, on the side lines of these works, parallel corpus development, non-parallel corpus development as well as the study of rules of usage of Saraiki language too have been studied and presented. These in turn have proved to be useful in current times for typical Saraiki NLP activities like study of phonetics, study of nasality, MTS, POS as well as OCR. The current work strongly recommends and predicts the core Saraiki NLP activities dealing with development of Saraiki language resources and approaches for WordNet, Named Entity Recognition (NER), Coreference Resolution, Morphological Segmentation and Word Sense Disambiguation (WSD), to name a few.

Through the current research work, it is concluded that the area of Saraiki language phonetics has drawn maximum attention of researchers so far while MTS and OCR are the most upcoming areas. It is further concluded that in spite of various formal and informal attempts to research the Saraiki language, it still remains

a resource-poor language slowly proceeding towards decline. It is also concluded that the last decade has seen a remarkable increase in the Saraiki NLP activities, though.

The current work laments the lack of sufficient research activities for Saraiki NLP but it does not lament the lack of NLP activities of any other language. It does not advocate the inadequacy or pre-eminence of any one language over any other language. It is just a research work undertaken with an academic perspective to communicate the first results in the area of Saraiki language NLP activities. One of the objectives of the paper is to definitely promote and encourage the NLP research activities in Saraiki language. It is believed that the current work will have an excellent ripple effect in the domain of NLP research activities of languages of Pakistan, Afghanistan as well as North India. In the current work, best attempt has been made for presenting as well as including exhaustive coverage of the NLP activities of the Saraiki language. The current work is intended to be a stepping stone and one-stop resource for NLP research in the Saraiki language. This paper, being authored in English language unlike many previous papers in domain of Saraiki NLP, is also an attempt to bring the Saraiki language into main stream research and put it on the international platform for further exploration and research.

References

- [1] Abbasi A.W., Quran Kareem Ma tarjama Manzoom (Translated from Urdu: The Holy Quran Translation), Saraiki hand written manuscript
- [2] Adward, A Glossary of Multani Language, 1881
- [3] Akhter N., Saraiki Zarb-ul-amsaal aur Saraiki Vasaib: Aik Tehqeeqi-o-tanqeedi Jaiza (Translated from Urdu: Saraiki Proverbs and Saraiki Culture: A Research and Critical Survey), published by Multan's Saraiki Adabi Board, MPhil thesis accepted by Bahauddin Zakariya University, Multan
- [4] Alam M.F., Gill A., Pragmatic Transfer: An Interlanguage Study Of Pashto And Saraiki English Language Learners, Gomal University Journal of Research, Special Issue III, ISSN: 1019-8180, pages 143-151
- [5] Anonymous, Download Saraiki font and keyboard for Windows and Android, Available online: <http://saraiki.jimdo.com>
- [6] Anonymous, Saraiki Alphabet with Gurmukhi equivalents, Available online: http://www.user.unihannover.de/nhtcapri/wester_npanjabialphabet.html

