

8.17 Art of Hindi Dictionary Making: An Historical Exploration

Ravikant Shrama, IndLinux

Abstract—Now that we are setting out to build new digital lexicons, it is perhaps useful to revisit the ways in which modern dictionaries were built in North India from the 19th century onwards. My paper will examine the shifting ambitions, cultural assumptions, politico-linguistic, lexical methodologies and usage of resources in the process of making various kinds of corpora that we have today. What has been the degree of community participation, if any, in projects that are by their very nature mammoth. How much have we really travelled from the days of Fallon and Platt through the age of Nagri Pracharini Sabha to that of Hardev Bahri and McGregor? Do we really see a shift in the post-independence period, especially after the setting up of the Rashtrabhasha Vibhags and expert bodies like CSTT? What relationship do these experts have with the languages of the society? What are their dreams and ambitions? Further, do we have anything to learn from the projects of the past? Or the art of digital dictionary making is an exercise entirely different from the print experience? Does it help to think in terms of computers on the net as a unique, interactive mass media? If so, what are its implications for dictionary making now?

A language is not learnt from Dictionaries.
- S W Fallon

A hundred-odd years are not much in the life of a language. Yet if one looks at the history of Hindi spanning over the last century and a quarter or so, one really marvels at the distance it has traveled. There can not be any doubt about its continued expansion and enrichment in certain registers and spheres within India and beyond. Yet there is something about it that worries any lover of the language. It has been the Rajbhasha for more than sixty years now, yet it remains powerless. We have spent crores on its capacity building yet it remains intellectually resourceless. Now, the moment of Globalisation has thrown up contradictory signals thanks to the media market, Hindi has acquired an unprecedented visibility, its newspapers and magazines are clocking all time high circulation, you have all kinds of technical and financial programmes being presented at the ever busy Hindi TV and Radio FM channels and for the first time it appears Hindi has taken off from the literary stronghold it had encased itself in. But if we do a reality check in terms of hardware resources, the state of dictionaries in the language is a good metaphor for the state of Hindi in general. This paper therefore tries to excavate the history of Kosh-making in modern Hindi with a clear agenda: Is there something in this history that can teach us? Are we in a position to learn before it is too late?

Ravikant Shrama is with Sarai-CSDS, New Delhi (e-mail: ravikant@sarai.net)

To be sure Kosh-making is an age-old practice in India and its history has produced any number of good grammarians and lexicographers - Amarkosha and Nighantu Shastras belong to ancient India. But all of that was in the distant past, the past of languages that are alas dead and gone from our midst. What is relevant for us, especially in the context of Localization, is the living languages the languages we still speak and write and hope to localize digital machines in. For this I would like to pick up the story from Nineteenth Century when colonial masters were trying to understand who speaks what so that they themselves could communicate with their diverse subjects. Although the making of Colonial Lexicons is almost as old as Colonialism's efforts to expand in India in the 18th century itself, let us start with a more definitive work published in 1879. I will spend some time talking about SW Fallon's A New Hindustani-English Dictionary, because it captures a fuzzy moment in the history of modern lexicography, and because Fallon himself spends great deal of time explaining to us his intent and purposes an act not too many people would sadly repeat in future.

So what does Fallon's Dictionary look like? It is a 1216-page thick long tome that uses four scripts, Arabic-Persian and Devanagiri and of course Roman for English. It showcases 'the spoken and rustic mother tongue of the Hindi speaking people of India; the exhibition of the pure unadulterated language of women; and the illustration given of the use of words by means of examples selected from the every day speech of the people, and from their poetry, songs, proverbs, and other folklore'. In his 'Preliminary Dissertation', Fallon expresses his dissatisfaction with the way linguists have understood language as a refined and pure repository of select users with the result that they have ended up describing and indexing languages on the basis of their religio-linguistic roots. He rejects both Pandit Hindi and Mullah Urdu and goes for the middling, inclusivist Hindustani written in two scripts. The illustrations provide an uncanny amalgamation of dohas and she'rs, ghazals and kavitaai, kajri, holi and Marsiya. A special emphasis is laid on the *Zanana boli* pr *Rekhti* and the basis for that choice is explained thus: "The seclusion of native females in India has been the asylum of the true vernacular, as pure and simple as it is unaffected by the pedantries of of word-makers. it is also the soil in which the mother-tongue has its most natural development.... Yet this true vernacular is not all confined to the narrow home in which it has been kept alive. The inherent vitality of living speech, and the all-pervading influence of women on language, strenuously strive to restore the deposed natives of the soil to their rightful inheritance...."

