

Extending Gurmukhi Script into the Twenty-first Century and Beyond

ARVINDER SINGH KANG^{1,3,4,*} AND AMANPREET SINGH BRAR^{2,3,4}

¹University of Mississippi

²Red Hat Inc.

³Punjabi Open Source Team

⁴Project Satluj

February 12, 2010

To appear in SIKHOLARS: Sikh Graduate Student Conference, Stanford University Campus, CA.
February 20, 2010.

*To whom correspondence should be addressed. Email: askang@olemiss.edu

List of Figures

1	Gurmukhi and Shahmukhi scripts.	2
2	Unicode range for Gurmukhi.	3
3	Photographs of Goindval Pothi.	4
4	Fedora Linux running in Punjabi interface.	6
5	Firefox browser in Punjabi.	7

Abstract

As we move forward into the digital age, the availability of digitized and standardized Gurmukhi is even more important to preserve our libraries and texts and to record our lives in the language of our thoughts.

However, without a standard for how an alphabet is encoded in a Punjabi font, different machines and browsers across the world interpret the alphabet differently. One of the solutions to this problem is embracing Unicode[14] standards. With Unicode, a specific font has one and only one code point or digital signature across all machines around the world. Our online group of volunteers at Punjabi Open Source Team[1][33][2][3] have been working since 2004 to create and absorb Unicode standards in Gurmukhi and translating open source softwares into Punjabi.

Through this paper we hope to highlight the development and achievements of a group of globally separated volunteers who have been able to come together to become one of the most successful open source Indic language communities. We also highlight challenges we face and our visions of future projects.

1 Introduction

Why is Gurmukhi literacy crucial to the future of Sikhs?

Texts record and preserve the cultural and spiritual framework of the people of individual languages for future generations. With an increasing number of Sikhs who cannot read Gurmukhi, in spite of it being ranked in the top 20 most spoken languages in the world[24], the threat to the interpretation of Sikh scriptures is imminent. It is difficult to interpret cultural, behavioral and social aspects of one language in another script, for languages are not just a way of expression, but a reflection of society and its way of life.

The year 1984 marked many unfortunate events for Sikhs. For educators, the most grim was the burning of the Sikh reference library at Amritsar and the resulting disappearance of over 1500 rare manuscripts [18], texts that helped define and interpret the framework for the 540 year-old religion.

To retain Sikh culture and identity, understanding Gurmukhi script is essential. It will be hard to understand the context of *Gurbani*, the holy scriptures from Guru Granth Sahib - the holy book of Sikh scriptures and revered as the final living Guru or teacher, without understanding the Gurmukhi script. As we move from nascent stages of digitization of Gurmukhi, it is highly important to adhere to universally adapted standards for Gurmukhi .

2 Background

2.1 Punjabi

Punjabi (or Panjabi) is the spoken language of people originating in the areas of Punjab. It is spoken by more than 90 million people[21, 27] in some form. According to Shackle(2003)[28], the word *Panjabi* was referred to as a language for the first time by Grierson[19]. In his 1916 Linguistic Survey of India [19], he coined the term “Panjabi” as the language of eastern side of Panj-Aab, or land of five rivers, and distinguished it from languages spoken in Western parts of Indo-Aryan territory, which he referred to as *Lehandi* [19].

Different forms of Punjabi known as “classical Punjabi” can be found as early as the 10th century AD. While some authors accept Grierson’s classification of Lehandi languages like Sairaki, Mirpuri and Derawali as separate from medieval Punjabi[11], others consider them as part of Punjabi[28]. It was in latter form of medieval Punjabi that Sufi movements and Bhakti movement writers like Bulleh Shah, Shah Hussain, Sultan Bahu, Waris Shah and Sikh gurus composed some of their work between the

16th and the 19th centuries. Various forms of prose like *Kafi* [13] as well as *Quissa* form of poetry [30] originated during this Punjabi era.

Gurbani in Guru Granth Sahib is composed in Punjabi as well as Sant Bhasha, a common name given to Braj, Khariboli, Sanskrit and occasionally Persian and other regional dialects[26].