- [7] Anonymous, Saraiki online transliteration, Available online: <http://www.sanlp.org/saraikiMT/saraikiMT.aspx>
- [8] Awan M.S., Baseer A., Sheeraz M., Outlining Saraiki Phonetics: A Comparative Study of Saraiki and English Sound System, *Language in India*, ISSN: 1930-2940, vol. 12, issue 7, July 2012; pages 120-136
- [9] Bahawalpuri Z., Saraiki Zuban ka Irteqa (Translated from Urdu: The Evolution of the Saraiki Language), Markaz-e-Saraiki Zuban-o-Adab, Bahawalpur, 1970
- [10] Bahl P., Étude de phonétique historique et expérimentale des consonnes injectives du Multani (Translated from French: A study of historical and experimental phonetics of injective consonants in Multani), MA thesis, A. Maisonneuve, Paris, 1936
- [11] Bakhsh H.F.E., Saraiki zabaan kee vajah e tasmiyyah (Translated from Urdu: The Name Story of the Saraiki Language), Daily Bashaarat, Muzaffargarh, December 1960
- [12] Bhaiya, Ahmad B., Saraiki Qavaid tey zubandani (Translated from Urdu: The Rules and Linguistics of the Saraiki Language), Saraiki Adbi Majlis, Bahawalpur, Pakistan, 1998
- [13] Brien O., A Glossary of the Multani Language, Punjab Government Press, 1988
- [14] Brien O., A Glossary of the Multani Languages, Revised and Re-arranged by Wilson, J and Kaul, Pandit Hari Kishen, Lahore, Pakistan, 1903
- [15] Brien O., Mughal S., Saraiki Urdu Dictionary, Jhoke Publishers, Multan, Pakistan, 2001
- [16] Farooq M., An Acoustic Phonetic Study of Six Accents of Urdu in Pakistan, a thesis accepted for M. Phil in Applied Linguistics, Department of English Language and Literature, School of Social Sciences and Humanities, University of Management and Technology, Lahore, Pakistan, 2014
- [17] Gardezi, Hassan N., Saraiki Language and its poetics: An Introduction, Sangat Publishers, London 1996
- [18] Gilani M.H., Historical Background of Saraiki Language, *Pakistan Journal of Social Sciences*, vol. 33, issue 1, 2013, pages 61-76
- [19] Gilani M.H., The Saraiki Translations of the Holy Quran with its Varieties, *Pakistan Journal of Social Sciences*, vol. 33, issue 2, 2013, pages 255-264
- [20] Haq A.M., Multani Zuban Aur Uska Urdu Se Taalluq (Translated from Urdu: Multani Language and its Relation with Urdu), Urdu Academy, Bahawalpur, 1967
- [21] Haq M.A., Saraiki Dian Mazeed Lisani Tehqeeqan (Translated from Urdu: More Insights into Saraiki Linguistics), Saraiki Adabi Board, Multan, 1985
- [22] Haq M.A., Saraiki Zaban dey Qaidey Qanoon (Translated from Urdu: The Grammatical Rules of Saraiki Language), Saraiki Adabi Board, Multan, 1984
- [23] Haq M.A., Saraiki aur us ki Hamsaya Ilaqai Zubanein (Translated from Urdu: The Saraiki Language and Its Neighboring Regional Languages), Saraiki Adbi Board, Multan, 1977
- [24] Hussain S.S., Farooq M., The Story of Language Conflicts in Pakistan: A Case Study of Saraiki and Punjabi Languages, *Pakistan Journal of Social Sciences*, vol. 35, issue 1, 2015, pages 287-295
- [25] Hussain S.S., Khan B.M., Integrated Linguistic Links of Pushto and Saraiki Languages, *Pakistan Journal of Social Sciences*, vol. 36, issue 2, 2016, pages 653-663
- [26] Hussain S.S., The Growth of Saraiki Language, *Pakistan Journal of Social Sciences*, vol. 36, issue 1, 2016, pages 387-396
- [27] Hussain Z., Asif S.I., Rehman R.A., Hanif M., Transferring Values to the Saraiki Farmers through Folk Tales, *Interdisciplinary Journal Of Contemporary Research In Business*, Institute of Interdisciplinary Business Research, vol. 4, issue 11, March 2013, pages 670-679
- [28] Hussain Z., Haneef M., Asif M., Naz S., Gender Construction in Saraiki Folktales: A Case Study from District Muzaffar Garh, *Applied Sciences and Business Economics*, ISSN: 2312-9832, vol. 2, issue 2, pages 8-15
- [29] Hussaini N.A.Z., Ma'aarif e Saraiki (Translated from Urdu: The Knowledge of Saraiki Language), Bahawalpur, 1972
- [30] Hussaini N.A.Z., Maarif Saraiki (Translated from Urdu: The Knowledge of Saraiki Language), Saraiki Abadi Board, Multan, 1972
- [31] Imtiaz A.S., Saraiki Language and Ethnic Identity, *Journal of Research*, Faculty of Languages and Islamic Studies, Bahauddin Zakariya University, Multan, Pakistan, vol. 7, 2005
- [32] Jampuri, Saraiki Poetry, Bazm-e-Saqafat, Multan, 1968
- [33] Jan M.T., Saleem Y., Optical Character Recognition (OCR) System For Saraiki Language Using Neural Networks, *University of Engineering and Technology Taxila. Technical Journal*, vol. 21, issue 3, September 2016, pages 106-109
- [34] Jumani N.B., Rehman R., Iqbal M.J., Chishti S., Effects of Native Language Saraiki on English Language Pronunciation, *International Journal of*