If we conclude on the basis of this description that Fallon was involved in a romantic project of excavating the past for its own sake, we would be terribly mistaken. He was trying to make a dictionary that could be used by everybody including the colonial officials who were

struggling hard to come to terms with Indian languages. To dispel any such notion he gives us a whole comparison of scientific terms translated into Aabic, Sanskrit and Hindustani. The results are breathtakingly useful, even today:

Retarded Velocity Valve	घटती चाल खुल मुंदनी	मोतनाकिस मिकदार-ए-हराकुत	न्यूनमान गति प्रमाण्य
Forces in equilibrium	तुले हुए ज़ोर	मुयुल-ए-मोतदिला:	तुल्यमान शक्ति
Rotatory Motion	चक्कर चाल	हरकते वज़ाई	चक्र गति
Perpendicular Line	खड़ी लकीर	आमूद	लंबक
Proportion	बराबर निस्बत	तनासुब	परस्पर संबंध
Right Angle	खड़ा कोना	ज़ाविया-ए-कायमा:	सम कोण
Obtuse Angle	फैला कोना	ज़ाविया-ए-मुन्फरेजक	अधिक कोण
Acute Angle	सुकड़ा कोना	ज़ाविया-ए-हदा:	न्यून कोण
Parellel Lines	बीच बराबर लकीर	खुतूते-मोतवाज़ी	समान अंतर रेखा
Diagonal	अध काट	कुत्र	भुज कर्ण
Sine	सामने की नाप	ज़ैब	भुजज्य
Cosine	साथ कोने की नाप	ज़ैबे मुस्तावी	कोटिज्य
Tangent	छूती नाप	ज़िल/मोमास	स्पर्श रेखा
Secant	काटती नाप	सहम	छेदन रेखा

Just look at the simplicity of the so-called rustic usage: Exactly the opposite of the contrived language of modern lexicography which is made exclusively from using the roots and derivative rules mostly if not exclusively from Sanskrit. More on that later. For now let us stick to some other lexicological efforts in the Colonial North India. We are all familiar with the story of how Hindi itself became a mission in the anti-colonial struggle and people wrote in it to liberate themselves. What I hope to show here is that the exercise of lexicography was pretty much a communitarian project till quite late. For this, the next text I choose to highlight is one concerned with Standardization of Hindi. Anant Choudhary's well researched work on the making of Hindi Language Script and Grammar in the early 20th century (*Nagari Lipi aur Hindi Vartani*, Bihar Hindi Granth Akademi, 1973) tells us that the task of standardization of Hindi orthography and grammar was actually taken up on a big scale. People sent out open calls in magazines, and those who knew language and could write, wrote in until there was a raging debate that went on for years. Fixers of language came up with models and each notation and grammar rule was hotly debated. This is a good contrast to our day and age when we see linguistic diversity and

controversy as problems and shy away from taking opinions, especially from the people who use language and therefore know how it works.

My third exhibit is the Sankshipt Hindi Shabdsagar, compiled from the Ramchandra Verma original Kosh and published by Nagri Pracharini Sabha in 1933. The kosh boasted of *lokprayukta* words taken from *deshi* and *videshi* sources and various *bolis*. Go to any page of the dictionary and you will any number of words from diverse, words that have become part of people's lexicon in the true blue *Bhasha bahta neer* style

The thing I want to emphasize is that it was a multilingual world, in spite of the parallel movement against the Urdu language and Script. There was an ambition to develop a script and register that could cater to all kinds of needs. That is how you had magazines like Devanaagar in the early decades of the last century, short-lived in the first instance and then revived again in the 1950s, that published material in various Indian languages written in the Nagri script. Granted that these efforts were limited and few and far in-between, but what we need to take from them is the idea that closing doors on languages would not make your own very prosperous.