The prevalent Punjabi or Modern Standard Punjabi(MSP) was shaped during the early twentieth century in reaction to the Hindi movement led by Arya Samaj and the Urdu Movement led by Aligarh Movement and the All India Muslim League, which demanded that Hindi and Urdu become the official languages of Punjab [30]. It was this religious-linguistic cohesion of standardized Gurmukhi script that Punjabi, consisting of mainly *Majhi*, and including the *Doabi* and *Makwai* dialects, started to identify with the Sikh religion[15]. Shahmukhi, a script based in Urdu, continue to be used to write Punjabi on the Pakistani side of pre-partitioned Punjab.

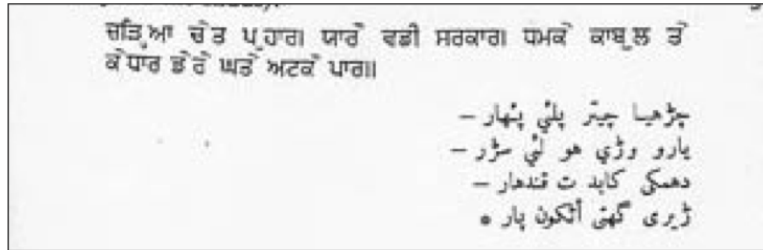


Figure 1: Punjabi in Gurmukhi and Shahmukhi scripts. (Williams, 2008)[34].

2.2 Gurmukhi

Gurmukhi means script from the mouth of the Guru. Gurmukhi was perfected and standardized by the second guru, Guru Angad Dev, with its roots in Landa script[11]. Although the Gurbani or the gurus' words in Sri Guru Granth Sahib are composed of spoken languages such as Braj, Punjabi, Khariboli, Sanskrit and Persian, the Sikh gurus chose a simple standardized script to write the texts based on thirty-five distinct consonants(ੳ - ੜ) and ten vowels(ਮੁਕਤਾ - ੋ), which unlike Sanskrit, the language of Vedas, was easier for common folk to understand.

There are a number of characters that are native to Gurmukhi and are not used in Modern Standard Punjabi(MSP). Characters like Adhak-Bindi(ੴ), Visarg(ੳ), Udaat(ੲ) and Yaksh(ੳ) are very much native to Gurbani and seldom used in MSP. Yaksh and Uddat were included in the Unicode 5.1[32]. Similarly, although MSP only supports (ੳ ੲ ੳ) alphabets as subjoined consonants to become sub-scripts when used

Gurmukhi																
Unicode.org chart (PDF)																
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
U+0A0x		ੴ	ੴ	ੴ	ੴ	ਅ	ਆ	ੲ	ਈ	ਉ	ਊ					ਏ
U+0A1x	ਐ			ੳ	ਐ	ਕ	ਖ	ਗ	ਘ	ਙ	ਚ	ਛ	ਜ	ਝ	ਞ	ਟ
U+0A2x	ਠ	ਡ	ਢ	ਣ	ਤ	ਥ	ਦ	ਧ	ਨ		ਪ	ਫ	ਬ	ਭ	ਮ	ਯ
U+0A3x	ਰ		ਲ	ਲ਼		ਵ	ਸ਼		ਸ	ਹ			.		ੴ	ੴ
U+0A4x	ੴ	-	=					~	~			~	~	~		
U+0A5x		~								ਖ	ਗ	ਜ	ਝ		ਛ	
U+0A6x						ੴ	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ
U+0A7x	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ	ੴ

Figure 2: Unicode range for Gurmukhi[14].

in conjunct form, Guru Granth Sahib also has occurrences of (ਯ) as a subjoined consonant[31]. Danda and Double Danda are used as part of Gurmukhi characters, although on Unicode table they are shared by many Indic scripts. Ik Onkar, the first alphabet of Sri Guru Granth Sahib, is used as individual character. Bindi (ˆ), tippi(ˆ) and addak(ˆ) were the additional signs used for nasalization and to indicate geminate consonant, respectively.