- Business and social Sciences, vol. 2, issue 8, 2011
- [35] Kalanchvi D., Saraiki Aur Uss ke Nasar (Translated from Urdu: The curriculum of the Saraiki Language), Academy Saraiki Adab, Bahawalpur, 1982
- [36] Kalanchvi D., Saraiki Lisaniat (Translated from Urdu: Saraiki Linguistics), Academy of Saraiki Literature, Bahawalpur, 2008
- [37] Kalanchvi D., Saraiki tarjameen wala Quran Shareef (Translated from Urdu: The Holy Book Quran in the Saraiki Language), Ibn-e-Kaleem publishers, Multan, 2000
- [38] Kalanchvi D., Saraiki Zaban tey Adab (Translated from Urdu: The Saraiki Language and Literature), Saraiki Adabi Majlis, Bahawalpur, 1987
- [39] Khan J. A., Saraiki Saqafat (Translated from Urdu: The Saraiki Culture), Roohani Art Press, Multan, 1995
- [40] Khokhlova L.V., Majority language death, Language Documentation & Conservation Special Publication No. 7, Language Endangerment and Preservation in South Asia, ed. by Cardoso H.C., January 2014, pages 19-45
- [41] Latif A., Phonemic Inventory of Siraiki Language and Acoustic Analysis of Voiced Implosives, Center for Language Engineering, Al-Khawarizmi Institute of Computer Science, University of Engineering and Technology, Lahore, Pakistan
- [42] Lewis P.M., Simons G.F., Fennig C.D., eds., Saraiki, Ethnologue, 19 ed., 2016, Available online: <http://www.ethnologue.com/language/skr>
- [43] Mahar M.A., Hussain S.S., Shah A., Diverse Mentions of Crane Bird in Saraiki Poetry, Pakistan Journal of Social Sciences, vol. 36, issue 1, 2016, pages 445-450
- [44] Malik M.G.A., Boitet C., Bhattacharyya P., Finite-state Scriptural Translation, published by Association for Computational Linguistics, ACM, USA in the proceedings of the 23rd International Conference on Computational Linguistics (COLING '10), Beijing, China, August 2010, pages 791-800
- [45] Molana S.M., Par Awal Mutarjam Ba muhawra Bazaboon-e-Saraiki (Translated from Urdu: The Translation of Idioms of the Farmers of Saraiki Language), Lahore Gilani Press, 1925
- [46] Monnier J.A., Glossary of the Multani language compared with Punjabi and Sindhi, printed at the Punjab Government Civil Secretariat Press, Lahore, 1881
- [47] Moran S., McCloy D., Wright R., eds., Saraiki sound inventory, Max Planck Institute for Evolutionary Anthropology, Leipzig, 2014
- [48] Mughal S., Saraiki diyaan khaas aavaazaan dee kahaanee (Translated from Punjabi: The story of Peculiar Sounds in Saraiki), Jhok Publishers, Multan, 2002
- [49] Mughal S., Saraiki Zuban De Hawaley (Translated from Punjabi: The Review Description of the Saraiki Language), (Hussain S.S., Intervieweur), February, 2010
- [50] Nasir N.K., Saraiki Shaire Daa Irtiqaa (Translated from Punjabi: The Evolution of the Sharayi style of Poetry of the Saraiki Language), Saraiki Adbi Board, Multan, 1977
- [51] Nasir S., Nasals and Nasality in Saraiki, Language in India, ISSN: 1930-2940, vol. 12, issue 7, July 2012; pages 323-347
- [52] Nasir S.A., Voice onset time of plosives in Saraiki: Implications for the acquisition of English aspiration contrast, ELF Annual Research Journal, Department of English Shah Abdul Latif University Khairpur, Sindh, vol. 14, 2012
- [53] Pandey A., Preliminary Proposal to Encode the Multani Script in ISO/IEC 10646, ISO/IEC JTC1/SC2/WG2 N4027, L2/11-124, Unicode, April 2011
- [54] Pandey A., Proposal to Encode the Multani Script in ISO/IEC 10646, ISO/IEC JTC1/SC2/WG2 N4159, L2/12-316, Unicode, September 2012
- [55] Parveen S., Anjum R.Y., Vocabulary Bank for English Text Books in Major Pakistani Languages, 2nd International Conference of the Linguistic Association of Pakistan organized at Institute of Languages, University of the Punjab, Lahore, 2016
- [56] Pervaiz S.H., Saraiki zabaan o adab kee muxtasar taareex (Translated from Urdu: A brief history of the Saraiki language and literature), National Language Authority, Islamabad, 2001
- [57] Qais F., Saraiki Zuban da Wewra (Translated from Urdu: The Description of the Saraiki Language), Nawa-e-Kaleem, Kot Mithan, 1993
- [58] Rahman T., The Siraiki language movement in Pakistan, Language Planning and Language Problems, vol. 19, issue 1, 1995
- [59] Rasoolpuri A., Saraiki Rasm-ul-Khat ki Mukhtasar Tareekh (Translated from Urdu: Brief History of the Saraiki Script), Multan: Bazm-e-Saqafat, 1976
- [60] Rasoolpuri M.A., Saraiki- Adab-Wich- Ma'aneeda- Pandh (Translated from Urdu: The Saraiki Literature) , Saraiki Publications, Rasoolpur, 2014
- [61] Rasoolpuri M.A., Saraiki zabaan ate lisaaniyaat (Translated from Urdu: Saraiki Language and Linguistics), Jhok Publishers, Multan, 2010