It is not as if things changed dramatically after the famous midnight of Independence. There were several academies at work and the task of resource building for Hindi was much more institutionally de-centralized than it is today. Hindustani Academy for example had stalwarts like Dheerendra Verma who inspired people like Amba Prasad Suman to compile a Compendium of Agricultural terms in Brajbhasha. The two-volume illustrated compilation remains a delight inasmuch as it covers all the important tools, technology, practices and actions associated with agrarian life. You have terms on masonry, carpentry, metallurgy, irrigation, weaving, dyeing, oil-pressing, animal husbandry, pottery, ironsmithy as well as those associated with agrarian trade. The lexicon went beyond its set goal of overtaking Grierson both in terms of quality and quantity, as in a number of words. Let us see who and it is really a who is who of Hindi - is saying what in the blurb of a book published in 1960:

1. हिंदी का क्षेत्र विशाल है. उसकी विशालता का रहस्य उसकी उपभाषाएँ हैं. निस्संदेह हिंदी की उपभाषाओं में ही उसकी प्रतिभा छिपी हुई है. प्रस्तुत खोज प्रबंध इस सत्य को स्पष्ट करता है तथा विद्वानों एवं भाषा-प्रेमियों का ध्यान उस असीम खजाने की ओर आकर्षित करता है, जिसका उपयोग यदि शीघ्र नहीं किया गया तो हिंदी का प्रकृत स्वरूप; उसका निजी स्वरूप विलुप्त हो जाएगा.

---विद्या भास्कर, मंत्री तथा कोषाध्यक्ष, हिंदुस्तानी एकेडेमी, इलाहाबाद, 1960.

2. मेरी निश्चित सम्मति है कि अलीगढ़ क्षेत्र की बोली के आधार पर 'कृषक जीवन संबंधी ब्रजभाषा शब्दावली' शीर्षक बृहत शोध-प्रबंध हिन्दी बोलियों की समृद्धि का ऐसा पक्का प्रमाण उपस्थित करता है जिसे देखकर हिन्दी की अभिव्यक्ति क्षमता के प्रति मन में नयी आस्था उत्पन्न होती है...हिन्दी के कल्याण के लिए यह ग्रंथ छपना ही चाहिए.

---डा. वासुदेवशरण अग्रवाल

3. जनता की बोलियों में तद्भव शब्द बहुत बड़ी संख्या में पाये जाते हैं. साहित्यिक हिन्दी में इनकी संख्या कम होती जा रही है, क्योंकि ये गँवारू समझे जाते हैं. वास्तव में ये असली हिन्दी-शब्द हैं और इनके प्रति विशेष ममता होनी चाहिए. कृष्ण की अपेक्षा कान्हा या कन्हैया हिन्दी का अधिक सच्चा शब्द है.

---धीरेन्द्र वर्मा, भूमिका से:

4. विविध कला-कौशलों तथा व्यावसायिक शिक्षा के क्षेत्र में पारिभाषिक शब्दों की समस्या को हल करने के लिए हमें एक दूसरी दिशा में भी खोज कार्य को प्रवर्तित करना है. किसानों, मज़दूरों तथा अन्य श्रमजीवियों की बोलचाल की भाषा में समाजशास्त्र, शिल्प तथा उद्योग-धंधों के बहुतेरे बढ़िया-बढ़िया शब्द मिलेंगे जो राष्ट्रभाषा की समृद्धि के पूरक हो सकते हैं. ऐसे शब्दों का सर्वे और संग्रह कराना परमावश्यक है; अन्यथा केवल अंग्रेजी की तालिका तैयार करके उनका पर्याय प्रस्तुत करते जाने की परिपाटी पर ही निर्भर करने से हम अपनी लोकभाषाओं के हजारों अर्थपूर्ण उपयोगी जीवित पारिभाषिक शब्दों से वंचित रह जाएँगे.

--- विश्वनाथ प्रसाद, भारतीय हिन्दी परिषद के दशम अधिवेशन, 1952, (आगरा) में 'हिन्दी गवेषणा और पाठ्यक्रम का पुनःसंगठन' शीर्षक से दिए गए भाषण से उद्धृत.

What we have here is a series of recommendations for using living resources whether in the tadbhav form or borrowed from the usages in the many sub-languages of Hindi. And these are no mean characters we are quoting they were the best people in the business. We also have similar sentiments from scholars like Hajariprasad Dwivedi and Rahul Sankrityayn who collected and wrote on the folk literature and culture of the Indian subcontinent and beyond.

