ڊاڪٽر احسان احمد عرساڻي

سنڌي اکرن جي معياري رمز جي رجسٽريشن ۽ يُونيڪوڊ جو معياري استعمال

The Standardized Registration of Sindhi Letters and the Use of Standardized Unicode

Abstract

This research paper examines the challenges of developing and standardizing the Sindhi language in the current era of computer and information technology. Over the last fifty years, the world has witnessed rapid technological progress, setting unprecedented milestones in digital invention and innovation. The evolution of electronic information delivery systems, which continue to change at an accelerating pace, underscores the urgent need for the standardization of the Sindhi language. This standardization is crucial, as technological advancements have turned the world into a global village, placing Sindhi in direct competition with other global languages. To fully leverage modern technological resources, establishing a national standard is imperative. Without it, Sindhi risks falling behind in its development compared to other languages. Currently, like hundreds of other languages, Sindhi relies on Unicode encoding for the exchange of information and ideas via the Internet. While Sindhi has successfully entered the Information Age, inconsistencies between the standard Sindhi alphabet and Unicode have caused confusion. A key issue is that Unicode has allocated multiple code points for the letter "ھـ". As a result, different keyboards and fonts use varying Unicode characters for the same letter, leading to discrepancies. Words that appear identical on screen or in print are, in fact, stored differently in digital formats. Moreover, two essential letters in the standard

Sindhi alphabet— Jhae (جھ) and Ghaaf (گھ)—are entirely absent from Unicode, hindering proper usage and making it difficult to search for words in online dictionaries. This issue is also evident in the Sindhi Language Authority's online dictionary, accessible at https://dic.sindhila.edu.pk/.These inconsistencies will pose challenges for future language processing systems, such as Text-to-Speech systems for Sindhi, and will affect the creation of essential reference materials, such as thesauri. In this paper, I argue that, beyond standardizing Unicode usage for Sindhi, it is essential to develop a dedicated code for Sindhi characters and register it with both the International Organization for Standardization (ISO) and the Internet Assigned Numbers Authority (IANA). Achieving this, however, requires the active involvement of the Sindhi Language Authority.

Keywords: Sindhi language, technological advancements, Unicode, Standardization, language characters and letters, ISO

تت

پوين پنجاهہ سالن ۾ دنيا ٽيڪنالاجيءَ ۾ ترقيءَ Technological progress جو سفر تيزيءَ سان طئي ڪيو آهي. ڄاڻ جي ترسيل جا برقياتي طريقا جنهن تيزيءَ سان تبديليءَ هيٺ آهن، تنهن جي تقاضا آهي تہ پاڻ سنڌي زبان جي معيارسازيءَ لاءِ تڪڙا قدم کڻون. اهو ان ڪري بہ ضروري آهي جو ان تڪڙي ترقيءَ دنيا کي عالمي ڳوٺڙو بنائيندي سنڌي زبان کي سڄيءَ دنيا جي ٻولين سان سڌي چٽاڀيٽيءَ ۾ آڻي ڇڏيو آهي. جديد ٽيڪنيڪي وسيلن جو ڀرپور فائدو وٺڻ لاءِ گهربل معيارسازيءَ جي ضرورت آهي، جنهن کان سواءِ سنڌي زبان پنهنجي ترقيءَ جي رفتار کي قائم نہ رکي سگهندي ۽ ان ڊوڙ ۾ ٻين ٻولين کان پٺتي رهجي ويندي.

هن وقت دنيا جي سوَن ٻين ٻولين جيان سنڌيءَ ۾ بہ ٻاهم ڄار (Internet) ذريعي ڄاڻ ۽ خيالن جي ڏي وٺ يونيڪوڊ رمز تي آڌاريل آهي. جيتوڻيڪ ان سان سنڌي زبان اثرائتي نموني اطلاعاتي دور (Information Age) ۾ داخل ٿي چڪي آهي، پر معياري

سنڌي الف بي ۽ يونيڪوڊ ۾ سَو سيڪڙو مطابقت نہ هئڻ سبب ان رمز جي استعمال ۾ ڪي مونجهارا ٿي پيا آهن. انهن مونجهارن جو وڏي ۾ وڏو ذريعو يونيڪوڊ ۾ "ھ" اکر لاءِ مختص ٿيل هڪ ڪان وڌيڪ جايون هئڻ آهي. ڪمپيوٽر (Computer) تي سنڌي اکرن جي داخلا لاءِ استعمال ٿيندڙ مختلف ڪي بورڊ ۽ مختلف فانٽس ۾ "ھ" اکر لاءِ مختلف يونيڪوڊ حرف استعمال ٿيڻ سبب پردي يا ڪاغذ تي ساڳيا نظر ايندڙ لفظ ڪمپيوٽر ۾ باهم چار اندر مختلف يونيڪوڊ حرفن سان سانڍيل آهن. ان ڪان سواءِ يونيڪوڊ ۾ معياري سنڌي الف بي جا بہ اکر "جھ" ۽ "گھ" مڪمل طور غير موجود آهن. ان سان سنڌي زبان جي معياري استعمال ڪي ڌڪ لڳو آهي ۽ آن لائن لغتن ۾ لفظن جي ڳولا مشڪل بڻيل آهي. سنڌي ٻولي اختياري (Sindhi Language Authority) جي ويب پتي https://dic.sindhila.edu.pk/ تي موجود لغت ۾ پڻ اهڙا مونجهارا پيدا ٿيا آهن. ان ڪان سواءِ سنڌي ٻوليءَ لاءِ مستقبل ۾ جڙندڙ زبان پردازڪاري سرشتن Language Processing Systems جهڙوڪ Text-to-Speech Systems ۾ پڻ مونجهارا پيدا ٿيندا. ساڳيا مونجهارا سنڌي زبان ۾ جديد حوالاجاتي (Reference) مواد جهڙوڪ ٿيسارس وغيره جي تياريءَ ۾ پڻ پيش ايندا.