There also exist characters which did not exist in Guru Granth Sahib that have been adopted as loan-words to accommodate sounds from other languages such as Persian. During standardization of Punjabi in the end of nineteenth century, modern Gurmukhi script adopted five additional consonants for loanwords (ਸ਼ - ਫ਼). An additional consonant (ਲ਼) was added in the twentieth century[34], bringing the total number of consonants to forty-one.

Gurmukhi remained a language confined to scriptures for centuries. Even during the Khalsa Raj, Persian enjoyed the status of script of the court. However the Gurmukhi script continued to evolve in calligraphy as the scribes made copies of Guru Granth Sahib. The scribes used reed pens and produced the shape of characters in a different form than what we observe today[34].

According to Williams (2008), the first metal type set in Gurmukhi can be traced back to 1811, when Serampore Missionary Press produced a Punjabi typeface to print *The holy bible in the Punjabee language* and *Grammar of Punjabee language* in 1812[34]. Many of the characters looked different than they do today, and latin punctuation made its way into the Punjabi language. It was not until 1900 that the Punjabi poet Dhani Ram Chatrik opened doors of his Sudarshan Printing Press in Amritsar and according to Williams(2008) “was the first to standardise the Gurmukhi type ... using modern technique at his SPP”[34]. Williams(2008) also mentions that Chatrik was “the father

of Gurmukhi printing ... the first to print saroops of Sri Guru Granth Sahib at his SPP”[34].

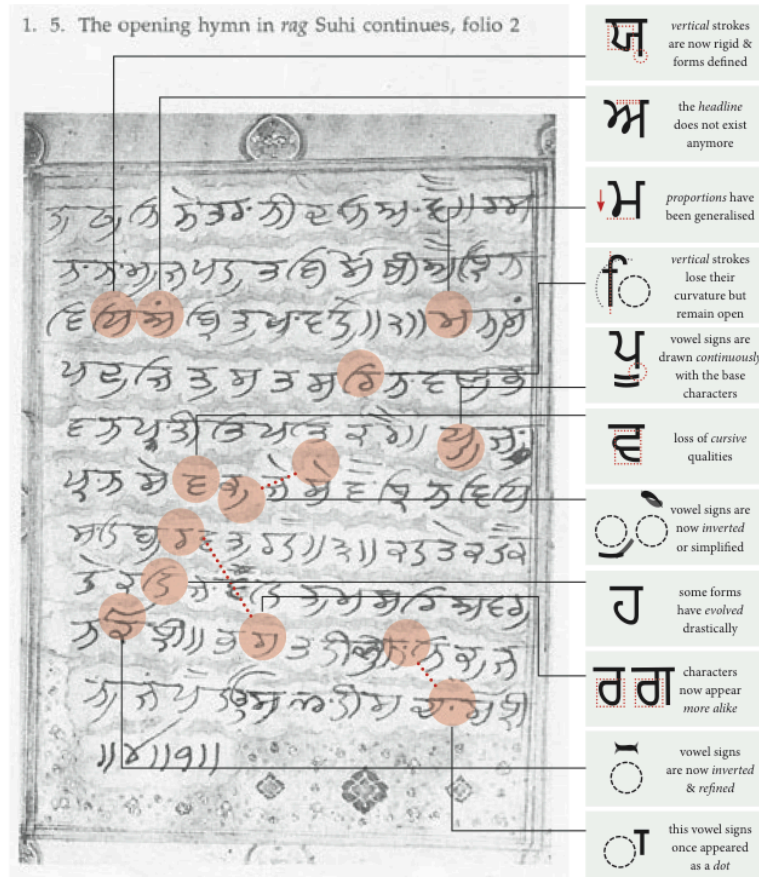


Figure 3: Analysis of writing style used in Goindval Pothis by Gurinder Singh Mann (Williams, 2008)[34].

In the early twentieth century, the dialect of written Punjabi was shifting from *Sheikhupuri* Punjabi or Lehnadi as used in the writings of Bhai Vir Singh, to a more Majhi-based dialect, although the influence of the former is reflected in literature created by writers born as late as the 1920’s[28].