- [62] Rassoolpuri M.A., *Saraiki Zaban Unda Rasmulkhat Te Awazan* (Translated from Urdu: The Script and Sounds of the Saraiki Language), 1980, page 72
- [63] Rauf S., Hameed A., Habib T., Hussain S., *District names speech corpus for Pakistani Languages*, published by IEEE in the proceedings of Oriental COCOSDA held jointly with 2015 Conference on Asian Spoken Language Research and Evaluation (COCOSDA/CASLRE), Shanghai, China, October 2015, DOI: 10.1109/ICSDA.2015.7357893
- [64] Raza G., *Etmology of the Saraiki Language Name*, *Journal of Linguistics & Literature*, vol. 1, issue 1, 2016, pages 61-81
- [65] Raza G., *Prospects of Machine Translation for Pairs of Pakistani Languages*, 2nd International Conference of the Linguistic Association of Pakistan organized at Institute of Languages, University of the Punjab, Lahore, 2016
- [66] Raza G., *Reduction of Compound Adpositions in Persian, Urdu and Saraiki*, presented at the Sixth International Contrastive Linguistics Conference, Berlin, Germany, October, 2010
- [67] Sajjad H., *Saraiki Zuba-o-Adab ki Mukhtasir Tareekh* (Translated from Urdu: A Brief History of the Language and Literature of Saraiki Language), Muqtadra Qaumi Zuban, Islamabad, 2001
- [68] *Saraiki Dialect*, Wikipedia Inc., Available online: https://en.wikipedia.org/wiki/Saraiki_language
- [69] Shackle C., *Siraiki language*, *Encyclopedia Britannica*, 2014
- [70] Shackle C., *Siraiki, A language movement in Pakistan*, *Modern Asian Studies*, 11-3, 1977
- [71] Shackle C., *The Multani Marsiya, Der Islam*, *Journal of the History and Culture of the Middle East*, De Gruyter, vol. 55, issue 2, January 1978, pages 281-311, DOI:10.1515/islam.1978.55.2.281
- [72] Shackle C., *The Siraiki Language of Central Pakistan A Reference Grammar*, School of Oriental and African Studies, University of London, London, UK, 1976
- [73] Shoukat M., *Aao Saraiki Parhoon te Saraiki Likhon* (Translated from Hindi: Come to Read and Write Saraiki Language), Jhoke Publishers, Multan, 2004
- [74] Singh A., Kaur J., *Review for Dialects in Punjabi Language*, *International Journal of Innovations & Advancement in Computer Science*, ISSN: 2347-8616, vol. 5, issue 8, August 2016, pages 25-30
- [75] Syed N. A., *Nasalization in Saraiki*, MA thesis accepted by University of Essex, UK, 2009
- [76] Syed N.A., *Quantity-sensitivity and left-headedness: A study of Arabic loanword adaptations in Saraiki*, 2nd International Conference of the Linguistic Association of Pakistan organized at Institute of Languages, University of the Punjab, Lahore, 2016
- [77] Syed N.A.R., Malik R., Abbas M.Y.A., *Role of markedness and a dominant language in L1 attrition*, Available online: http://www.phonology.info/OCP/abstracts/Syed_OCP.pdf
- [78] Wagha A., *The Siraiki Language Its Growth and Development*, Derawar Publications, Islamabad, Pakistan 1990