هن مقالي ۾ آئون اهو مقدمو پيش ڪندس تہ نہ رڳو يونيڪوڊ جي درست استعمال لاءِ معيار جوڙڻ جي ضرورت آهي، بلڪ سنڌي اکرن جي معياري رمز جوڙڻ ۽ ان ڪي معيارسازيءَ جي بين الاقوامي تنظيم ISO وٽ توڙي Internet Assigned Numbers Authority وٽ رجسٽر ڪرائڻ پڻ لازمي آهي.

تعارف

سنڌي زبان جي مختلف رسم الخطن يا لپين ۾ لکجندي رهي آهي [1]. تنهنڪري اڄ بہ ان جي معياري رسم الخط تي بحث هلندا رهن ٿا [2]. ان بحث جو هڪڙو وڏو سبب سنڌي قوم جو مذهب جي آڌار تي ورهاڱو ۽ هندستان ۾ سنڌي ٻوليءَ لاءِ عربيءَ ڪان سواءِ ديوناگري خط بہ اختيار ڪرڻ آهي. ان ڪان سواءِ عربي رسم الخط ۾ بہ سنڌي زبان جي ڪيترن لفظن جي تحريري صورت تي اهڙا بحث بہ ٿيل آهن

جيڪي معياري سنڌي الف بي کي تڪراري بڻائي ڇڏين ٿا. انهن مان ڪن مباحثن کي بدقسمتيءَ سان اجايو ڊگهو ڪيو ويو آهي [3]. انهن اجاين بحثن سان سنڌي ٻوليءَ جي معياري تحريري روايتن کي ڇيهو رسيو آهي ۽ نااتفاقي پيدا ٿي آهي.

پر موجوده دور ۾ واهپي ۾ ايندڙ ٻوليون ۽ علم و ادب جو ذريعو بڻيل ٻولين جو رڳو ڪاغذ تي ڇپجڻ ڪافي نہ آهي. هلندڙ دور ٻاهمجار Internet جو دور آهي، جنهن ۾ ٻوليون برقياتي ذريعن وسيلي پارمڪارجن transmit ٿيون ۽ برقياتي سنڀار ۾ سانڍجن Store ٿيون. بد قسمتيءَ سان اها سموري ٽيڪنيڪياسي ترقي سنڌ کان ٻاهر ٿي آهي، جنهن ۾ سنڌي سماج جي ڀاڱيداري نہ هئڻ برابر آهي. تنهنڪري ڪمپيوٽر ۽ انٽرنيٽ جي اوائلي دور ۾ سنڌي ٻوليءَ جو استعمال عدم موجود رهيو. نہ رڳو ايترو پر سنڌي ٻوليءَ کي رياستي سرپرستي نصيب نہ هئڻ سبب پڻ بين الاقوامي سطح تي نمائندگي ميسر ٿي نہ سگهي، ۽ بين الاقوامي ادارن وٽ سنڌي ٻوليءَ جي تحريري صورت جي معيارسازيءَ واري ڪم ۾ سنڌ سرڪار بہ ڪي جوڳا قدم نہ کنيا. ان سموري صورتحال ۾ سنڌي ٻوليءَ جي اهميت ۽ پاڪستان توڙي هندستان اندر ان جي استعمال کي نظر ۾ رکندي جڏهن عالمي سطح جي ادارن سنڌي ٻوليءَ جي معيارسازيءَ تي ڪم ڪيو تہ ڪي مونجهارا پيدا ٿي پيا. انهن مان هڪڙو مثال اکر "هہ" جو استعمال آهي. اهي مونجهارا ان ڪري پيدا ٿيا جو سرڪاري سطح تي سنڌي زبان جي پنيرائي نہ هئڻ سبب انهن ادارن وٽ سنڌي لسانياتي مهارتن جي جوڳي نمائندگي غير موجود هئي.

آءِ ايس او 8859 رمزون

معيارسازيءَ جي بين‌الاقوامي تنظيم آءِ ايس او ISO ۽ انٽرنيشنل اليڪٽروٽيڪنيڪل ڪميشن IEC جي گڏيل ٽيڪنيڪي ڪاميٽيءَ 8859 جي سجاٽپ نمبر سان ڪيترين ئي ٻولين لاءِ علامت رمزون رجسٽر ڪيون آهن [4]. انهن مان ڪي مثال جدول 1 ۾ ڏنل آهن.

جدول 1: آءِ ايس او 8859 هيٺ رجسٽر ٿيل علامت رمزون [4]

شمار نمبر	ٻوليون / رسم الخط	معيار	پهريون ڇپيرو	آخري ڇپيرو
1	لاطيني 1: الهندي يورپ جون ٻوليون	ISO-8859-1	1987	1998
2	سِرِلڪ	ISO-8859-5	1988	1999
3	عربي	ISO-8859-6	1987	1999
4	يُوناني	ISO-8859-7	1987	2003
5	ابراني	ISO-8859-8	1988	1999
6	ترڪي، مالٽيز ايسپرانتو (لاطيني)	ISO-8859-9	1989	1999
7	ٿاءِ (لاطيني)	ISO-8859-11	2001	2022
8	وچ يُوربي ٻوليون	ISO/IEC 8859-16	1998	2001