3 Unicode and its Need

Although Gurmukhi script fonts were made available in the early days of computing by Dr Kulbir Singh Thind in 1985[22], script required that specific fonts to loaded on

the machine.

Without a standard for how an alphabet is encoded, different machines, softwares and browsers across the world interpret the alphabet differently. Encoding was a problem with the earlier systems using parallel standards for Indic fonts. While some systems used American Standard Code for Information Interchanger (ASCII), others used Indian Standard Code for Information Interchanger (ISCII), Latin-1, ISO or one of many other encodings. The encoding differences led to discrepancies in the original text and the output on different machines, as fonts created in different mappings did not agree on what *glyph* or font character should be loaded[35]. The Gurmukhi Unicode range as approved by Unicode Consortium is between 0A00–0A7F, so any Unicode compliant Punjabi font character has a fixed code point in this range.

With Unicode, a specific font has one and only one codebase or digital signature across all machines around the world. Such an agreement across the fonts makes it easy for browsers to display correct characters, using the correct range defined for most languages. Major common Operating Systems have been very welcoming in adapting Unicode standards. Most Linux 2.6.32 based operating systems and Windows 7 are fully Unicode 5.2 compliant. OS X 10.6.1 claims to be Unicode 5.1 compliant, but some rendering bugs for five Gurmukhi alphabets were found and reported in 2009.

This means to users that the OS of the future would be able to reproduce Punjabi text without the need of additional software or font installations. This also means that users would be able to write and read Gurmukhi Punjabi correctly, without the need of additional fonts or software installations.

4 Punjabi Open Source Team

In 2003, a group of two undergrad students¹ discovered that Linux, an open source operating system, was available in Indic Languages like Telugu, Tamil and Hindi but not in Punjabi. Linux is open source, which means that unlike commercial computer softwares, a farmer in Punjab can buy a cheap machine and not have to pay a license fee to run the operating system on his computer. The project had the potential to enable a Gurmukhi reading villager in remote Punjab to be able to work on the computer and connect with the internet. It gave them access to the information and to connect with the realities of the world beyond what is portrayed by the local media and to create open market opportunities for themselves and their fellow villagers.

In 2004, Sukhjinder Singh Sidhu and Bhupinder Singh created the first Unicode 4.0 compliant font, *Saab*. In 2004, Sidhu also created Gurmukhi Unicode Conversion

¹Amanpreet Singh Alam & Gursharan Singh Khalsa

Application(GUCA)[29] to help conversion of Dr. Thind's ASCII text into Unicode text. The same year Sidhu, along with Dr Thind, successfully incorporated characters - Udaat and Yakash - into Unicode.

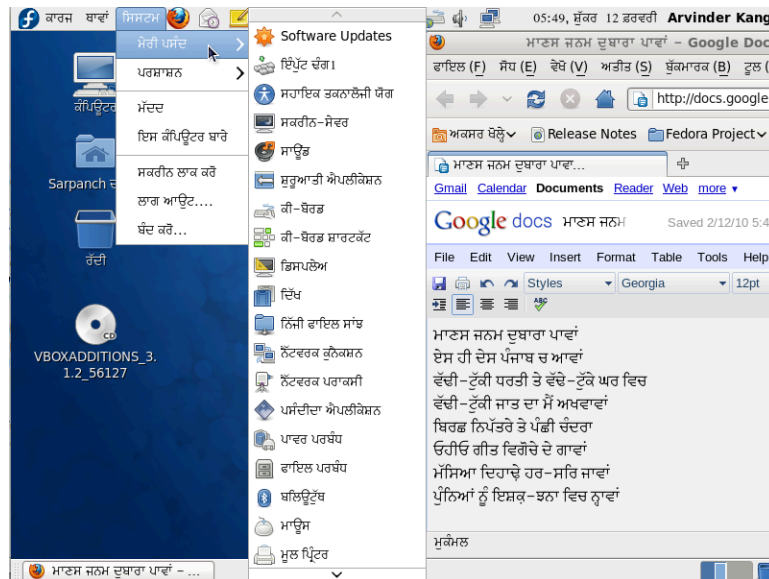


Figure 4: Fedora Linux in Punjabi interface [5].