- [79] Zami B.A., *Saraiki zabaan kaa irtiqaa* (Translated from Urdu: Evolution of the Saraiki language), Markaz Saraiki Zaban ka Adab, Bahawalpur, 1970
- [80] Khan A.H., Shami J.U., *The European Concept of the Siraiki Language and its Political Significance*, *The Journal of Political Science*, Department of Political Science, Government College, Lahore, Pakistan, 1996, vol. 19, issue 1 and 2, pages 97-105
- [81] Perveen R., *Mind your language: HIV communication in Pakistani languages*
- [82] Anonymous, *The Translation of the Holy Bible in Siraiki*, 1818
- [83] Skemp F.W., *Multani Stories: Collected and Translated by Skemp*, *Journal of the Royal Asiatic Society*, vol. 50, issue 3-4, October 1918, pages 626-628, published online March 2011, DOI: 10.1017/S0035869X00052205
- [84] Amritsari M., *Multani Proverbs, Sayings and Verses (Saraiki Muhaware)*, Available online: <http://gupshup.org/gs/cultureartsliteratureandlinguistics/612611multaniproverbssayingsversessaraimuhaware.html>
- [85] Khawaja Z.F., *Saraikipedia*, Available online: <http://saraikipedia.blogspot.in>
- [86] Shackle C., *Styles and themes in Siraiki Mystical Poetry of Sind*, *Bazme-Saqafat*, Multan, pages 5-6
- [87] Samarnov U.A., *The Lehnda Language* (originally in Russian; translated into English by Sipan E.H.T. and translated into Saraiki by Kalanchavi D., *Saraiki Bagho Bahrn*, Bahawalpur Academy, Siraiki Adab, 1995)
- [88] Shackle C., *A Century of Siraiki Studies* (originally in English; translated into Saraiki by Kalanchavi D., *Siraiki Mutalay de So Saal*, Bahawalpur Academy, Siraiki Adabi Majlis, Bahawalpur, 1986)
- [89] Khan H.A., *Siraiki Identity Construction in South Punjab*

- [90] Singh H., A Comparative Study of Majhi and Multani, PhD thesis submitted to and accepted by Punjab University, Chandigarh, December 1968
- [91] Shackle C., Siraiki and Siraiki Literature, c. 1750-1900, in Upper Sind and South-West Panjab, an unpublished Ph.D. thesis, School of Oriental and African Studies, University of London, 1972
- [92] Duong L.T., Natural Language Processing for Resource-Poor Languages, a PhD thesis accepted by The Department of Computing and Information System, The University of Melbourne, Melbourne, Australia, October 2017
- [93] Afsheen, Irtza S., Farooq M., Hussain S., Accent Classification among Punjabi, Urdu, Pashto, Saraiki and Sindhi Accents of Urdu Language, Center for Language Engineering, Al-Khwarizmi Institute of Computer Science, University of Engineering and Technology, Lahore, Pakistan, Available Online: http://www.cle.org.pk/Publication/papers/2014/clt14_submission_25.pdf, 2014
- [94] Farooq M., Hussain S., Acoustic Analysis of Corner Vowels in Six Indigneous Languages of Pakistan, Journal of Research in Social Sciences, ISSN: 2305- 6533, June, 2018 vol. 6, issue 2, pages 18-46, June 2018