ڏسي سگهجي ٿو تہ انهن مان اڪثر 1980ع واري ڏهاڪي جي پڇاڙيءَ ۾ ئي رجسٽر ٿيا آهن ۽ تازي ترين رمز ٿاءِ زبان جي آهي، جنهن جو تازو ترين ڇپيرو Revision سن 2022ع ۾ ٿيو آهي. ان جدول ۾ ديوناگري موجود نہ آهي. ديوناگريءَ جي رمز ISO-8859 تحت رجسٽريشن جي مرحلي ۾ هئي تہ 1990ع ڌاري آءِ ايس او پنهنجي يونيورسل علامتي رمز 10646 شايع ڪرائي. تنهنڪري هندستاني سرڪار 1997ع ڌاري ISO-8859 تحت ديوناگري خط جي رجسٽريشن تان هٿ ڪڍي ورتو ۽ هنن هندستاني معيارسازيءَ هيٺ ٻہ معيار تيار ڪري انهن کي ISO-10646 جو حصو بڻارايو. انهن ٻن معيارن مان هڪڙي جو نالو (Indian Script Code for Information Interchange (ISCII [5] آهي تہ ٻئي جو نالو Perso-Arabic Script Code for Information Interchange [6] آهي. انڊين اسڪرپ ڪوڊ

۾ ديوناگري، بنگالي-آسامي، گجراتي، گرمکي، ڪنڙ، ملايالم، اوڙيا، تامل ۽ تيلگو رسم الخط جون نشانيون شامل آهن. جڏهن تہ پاسڪِي ۾ ڪشميري، سنڌي، اردو ۽ پشتو رسم الخط جون نشانيون شامل آهن. معيارسازيءَ جو اهو ڪم هندستان جي مرڪزي سرڪاري اداري Centre for Development of Advanced Computing (CDAC) جي نظرداريءَ هيٺ ٿيو.

يونيڪوڊ 14.0

ساڳئي ئي وقت جڏهن آءِ ايس او پنهنجي معيار ISO-10646 تحت ڪم ڪري رهيو هو ۽ يونيڪوڊ ڪنسا ٽيئم يونيڪوڊ تي ڪم ڪري رهيو هو. اڳتي هلي آءِ ايس او ۽ يونيڪوڊ ڪنسا ٽيئم گڏجي ڪم ڪيو ۽ نتيجتاً ISO-10646 جو 2021ع ۾ شايع ٿيل ۽ يونيڪوڊ 14.0 بلڪل هڪجهڙا آهن. نتيجتاً اڄ هندستان جون ڪيتريون ئي ٻوليون ISO-10646 يعني يونيڪوڊ وسيلي ويب تي لکي پڙهي سگهجن ٿيون. هن وقت انٽرنيٽ تي جاڻ جي ڏي وٺ جو تمام وڏو ڀاڱو يونيڪوڊ وسيلي ئي عمل ۾ اچي ٿو. يُونيڪوڊ ۾ هر سنڌي اکر ٻن بائٽس Bytes جي جاءِ والاري ٿو [8-7]. جيڪا انگريزي اکرن جي مقابلي ۾ ٻيڻي آهي. يونيڪوڊ ۾ سنڌي ڏار ٻوليءَ جي جاءِ نہ ٿي رکي. سنڌي صورتخطيءَ جا اهي اکر جيڪي عربيءَ ۾ نہ ٿا اچن، تن کي يونيڪوڊ اندر عربي زبان جي اضافت Extension طور جاءِ ڏنل آهي.

معياري سنڌي اکر ميڙ

سنڌي-عربي رسم الخط يعني هاٽوڪي سنڌي الف بي جو معياري صورت ۾ استعمال تمام مضبوط روايت رکي ٿو. برطانوي راڄ کان وٺي هن مهل تائين ان ۾ تمام گهڻو مواد شايع ٿي چڪو آهي ۽ هن وقت بہ سنڌي زبان تيزيءَ سان وڌندڙ آهي ۽ سنڌي زبان ۾ ڪيترو ئي تحريري مواد شايع ٿي رهيو آهي. ان ۾ 52 اکرن کان سواءِ چند ٻيون نشانيون پڻ شامل آهن، جهڙوڪ "آ"، "ۓ"، "۾". ساڳيءَ ريت سرڪاري سرپرستيءَ ۾ پڻ سنڌي زبان جي تدريس جي ڏيڍ سَو سال ڊگهي تاريخ آهي. ان جي هِجي جا اصول

باضابطہ طور جوڙيل ڪِڪِپٽيءَ تي آڌاريل آهن، جنهن جي سکيا نہ رڳو درسي ڪتابن ۾ ڏني پئي وئي آهي، بلڪ ان جي پنيان تدريسي عملي لاءِ تربيتي ادارا پڻ ڪارفرما رهندا آيا آهن. پوين پنجاهہ ورهين ۾ مجموعي طور تعليمي سرشتي جي ڪمزور ڪارڪردگي ۽ استادن جي تربيتي ادارن جي غير متحرڪ رهڻ سبب لساني ميدان ۾ معيارسازيءَ واري ڪم کي ڪاپاري ڌڪ لڳو آهي، ۽ ڪِڪِپٽيءَ کي آڌار بنائيندي هِجي جا اصول اختيار ڪرڻ ۾ ڪوتاهي ٿيڻ سبب سنڌي تحرير ۾ مونجهارا حل ٿي نہ سگهيا آهن. پر هن مضمون جو مقصد انهن سمورن مونجهارن جو ذڪر ڪرڻ نہ پر يونيڪوڊ وسيلي سنڌي زبان جي ترسيل ۾ پيدا ٿيل مونجهارن جو ذڪر ڪرڻ ۽ انهن جو حل تجويز ڪرڻ آهي.