Since then, the Punjabi Open Source Team has been translating open source software into Unicode. Users have joined and left the group and only a few actively committed users remain. In spite of being a very small team, a number of major milestones achievements have been made.

- PunLinux, the free Punjabi version of Linux Operating System has been available in Gurmukhi interface since 2004. This gives computing capability to anyone in remote Punjabi villages with minimal hardware and almost no knowledge of English. It is the only fully Punjabi Operating system, with even the installation interface available in Gurmukhi. PunLinx has now been adapted into various Linux distributions including RedHat[7] Fedora[5].
- Firefox, a browser used by a large number of internet users has been available in Punjabi interface ever since its version 1.6 release in January 2005. A Firefox mobile browser is available in beta.
- Moblin 2.1, the open Operating system by Linux Foundation for Intel based netbooks, released on Nov 4, 2009, had Punjabi as the only Indic language supporting it.

- Various other smaller projects have sprung from the core translation over time, e.g. localization of Drupal and Wordpress, a content management system, Google Punjabi translation and translation of the Facebook social network interface to Gurmukhi.
- The group has helped various vendors to test their Gurmukhi Unicode offerings such as improving OS X Snow Leopards's Gurmukhi support by filing bugs.
- Over the years we have helped direct the adaptation of Gurmukhi in Open standards such as Unicode and ISO.
- We have helped build tools and resources for software developers to adapt Gurmukhi for software and web development. These include and Internet Relay Chat (IRC) support channel(#punjabi at irc.freenode.net), a documentation library satluj.com[3], two Google groups[2][1] and a new Punjabi keyboard standard layout *jhelum*.
- Android OS, an open source operating system for mobile phones, has been successfully tested for enabling Punjabi.



Figure 5: Firefox browser has been available in Gurmukhi interface since 2005.

The Punjabi Open Source Team now focuses in four areas:

- Translation(or localization) of open source software like Linux operating system, Firefox browser and many other softwares to Unicode Gurmukhi.

- Providing open support for additional vendors to absorb Unicode standards into their open and proprietary softwares.
- Acting as guardians and watchdogs for Gurmukhi standards being absorbed into Unicode and working closely with open vendors to test the absorption of Gurmukhi standards.
- Development of tools for anyone to produce software in Punjabi, without reinventing the wheel.

5 Future And Challenges

The world is gradually moving towards Unicode and almost 50% of the web is already using Unicode standards[12], but the efforts in Gurmukhi remain limited and decentralized. For example, although we have seen a large number of Gurmukhi fonts[17] developed in the last two years[22], there are not many strictly Unicode 5.2 compliant GPL-licensed fonts freely available online for download. Secondly, most freely available fonts are the products of hobby development[20]. From a typographer's point of view[34], the aesthetic quality and legibility of available Gurmukhi fonts have a lot of room for improvement. Proprietary works done in this Punjabi/Gurmukhi domain are seldom shared.

In spite of large number of Sikhs and/or Punjabis worldwide, there seems to be limited enthusiasm for open source work in enabling tomorrow's devices for Gurmukhi. The translation has been a product of a group never larger than 40 members with only five or less active members at any given time. Mobile browser support for Gurmukhi is very limited. Optical Character reader support is available but presently does not work beyond Windows XP. The Advanced Centre for Technical Development of Punjabi language is doing some great work in Punjabi domain[23]; however their research and code mostly remains within walled gardens and focused on proprietary operating systems.

Until there is better mutual communication and understanding between discussed traditional and non-traditional groups engaged in the development of Gurmukhi/Punjabi solutions, we are going to see repetition of efforts, disagreement on standards and delay in adoption of Gurmukhi in this fast changing digital era. It can only occur if institutions involved embrace sustainable business practices, embrace transparency of efforts and take steps towards the philosophy of *share, grow and benefit*.