But things changed slowly and surely for worse in the post-Independence India, at the 'moment of arrival' for Indian Nationalism. Institutionally, it happened with the making of expert bodies at the top, it happened with the Rajbhasha Vibahags in various central and state government institutions. But there was also a deeper social, political and philosophical reason for the kind of Hindi we thought was good. Hindi had won its cause in the Constituent Assembly and the fact of political Partition prepared the ground for a larger linguistic separation in the remaining part of North India. There was a definitive campaign launched by Hindi politicians like Ravishankar Shukla and others against the use of middling language Hindustani in the government-owned mass media such as All India Radio. The advocates for the exclusivist and pure type of Hindi were actually a paranoid lot they feared any affinity with Urdu or the so-called other dialects and espoused a direct lineage with Sanskrit. In this imagination, all the 'foreign' as well as rustic elements in a language had to be weaned out. And so the new Hindi that was constructed in the government manufactories was sought to be exclusively Sanskrit-based.

This created a massive rift in the public body of language and knowledge and created a curious paradox those who worked with living people's languages acquired the status of antiquarians and those who

recycled a 'dead' language in the name of the 'language of the nation-state became its leaders. But that was the official approach and their success was as far as it could go! In spite of the omnipresent hoardings and road-signs, people did not really take to the officialese like 'bhoomigat paidal paar path' and 'durgatna aashankit Kshetra' for they were fond of another kind of language, a mixed language, which for example, was used in the Hindi films. Films worked with a different logic of market and profit and access was their mantra. They were not driven by an ideology. If films did not speak to people they would reject them they would be declared flops at the Box Office. Look at the picture now. Where is the Official Hindi now? Can you find much similarity between what is being spoken 24X7 on these new channels and the one we had come to identify as doordarshan Hindi: let's listen to Hindi in news rather than Hindi news variety? Perhaps not. Let us look at the Hindi content on the Internet, at the various wiki-like efforts and see if there is departure from purist cananization of Hindi literature. There definitely is.

So that is the point really of this paper. Computers connected to the Internet are mass media. The mass media has to speak to, with and through the people. It is the people who would use these tools. Experts are best deployed at the back end. To localize tools is a welcome idea but to present the interface in a dead language is worse. It is better left the way it is in English or we try and be generous and innovative in our resource-building exercises. Time to hear S W Fallon and Hindi Cinema. Time to de-stigmatize Doordarshan and Akashvai Hindi. Time also to take the public domain route so that users of computers and languages have a say in the way their machines talk to them. In other words, time to ask people to make tools in their language.

ACKNOWLEDGMENT

I must thank the organisers and participants of LRIL 2007, especially Dr. Alka Irani and Ms. Shashi Palekar for their hospitality and warmth. I also owe huge debt to Prof. Shahid Amin for pointing me towards Fallon a decade or so back. My colleagues on the the Indlinux Project Gora Mohanty, Guntupalli Karunakar and Ravishankar Shrivastav have worked tirelessly to take it to where it stands now.

SELECT READINGS

1. S. W. Fallon, *A New Hindustani-English Dictionary*, Uttar Pradesh Urdu Akademi, Lucknow, 1955 Edn.
2. Amba Prasad 'Suman', *Krishak Jeevan Sambandhi Brajhasha Shabdavali, vols.1 and 2*, Hindustani Ekedemi, Allahabad, 1960.
3. Alok Rai, *Hindi Nationalism: Tracts for the Times, #13*, Orient Longman, Delhi, 2001.
4. Ramchandra Verma, *Sankshipt Hindi Shabdsagar*, Nagri Pracharini Sabha, Kashi, 1981, Edn.
5. Ramchandra Verma, *Manka Hindi Kosh*, Hindi Sahitya Sammelan, Prayag, 1953.
6. I have also seen and consulted the various volumes prepared by CSTT.

This paper was presented at LRIL-2007: National Seminar on Creation of Lexical Resources for Indian Language Computing and Processing at C-DAC Mumbai (26th to 28th March 2007), jointly organized by the Commission for Scientific and Technical Terminology (CSTT), New Delhi, MHRD, Govt. of India and the Centre for Development of Advanced Computing (C-DAC), Mumbai, Department of Information Technology, MC&IT, Govt. of India.