مسئلا

جيتوڻيڪ هن وقت سنڌي زبان باهمجار Internet تي يونيوڪوڊ ڪنسائٽيم جي UTF-8 ۽ UTF-16 نالي رمزسازين Encodings ذريعي اطلاعت Information جي ڏي وٺ لاءِ گهڻي ڀاڱي آسانيءَ سان استعمال ٿي رهي آهي، تنهن هوندي بہ ڊجيٽل سنڌي ٽيڪسٽ ۾ معيارسازيءَ جي کوٽ ۽ انيڪ موجهارن [9] سبب لغت سازي، ٻولا جي آساني، خودڪار ترجمي، ۽ معنوي ويب Semantic Web لاءِ سنڌي زبان اجا تيار ٿي نہ سگهي آهي. انهن مونجهارن جو وڏي ۾ وڏو سبب يونيڪوڊ ۾ ٻن سنڌي اکرن ”جهہ“ ۽ ”گهہ“ جي عدم موجودگي، ۽ انهن جي پورائي لاءِ اکر ”ھ“ جو استعمال آهي. يونيوڪوڊ جو استعمال ڪندي سنڌي زبان جي ترسيل لاءِ ڪيترن ئي هنڌن تي ۽ ڪيترن ئي مختلف مقصدن سان ”ھ“ اکر جو استعمال ڪيو ويو آهي. پاڻ ان اکر جي مصرفن کي سمجهنداسين ۽ ان ۾ پيدا ٿيل مونجهارن کي سمجهنداسين. جيئن تہ يونيڪوڊ ۾ سنڌيءَ کان سواءِ عربي خط ۾ لکجندڙ ڪيترين ئي ٻين ٻولين جا اکر ۽ نشانيون پڻ شامل آهن، تنهنڪري ان ۾ هڪ کان وڌيڪ جائيون شامل آهن جن جو مقصد مختلف ٻولين ۾ متفرق ”ھ“ اکرن کي ظاهر ڪرڻ آهي. يونيڪوڊ معيار 16.0 ۾ شامل عربي اکرن جي رينج [10] تان ورتل جدول 2 ڏيکاري ٿي تہ

يونيڪوڊ ڪنساٽيئيم طرفان "ھ" جي مختلف صورتن لاءِ ڪل 7 ڪوڊ پوائينٽس تجويز ڪيل آهن. ياد رهي تہ انهن مان هر ڪنهن جي اڳين، وچين، ۽ پڇاڙڪي صورت لاءِ ڌار ڪوڊ مختص آهن. جيئن [11] ۾ سمجهايل آهي.

جدول 2: يونيوڪوڊ معيار 16.0 [10] ۾ "ھ" جي مختلف صورتن لاءِ مختص ٿيل ڪوڊ پوائينٽس

شمار نمبر	ڪوڊ پوائنٽ	پردي تي ظاهري صورت
1	0647	ه (اکر ھ)
2	06BE	ھ (دوچشمي ھ)
3	06C0	ۀ (عربي ھ مٿان ي)
4	06C1	ہ (گول ھ)
5	06C2	ۂ (عربي ھ مٿان همزو)
6	06D5	ە (ھ جهڙو ڏسجندڙ عربي آي)
7	06FF	ۿ (عربي اکر ھ ٽوپيءَ ساڻ)

ڪمپيوٽر تي سنڌي ٽيڪسٽ جي داخلا ۾ "ھ" جو اکر ڇهن جايُن تي استعمال ۾ اچي ٿو: i) لفظن جي شروع ۾ ۽ وچ ۾ ھ جو اکر داخل ڪرڻ لاءِ، ii) زنانہ نالن جي پڇاڙيءَ ۾، iii) هڪ اکرا لفظ، iv) ڪن صفت لفظن جي پڇاڙيءَ ۾، v) "جهہ"، "گهہ"، ۽ "ھ" سان ختم ٿيندڙ لفظن جي پڇاڙيءَ ۾ ۽ vi) سنڌي اکر "جه" ۽ "گه" لکڻ لاءِ. انهن ڇهن ئي مصرفن سان جيڪي مسئلا لاحق آهن، تن تي بحث هيٺ ڪجي ٿو.

پهرين مصرف جا مثال ”هڏو“ ۽ ”هاٿي“ جهڙا لفظ آهن. ان ۾ جيتوڻيڪ پردي تي صورت اڪثر فانٽس ۾ ساڳي رهي ٿي، پر اهو عين ممڪن آهي ته مختلف فانٽ جوڙيندڙ يونيڪوڊ جا مختلف ڪوڊ پوائينٽس استعمال ڪن. تنهنڪري متفرق فانٽس ۾ جيتوڻيڪ پردي تي ”هڏو“ ۽ ”هاٿي“ جهڙا لفظ پلي هڪجهڙا نظر اچن، پر ڪمپيوٽر ۾ انهن جي نمائندگي يونيڪوڊ جي مختلف ڪوڊ پوائنٽس سان ٿيل ٿي سگهي ٿي؛ ڇاڪاڻ ته هر منطقگر Software Developer ۽ فانٽ ڊيولپر مختلف ڪوڊ پوائينٽس استعمال ڪري ٿو.

ٻين مصرف جا مثال ”رضيہ“، ”رخسانہ“، ۽ ”راحيلہ“ جهڙا زنانہ نالا آهن. ڪمپيوٽر ۽ انٽرنيٽ تي پاڻ کي اهي نالا ”رضيہ“، ”رخسانہ“، ۽ ”راحيلہ“ جي صورت ۾ بہ نظر ايندا آهن. ماضيءَ ۾ انهن نالن کي ”رضيہ“، ”رخسانہ“، ۽ ”راحيلہ“ ئي لکبو هو جڏهن تہ ڪي ماڻهو انهن کي ”رضيا“، ”رخسانا“، ۽ ”راحيلا“ ڪري بہ لکندا هئا.