6 Conclusion

Data is accessible to us at a much faster rate than it has ever been. New frontiers of human communications are opening up. This can be the perfect opportunity for the renaissance of Gurmukhi.

There seems to be rekindled interest in students in the research and in the study of Gurmukhi[25, 16, 34]. However, for non-technical students to carry out research and for media to reach Punjabi audiences, there is a strong need of culture of collaboration and openness among Gurmukhi developers. Also needed is the involvement of experts from domains like design, typography, sociology, cultural studies and linguists to be the part of the debate of adaptation of the language into the newer frontiers of technology.

We feel strongly that our focus should include research on development of script in cultural and social domains. The Sikh scholars should not tie themselves only to the research in Gurbani but branch out to unearth and bring in the rich flavors of *quissa, vaar, kafi, lok bolian, lok geet, akhaan, mubaavre* and other forms of folk literature, which usually fall outside the realms of religious study but are important for the context in which we understand Gurbani. We feel a language based solely on religious aspects and out of touch with cultural heritage and a changing society will, like Sanskrit, be limited to academia, eventually losing its common touch and becoming archaic. We further feel there needs to be better collaboration between groups interested in the scholarly aspects of the language[4, 8] and groups interested in exploring the religious side of the script[6, 9].

We also believe that the growth of Punjabi language depends on the Gurmukhi script, and by providing access to the script across the devices of tomorrow, we will be able to attract newer generations towards the language and create an environment of its growth.

We hope that the presentation of this paper at the Sikhs in the World conference[10] will raise awareness about the need to learn, explore and develop in Gurmukhi. We hope it will also encourage the developers to adhere to Unicode standards when developing projects for the Punjabi or Gurmukhi literate audiences. The group hopes to make helpful connections to keep the project alive and attract more contributors to the open development of Gurmukhi Script.

7 Acknowledgements

The authors thank Dr Kulbir S Thind for his contributions to the language and for producing the first Punjabi computer fonts. The authors thank Gursharan Singh Khalsa, Narinder Singh, Sushil Kumar, Jaswinder Singh Phoolewala and Gurpartap Singh, with-

out whom these projects would have never taken wing. The authors would also like to thank Sukhjinder Sidhu, Bhupinder Singh and all the contributors of Punjabi-user group, Punjabi-translation groups and Satluj.com, who have helped in many ways. Amanpreet would like to thank his parents and the people of his village Alamwala, who inspire him in every step he makes towards fulfillment of his Punjabi dream. Arvinder would like to thank Brenda Robertson, Director of the Writing Center at the University of Mississippi, whose love for English language transmutes into his quest for Punjabi. Arvinder thanks his brother Harvinder Singh Kang for his advice and his admirable knowledge of Sikh history, and his parents for providing an environment of Punjabi and Sikhism from early on. The contributions to these projects are the result of personal time given by the contributors. Any opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views or policies, either expressed or implied, of our parent institutions, or any other entity.

References

- [1] Gurmukhi project for punjabi language support - internationalization and localization. <http://code.google.com/p/gurmukhi/>, Nov 2009.
- [2] "punjabi-users." punjabi-users google group. <http://groups.google.com/group/punjabi-users>, Nov 2009.
- [3] Satluj | ਆਪਣੀ ਬੋਲੀ, ਆਪਣਾ ਮਾਣ! <http://satluj.com/>, Nov 2009.
- [4] APNA, Academy of the Punjab in North America. <http://www.apnaorg.com/>, Feb 2010.
- [5] Fedora project a linux based operating system. <http://fedoraproject.org/>, Feb 2010.
- [6] Panjab Digital Library - Revealing the Invisible Heritage of Panjab through Digitization - Punjab Digital Library, First Sikh Digital Library. <http://www.panjabdigilib.org/>, Feb 2010.
- [7] redhat.com | the World's Open Source Leader. <http://www.redhat.com/>, Feb 2010.
- [8] UKPHA - UK Punjab Heritage Association. <http://www.ukpha.org/>, Feb 2010.