ٽين مصرف جا مثال سنڌي لفظ ”ٻِہ“، ”ٻَہ“، ”ٽِہ“، ”ٽَہ“، "نَہ" آهن. جيڪي هڪڙي ئي اکر تي ٻڌل آهن. انهن کي ڪمپيوٽر جي سنيار ۾ ساڳين اکرن جي اڪيلين صورتن مطلب تہ ”ب“، ”ٻ“، ”ت“، ۽ ”ن“ کان ڌار نمائندگي ملڻ گهرجي، تہ جيئن انهن کي پردي تي بہ ڌار طريقي سان نمائي سگهجي. انهن کي بہ ٽائيپ ڪرڻ لاءِ بہ انهن جي پڇاڙيءَ ۾ جدول 2 ۾ ڏنل ”ه“ جي ڇهن ڪوڊپوائينٽس مان ڪا هڪ ڪوڊ پوائينٽ استعمال ٿئي ٿي. فانٽ تي دارومدار رکي ٿو تہ اهي پاڻ کي پردي تي ڪيئن ٿا نظر اچن. تنهنڪري اهي پاڻ کي ڪڏهن ڪڏهن ”ٻِہ“، ”ٻَہ“، ”ٽِہ“، ”ٽَہ“، ”نَہ“ بہ نظر ايندا آهن.

چوٿين مصرف جا مثال ”مردانہ“ ۽ ”زنانہ“، ”ظالمانہ“ وغيره آهن. انهن کي بہ خالص ڪاغذ جي دور ۾ ”مردانہ“، ”زنانہ“، ۽ ”ظالمانہ“ ئي ڪري لکبو هو. ٻنهي صورتن ۾ ڪمپيوٽر تي ٽائيپ ڪندي انهن لفظن جي پڇاڙيءَ ۾ جدول 2 ۾ ڏنل ”ه“ جي ڪا نہ ڪا ڪوڊ پوائينٽ لڳائجي ٿي.

”ه“، ”جه“، يا ”گه“ تي ختم ٿيندڙ سنڌي لفظن جهڙوڪ ”پرواه“، ”شاه“، ”گاه“، ”راه“، ”سمجه“، ۽ ”سگه“ جي پڇاڙيءَ ۾ بہ ”ه“ جي ڪا نہ ڪا ڪوڊ پوائينٽ

استعمال ٿئي ٿي. ڪڏهن ڪڏهن اهڙا لفظ پردي تي ”پرواھ“، ”ساھ“، ”شاھ“، ”گاھ“، ”راھ“، ”سمجھ“، ۽ ”سگھ“ جيان نظر ايندا آهن. ٻہ ٻہ ”ھ“ ڏسندي، ڪيترا ٽائيپسٽ انهن لفظن ۾ ”پرواه“، ”ساه“، ”شاه“، ”گاه“، ۽ ”راه“ جي صورت ڏيڻ ئي پسند ڪندا آهن. تنهنڪري اهڙا لفظ پردي ۽ ڪاغذ تي مختلف صورتن ۾ ظاهر ٿين ٿا ۽ ڪمپيوٽر جي سنڀار Memory ۾ بہ مختلف ڪوڊ سان سانڍيل ٿين ٿا.

ڇهين مصرف جا مثال اهي لفظ آهن جن ۾ ”جه“ يا ”گه“ جا اکر اچن ٿا، جهڙوڪ ”جهرڪ“، ”جهونو“، ”گهرو“ وغيره. ان جو سبب اهو آهي جو ٽوٽي جو اهي بہ اکر معياري سنڌي الف بي ڌار اکرن جي حيثيت رکن ٿا، يونيڪوڊ ۾ انهن کي پنهنجي تختيءَ ۾ شامل نہ ڪيو آهي. ان جو سبب شايد اهو ئي آهي جو جڏهن يُونيڪوڊ تي ڪم هلي رهيو هو، تڏهن سنڌي زبان جي معياري رمز ڪٿي بہ رجسٽر ٿيل نہ هئي. ان وقت ڪيترين ئي ٻولين جون معياري رمزون نہ رڳو معيارسازيءَ جي بين الاقوامي اداري ISO وٽ رجسٽر ٿيل هيون، بلڪ انهن کي ڪيترا ئي منطقگر Software Developers پنهنجي ويب برائوزرس يا ٻي منطقگريءَ Software ۾ شامل ڪري چڪا هئا. اڄ بہ ڪيترا ئي انٽرنيٽ برائوزر آءِ ايس او 8859 معيار هيٺ رجسٽر ٿيل رمزُن کي هٿي وٺائن ٿا. سنڌي زبان جي غير عربي فارسي اکرن جهڙوڪ ٻ، ڀ، ٺ، ٽ، ٿ، ڇ، ڃ، ڄ، ڏ، ڳ، ڱ، ڙ ڻ، وغيره کي يونيڪوڊ ۾ ان ڪري جاءِ ملي وئي جو اهي اکر هندستان سرڪار وٽ رجسٽر ٿيل پرسو عربڪ اسڪرپٽ فار انفارميشن انٽرچينج (پاسڪِي) PASCII نالي رمز ۾ شامل هئا [7]. جيئن تہ پاسڪِي ۾ بہ ”جه“ ۽ ”گه“ جا اکر شامل نہ هئا تنهنڪري يونيڪوڊ ۾ بہ انهن کي جاءِ ملي نہ سگهي. ياد رهي تہ اردو لپيءَ ۾ ڪنهن بہ وسرڳ آواز لاءِ ڌار اکر موجود نہ آهي. پاسڪِي رمز تيار ڪندي بہ هندستاني ماهرن آڏو اردوءَ جي حيثيت حاوي رهي، ۽ هنن ”جه“ ۽ ”گه“ جي ظاهري صورت کي ڏسندي انهن کي ڌار اکر واري حيثيت ڏئي نہ سگهيا. نتيجتاً، يونيڪوڊ ۾ بہ ”جه“ ۽ ”گه“ ڌار اکرن جي حيثيت ۾ شامل نہ آهن.

يُونيڪوڊ ۾ ”جه“ ۽ ”گه“ جي غير موجودگيءَ سبب، يونيڪوڊ کي استعمال ڪندي سنڌي زبان کي ڪمپيوٽر ۾ لکندڙ فرد مجبوراً ترتيبوار ”ج“ ۽ ”گ“ جي پٺيان ”ه“ جو اکر لڳائين ٿا. تاهم فانٽس سنڌي ٻولي اختياريءَ طرفان فانٽس ڊيولپرس يا ٻي منطقگري (Software) جوڙيندڙن لاءِ مخصوص هدايتون جاري نہ ڪيل آهن تہ سنڌيءَ ۾ ڪهڙي استعمال لاءِ ڪهڙي ڪوڊ پوائينٽ استعمال ڪرڻي آهي. تنهنڪري مختلف منطقگر انهن مان مختلف ڪوڊن جو استعمال ڪندا رهيا آهن. نتيجتاً معيار طي نہ ٿي سگهيو آهي ۽ سنڌي ٻوليءَ لاءِ جوڙيل مختلف ايپس Apps ۾ ڪمپيٽيبلٽيءَ Compatibility جا مسئلا پيدا ٿي پيا آهن.

هر زبان وانگي سنڌيءَ ۾ ڪي آواز ڳالهائجن ٿا، جن لاءِ ڏار اکر موجود نہ آهي. انهن ۾ وسرڳ آواز شامل آهن، جهڙوڪ ڙهہ، لهہ، مهہ، نهہ، ٽهہ [12]. انهن آوازن جو تحريري اظهار ٻن اکرن کي ملائي ڪرڻ آسان هو. تنهنڪري انهن لاءِ ڏار اکر يا نشاني رکڻ ضروري نہ هو ۽ معياري لپيءَ ۾ شامل نہ ڪيو ويو. ياد رهي تہ انهن جي ابتڙ سنڌي صورتخطيءَ ۾ ڪيترن ئي وسرڳ آوازن لاءِ ڏار اکر شامل آهن، جهڙوڪ ڀ، ٿ، ٺ، جه، ک، گه، ۽ ق. انهن جي شامل هجڻ جو يقيناً علمي سبب موجود آهي. ڪن سنڌي ماڻهن ڏيڍ سَو ورهين کان معيار بڻيل سنڌي الف بي کي نظر انداز ڪندي نہ رڳو ”ڪلهہ“، ”ملهہ“، ۽ ”سنهو“ جهڙن لفظن کي ”ڪلهو“، ”مله“، ۽ ”سنهو“ ڪري لکيو آهي ۽ بلڪ ان غلط هِجي لاءِ مُهمجوئي پڻ ڪئي آهي [11]. ان غلط مُهمجوئيءَ جا نہ رڳو منفي اثر پيا آهن بلڪ يونيڪوڊ ڪنسائيٽم وٽ ان سان مونجهارو وڌيو آهي، جنهن سبب جها ”جه“ ۽ گهاف ”گه“ جي اکرن لاءِ ڏار ڪوڊ پوائينٽس تفويض ٿي نہ سگهيون آهن. جيتوڻيڪ SIL International جي طرفان جي طرفان Lorna Priest Evans يونيڪوڊ ڪنسائيٽم جو ڌيان ”هہ“ جي مختلف ڪوڊ پوائينٽس سبب پيدا ٿيندڙ مونجهاري ڏانهن ڇڪايو آهي [13]، پر هن ان معاملي کي اهو چئي ويتر منجهائي ڇڏيو آهي تہ سنڌيءَ ۾ جهاءِ ”جه“ ۽ گهاف ”گه“ جا ڏار اکر موجود نہ آهن. ان جو

ٻيو مثال نسبتاً تازو شايع ٿيل مضمون آهي [14]. جنهن ۾ دعويٰ ٿيل آهي تہ سنڌيءَ ۾ "جه" ۽ "گه" جا ڏار اکر موجود نہ آهن. ياد رهي تہ جيتوڻيڪ سنڌي ٻوليءَ ۾ "ڙه"، "له"، "مه"، "نه"، ۽ "ڻه" جا آواز ضرور موجود آهن. پر معياري سنڌي لپيءَ ۾ انهن لاءِ ڏار اکر مختص نہ آهن. تنهنڪري انهن کي "ڙ"، "ل"، "م"، ۽ "ن" جي پٺيان سادي "ھ" لڳائي ئي لکڻ گھرجي. ايئن دنيا جي ڪيترين ئي لپين ۾ ٿيندو آهي. جھڙوڪ انگريزي ٻوليءَ ۾ "چ"، "ش"، "ڱ"، ۽ "ژ" جا آواز موجود آهن. پر انگريزي معياري الفابيٽ ۾ انهن لاءِ ڪو ڏار اکر موجود نہ آهي [15].

حل ۽ نتيجا

مٿي نشاندهي ڪيل مسئلن جو حل آسان آهي. سنڌي زبان جي ڏيڍ صدي پراڻي معياري الف بي کي آڌار بڻائيندي، سنڌي ٻولي اختياريءَ کي سنڌي زبان جي معياري رمز تيار ڪري پڌري ڪرڻي پوندي ۽ پهرين مرحلي طور ان کي بحيثيت معيار جي، پاڪستان جي معيارسازيءَ واري قومي اداري Pakistan Standards and Quality Control (PSQCA) Authority وٽ ان کي رجسٽر ڪرائڻو پوندو. ڇاڪاڻ تہ معيارسازيءَ جي بين الاقوامي اداري ISO وٽ پاڪستان جي نمائندگي PSQCA جو ادارو ڪري ٿو. تنهن کان پوءِ ساڳي معيار کي معيارسازيءَ جي عالمي اداري ISO وٽ رجسٽر ڪرائيندي ان کي ISO-10646 جو حصو بڻائڻو پوندو. ان کان سواءِ سنڌي زبان جي معياري رمز کي وٽ بہ رجسٽر ڪرائڻو پوندو جيئن تاملي ٻوليءَ جي معياري رمز TSCII رجسٽر ٿيل آهي [16]. تہ جيئن باھمچار Internet جي MIME معيار ذريعي سنڌيءَ ۾ اِي ميلز ڪري سگھجن. ان سان سنڌي اکرن "جه" ۽ "گه" کي ڏار ڪوڊپوائينٽس ملي ويندا.