- [9] Sikh Research Institute. <http://www.sikhri.org/>, 2010 Feb.
- [10] SIKHOLARS: Sikh Graduate Student Conference. <http://www.sikholars.org/>, 2010 Feb.
- [11] Tej K Bhatia. *Punjabi : a conginitive-descriptive grammar*. Routledge, New York, 1993.
- [12] The Official Google blog. Unicode nearing 50 <http://googleblog.blogspot.com/2010/01/unicode-nearing-50-of-web.html>, 01 2010.
- [13] Peter J Claus. *South Asian folklore : an encyclopedia : Afghanistan, Bangladesh, India, Nepal, Pakistan, Sri Lanka*. Routledge, New York, 2003.
- [14] Unicode Consortium. What is unicode? <http://www.unicode.org/standard/WhatIsUnicode.html>, Nov 2009.
- [15] Harnik Deol. *Religion and Nationalism in India : the Case of the Punjab*. Routledge, London ;;New York, 2000.
- [16] Erin McLaughlin. HINDI RINNY - Gurmukhi. <http://hindirinny.blogspot.com/search/label/Gurmukhi>, Feb 2010.
- [17] Center For Development and Advanced Computing. C-dac: Gist - downloads - typeface catalogue. <http://www.cdac.in/html/gist/down/typeface.asp>.
- [18] Dalvinder Singh Grewal. Where are the 1984 Sikh Manuscripts and Articles?, 2009.
- [19] George Abraham Grierson. *Linguistic Survey of India*. Supt. Govt. Printing India, Calcutta, 1916.
- [20] Paul Grosse. Home :: Billie the cat. <http://www.billie.grosse.is-a-geek.com/>, Feb 2010.
- [21] Registrar General Census Commissioner India. Census of India - Statement 1. http://www.censusindia.gov.in/Census_Data_2001/Census_Data_Online/Language/Statement1.htm, 2007.
- [22] Gurpreet Singh Lehal. A survey of the state of Art in Punjabi Language Processing. *Language in India*, 9(10):9-23, October 2009.

- [23] Gurpreet Singh Lehal. Advancedcenterpunjabi.org Punjabi University, Patiala (Punjab), india. <http://www.advancedcentrepunjabi.org/>, Jan 2010.
- [24] M. Paul (ed.) Lewis. *Ethnologue: Languages of the World*, Sixteenth edition. Online version: <http://www.ethnologue.com/>, 2009.
- [25] Gurinder S. Mann. The Teaching of Punjabi in American Universities: Present Situation and Future Prospects. In *International Conference on the Teaching of Punjabi*. Punjabi University, Patiala, March 2008.
- [26] Gurinder Singh Mann. *The Making of Sikh Scripture*. Oxford University Press, New York, 2001.
- [27] Population Census Organization. Population by Mother Tongue. Technical report, Islamabad, Pakistan, 1998.
- [28] Christopher Shackle. *The Indo-Aryan languages*, pages 581–621. Routledge, London; New York, 2003.
- [29] Sukhjinder Sidhu. Punjabi computing resource centre. <http://guca.sourceforge.net/applications/guca>, Jan 2005.
- [30] Atamjit Singh. The Language Divide in Punjab. *South Asian Graduate Research Journal*, 4(1), Spring 1997.
- [31] Kulbir S Thind. Information Regrading Gurmukhi Unicode Fonts. <http://www.gurbanifiles.org/unicode/unifonts.html>, Feb 2005.
- [32] Kulbir S Thind. Issues Regarding the Use of Unicode Gurmukhi fonts. http://www.gurbanifiles.org/unicode/unicode_issues.htm, March 2006.
- [33] Punjabi Translation Google Groups. Punjabi Translation. <http://groups.google.com/group/punjabi-translation>, Nov 2009.
- [34] Emma Williams. A comparative study of the development of the Gurmukhi Script: from the handwritten manuscript to the digital typeface. Masters thesis, University of Reading, United Kingdom, September 2008.
- [35] Dik T Winter. Indic standards. <http://homepages.cwi.nl/~dik/english/codes/indic.html>, Nov 2009.