جدول 3 سنڌي زبان لاءِ مجوزه رمز تختي ڏيکاري ٿي، جنهن کي سنڌي ٻولي اختياري صلاح مشوري کان پوءِ پاڪستان جي معيارسازيءَ واري اداري ڏانهن رجسٽر ڪرڻ لاءِ اماڻي سگھجي ٿي.

جدول 3: سنڌي زبان لاءِ مجوزه رمز تختي

				B8	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
				B7	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
				B6	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
				B5	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
B4	B3	B2	B1		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	0	0	0			SP	0	@	P	`	p			NBSP	ث	ر	گ		٠
0	0	0	1	1			!	1	A	Q	A	q			ٖ	پ	ڙ	ڳ		١
0	0	1	0	2			"	2	B	R	b	r			َ	ج	ز	گھ	ژ	٢
0	0	1	1	3			#	3	C	S	C	s			ِ	ڄ	س	ڱ		٣
0	1	0	0	4			$	4	D	T	d	t			ُ	جھ	ش	ل		٤
0	1	0	1	5			%	5	E	U	E	u			ً	ڃ	ص	م	۾	٥
0	1	1	0	6			&	6	F	V	F	v			ٰ	چ	ض	ن	ں	٦
0	1	1	1	7			'	7	G	W	g	w			آ	ڇ	ط	ڻ		٧
1	0	0	0	8			)	8	H	X	h	x			ا	ح	ظ	و		٨
1	0	0	1	9			(	9	I	Y	i	y			ب	خ	ع	ھ		٩
1	0	1	0	10			*	:	J	Z	j	z			ٻ	د	غ	ء	۽	×
1	0	1	1	11			+	;	K	[	k	}			ڀ	ڌ	ف	ي	ئي	؛
1	1	0	0	12			،	>	L	\	l	|			ت	ڏ	ڦ		ڤ	،
1	1	0	1	13			_	=	M	]	m	{			ٿ	ڊ	ق			
1	1	1	0	14			.	<	N	^	n	~			ٽ	ڍ	ڪ			
1	1	1	1	15			/	؟	O	_	o				ٺ	ذ	ک			

جدول 3 ۾ ڏيکاريل اڌ دم Comma ۽ واڪ جي نشاني Semicolon سنڌيءَ وارا نہ پر عربيءَ وارا آهن. سنڌي فانٽس جوڙيندڙن هميشہ سنڌيءَ بجاءِ عربي نشانيون ئي استعمال ڪندا اچن. ياد رهي تہ SIL International جي طرفان يونيڪوڊ ٽيڪنيڪل ڪاميٽيءَ وٽ اهڙي عرضي جمع ٿيل آهي تہ سنڌي اڌ دم ۽ واڪ جي نشانيءَ لاءِ ڪوڊ پوائينٽس مختص ڪيا وڃن [17].

تنهن کان سواءِ سنڌي ٻولي اختياريءَ کي عورتاڻن نالن جي معياري لکت لاءِ بہ سفارشون تيار ڪري عام ڪرڻيون پونديون تہ جيئن انهن نالن جي داخلا NADRA جي قومي اعدادخاني National Database ۾ بہ معياري صورت اختيار ڪري سگهي تہ ٻي سموري لکپڙهہ ۾ بہ. منهنجي راءِ ۾ زنانہ نالن، "مردانہ"، "ظالمانہ" جهڙن صفت لفظن، ۽ "وغيره" جهڙن عربيءَ کان اڌاريل لفظن جي پڇاڙيءَ ۾ "هہ" جي ساڳي ئي ڪوڊ پوائينٽ يعني 06C1 استعمال ٿيڻ گهرجي، جيڪا گول "ه" آهي.

هڪ اکرا سنڌي لفظ جهڙوڪ "ٻہ"، "ڀہ"، تہ، ۽ "نہ" اصولاً "هہ" جو اکر نہ ٿا رکن. تنهنڪري انهن جي لاءِ مختلف طريقا اختيار ڪري سگهجن ٿا. منهنجي راءِ ۾ اهڙن لفظن ۾ "هہ" جي ڪا بہ ڪوڊپوائينٽ استعمال نہ ڪرڻ گهرجي. اهڙن لفظن کي لکڻ لاءِ ب، ڀ، ت، ۽ ن جي پنيان يونيڪوڊ جي ڪوڊپوائينٽ 200C، ۽ ان پنيان 0020 يعني عام وٿيءَ SPACE جي ڪوڊ پوائينٽ استعمال ڪرڻ گهرجي. ڪوڊپوائينٽ 200C دراصل Zero Width Non-Joiner (ZWNJ) چوائيندو آهي، ۽ ڪنهن بہ سنڌي ڪِي بورڊ جو لازمي جُز آهي. اهو تڏهن استعمال ٿيندو آهي جڏهن پاڻ "شمس الدين" جهڙن نالن ۾ "س" کي "الف" سان ملائڻ نہ چاهيندا آهيون. مطلب تہ پاڻ چاهيندا آهيون تہ "س" جو اکر پنهنجي پڇاڙيءَ واري صورت ۾ ۽ الف جو اکر پنهنجي اڪيلي صورت ۾ ئي پردي تي ظاهر ٿئي. ساڳيءَ ريت "جه"، "گه"، ۽ "هہ" جي اکرن تي ختم ٿيندڙ لفظن جهڙوڪ "سمجهہ"، "سگهہ"، ۽ "شاهہ" جي پنيان بہ ساڳي ئي ڪوڊ پوائينٽ 200C يعني ZWNJ، ۽ ان پنيان عام وٿيءَ واري ڪوڊپوائينٽ 0020 لڳائڻ گهرجي تہ جيئن اهڙن لفظن جي پڇ ۾ لاڳاپيل اکرن جي پڇاڙيءَ واري صورت يعني "جهہ"، "گهہ" ۽ "هہ" کي ظاهر ڪري سگهجي. اهڙي معيارسازيءَ جي نتيجي ۾ سنڌي ٽيڪسٽ کي

برقياتي ذريعن ۾ سانڍڻ Store، پارمڪاري transmission، ۽ پردازڪاري Processing ۾ سهنجائي پيدا ٿيندي ۽ سنڌي زبان لاءِ مصنوعي ذهانت واريون ايپس سرجڻ جي راهہ هموار ٿيندي. ان کان سنڌيءَ ۾ برقياتي لغتون ٺاهڻ ۽ موجوده لغتن کي وڌائن توڙي انهن جي استعمال ۾ آساني پيدا ٿيندي.

ماحاصل Conclusion

سنڌي ٻوليءَ جو بااختيار ادارو هڪڙو حتمي اکر ميڙ Character Set مشتهر ڪري ۽ آءِ ايس او وٽ معياري رمز Standard Code طور رجسٽر ڪرائڻ جي ضرورت آهي تہ جيئن سنڌي ٻوليءَ کي پنهنجي منطقگري (Software) ۾ شامل ڪندڙ هر منطقگر Software Developer ان معيار جي پاسداري ڪرڻ جو پابند ٿي سگهي. جيتوڻيڪ هن وقت ڪمپيوٽر ۽ انٽرنيٽ تي سنڌي زبان ۾ ڄاڻ جي ڏي وٺ تواتر سان ٿي رهي آهي، ان هوندي بہ اڃا تائين معيارسازيءَ جا ڪيترا ئي مونجهارا موجود آهن، جن سبب پردي تي ۽ ڪاغذ مٿان سنڌي ٽيڪسٽ جي نماوَ ۾ اهڙيون صورتون ظاهر ٿينديون رهن ٿيون جيڪي معياري صورتن کان هٽيل آهن، يا جن تي نااتفاقي رهندي اچي. انهن مونجهارن کي ختم ڪرڻ لاءِ سنڌي اکرن جي معياري رمز جي رجسٽريشن ۽ يُونيڪوڊ جي معياري استعمال لاءِ سفارشون مرتب ڪرڻ جي ضرورت آهي. اهي ٻئي قدم صرف سنڌي ٻولي اختياري ئي کڻي سگهي ٿي.

حوالا

1. Iyengar, A. (2021). A diachronic analysis of Sindhi multiscriptality. *Journal of Historical Sociolinguistics*, *7*(2), 207-241.
2. Wadhwa, S. (2023). The question of script for Sindhi in India: reflections on postcolonial grammatology. *Interventions*, 1-16.
3. Kumbhar, Shabeer. (2019). Orthographic complexities and Pre-requisites in Sindhi composing. Sindhi Boli, 12(1), 83-102.
4. Comité Européen de Normalisation and Europäisches Komitee für Normung. "Guide to the use of character set standards in Europe." (1999).

5. Indian Standard - Indian Script Code for Information Interchange – ISCII Bureau of Indian Standards Manak Bhavan, 9 Bahadur Shah Zafar Marg New Delhi 110 002, https://www.services.bis.gov.in/php/BIS_2.0/bisconnect/knowyourstandards/Indian_standards/isdetails_mnd/4722
6. Perso-Arabic Script Code for Information Interchange, Centre for Development of Advanced Computing, http://parc.cdac.in/PASCII_V10.pdf
7. Unicode Consortium (2011). The Unicode Standard-Version 6.0–Core Specification-. Unicode Consortium.
8. Keiser, J., & Lemire, D. (2021). Validating UTF-8 in less than one instruction per byte. *Software: Practice and Experience*, *51*(5), 950-964.
9. Iyengar, A. (2018). Variation in Perso-Arabic and Devanāgarī Sindhī orthographies: An overview. *Written Language & Literacy*, *21*(2), 169-197.
10. Unicode Standard ver 16.0: Arabic Range: 0600–06FF. https://www.unicode.org/charts/PDF/U0600.pdf
11. Kumbhar, Shabeer. (2018). Limitations of Usage of Aspirated Letters and their Solutions. Sindhi Boli, 11(1), 19-42.
12. جتوئي، علي نواز حاجن خان، (1983) سنڌي زبان ۽ علم صوتيات، (ڇاپو پيون)، انسٽيٽيوٽ آف سنڌالاجي
13. Lorna Priest Evans. (2021). Regarding the Sindhi Heh. SIL International. https://www.unicode.org/L2/L2022/22052-regarding-sindhi-heh.pdf
14. Doctor, R., Gutkin, A., Johny, C., Roark, B., & Sproat, R. (2022). Graphemic normalization of the Perso-Arabic script. *arXiv preprint arXiv:2210.12273*.
15. Ogden, R. (2017). *Introduction to English phonetics*. Edinburgh university press.
16. Manivannan, Mani M. Very long-term digital preservation and archival strategies for Tamil documents.
17. Request for Annotations for Sindhi and Behdini Kurdish, Lorna Priest Evans (SIL International), https://www.unicode.org/L2/L2020/20288-kurdish-sindhi-annot.pdf