Lexicon Building and its Challenges in Language Teaching

Udaya Narayana Singh

**Chair Professor of Linguistics and Dean Faculty of Arts
Amity University Haryana, India**

(A Reflective Article)

Although a lexicographer is generally described as "A… harmless drudge that busies himself in tracing the original and detailing the signification of words" (Samuel Johnson. 1755. *Dictionary of the English Language*), thanks to the advancement in Language Technology as well as due to emergence of alternative models of dictionary-making (such as '*Community Glossary programmes*'), it is high time we take out the 'drudgery' part from the compulsive description of a lexicographer. Whereas we had devoted a considerable time to battle out the structure and nature of the *lemma*, or the canonical form of a *Lexeme*, when we have a set of many forms that have the same meaning, the *lemma* being that particular form which is chosen by convention to represent the lexeme, we now have new kinds of dictionaries like the *Longman Basic English*, or the *Longman-CIIL Bilingual Dictionaries* in ten Indian languages under my general editorship (Cf. Singh et al, 2011ab, 2012a-d), where the source of selecting head-entries is dependent purely on the frequency of occurrence of a given form in a large corpus- in this case, either the British National Corpus (or BNC) or the CIIL corpus. Consequently, we have '*do*', '*don't*', and '*doesn't*' as independent head-entries as well as function words such as '*d* or *ain't*.

In this context, works of Vincent Oi (2000 at the 19th ACAME Conference, Amsterdam, Rodopi, and reprinted 2007) is worth-mentioning as he was also trying to discover the differentiate between the' Western' and the 'Asian' realities that must get reflected in their differential language use as available in the 'New Englishes' versus the 'Old Englishes' corpora. Tabloid and Newspaper corpora

and the *Collins-Cobuild Corpora* were the two main sources electronic resources. Since 'collocations' are a dynamic and recurring set of expressions and 'idioms' are a frozen lot, what better alternative can there be to find texts where both have high chances of occurrences. These also take much less time and prepare the statistic before one can bat one's eye-lid. Which means, it removes the drudgery, to a great extent, because one is not merely dealing with the stabilized world of 'inner circle' variety of English focusing on the Standard British and the American varieties (cf. Kachru, Braj B. 1986), but also on the large number of vibrant 'outer circle' Englishes here. Thus, the lexicographer here is acting like a sociolinguist discovering the dynamicity of the language she is working with. She is also acting like a sociologist of language discovering the value systems of a given society as reflected in their linguistic expressions. The question is: To this already changed scenario of modern-day lexicography, how does one bring in a greater amount of color and fresh wind – some joy, as one struggles hard with the arduous task of dictionary-making? We need to think together about the possibilities. But even before we do so, let us look into the possible devils and detractors.

1. Politics of Lexicography

Many complain about a dictionary in terms of its politics. It is the *Politics of Lexicography* and the resultant *biases* that define the stand of each one of us. Are we leaving out entries as we are a sexist lot? (Many feminist theoreticians would like to believe so.) Or, are we somehow censoring certain words that we consider 'vulgar' or 'imbued in colloquialism as reflected in the so-called 'restricted codes' of a certain socio-economic class, or ethnic agglomerate or a color group? We are obviously not talking about deletions dictated by constraints of space often quoted as excuses by the publishers. At times, our purpose or definition of domains of use can dictate what to take and what to leave out. We have no quarrel over such decisions. But when we decide to show (or hide) our biases in what we accept I would like to use the example of the ever-popular dictionary called '*Calantika*' (meaning 'On-going/ever-moving') created by a scientist-turned author-lexicographer, Rajsekhar Basu (1937) in Bangla – who was a humorist drawing enormously from colloquialism in his short fictional writing which used a loosely outwardly false scholarly style to reel out the most hilarious tales. He left out a large number of classical borrowings in Bangla, later included in voluminous dictionaries such as '*Bangiya Sabdakosa*' (1933-1947/1966) by the celebrated lexicographer Haricharan Bandyopadhyay (originally a school teacher but later accommodated by Rabindranath Tagore in Viswa-Bharati, because by then he had sold all his property and belongings to

bring out the monumental work in 104 parts). However, if Basu left out words, he was trying to christen many untouched and untouchable expressions by being inclusive. That was his politics. Haricharan was merely being all-inclusive, but his scholarly bias might be proven if one took up the task of bringing out a comparison of which lexicographer in Bangla had included or excluded which expressions, and prepares a huge database of this information. This is surely an interesting area of research. But I would still like to echo Jonathan Green (1996) where this British historian of lexicography admits that "while no one of us can escape completely our own prejudices and assumptions, at least attempting to do so is the first obligation of any historian – even while acknowledging the truth that much of the joy and beauty of history is provided by the authors' biases" (Cf. Trageser, 1996). It emerges from this discussion that each lexicographer has her bias – even though some do not talk about it in the introductory remarks of the work, but that it becomes a research question again to discover them in the compilations that have been published in our languages.

A similar question arises in another area of dictionary-making: Are we using (or misusing) 'etymology' to justify certain late intrusions into the Standard Vocabulary culled out of a non-standard variety? Sukumar Sen had given such fanciful etymology example in Bangla, where the dictionary-maker would try and 'create' a source such as '*nakha-haranika*' for a colloquial expression (picked up from a dialectal source obviously) such as '*narun*' (meaning 'a file – to cut and polish nails') – got from collocations like '*naker badale narun pelam*' meaning 'Asked for the nose, and got a file!' In this context, I think it would be good to underscore the etymology of '*joy*' that we are talking about:

Joy c.1225, "feeling of pleasure and delight," from O.Fr. *joie*, from L. *gaudia*, pl. of *gaudium* "joy," from *gaudere* "rejoice," from PIE base **gau-* (cf. Gk. *gaio* "I rejoice," M.Ir. *guaire* "noble"). [*Online Etymological Dictionary* at http://www.etymonline.com/]

It looks like 'pleasure', 'delight' and 'rejoicing' are the key-words here. One source of this pleasure comes from our encounters with new words and expressions, or neologisms. One question arises in the minds of language educators, teachers of languages, writers, translators and even general (informed) public as to what is or should be the acceptable way of introducing new words in a language. At times neologism comes about because a great author or an important text uses the word in question, either by borrowing (including classical borrowing) or through combination of generally unattached word-elements or through abbreviation and acronym-formation. Alternatively, someone coins a word or phrase that happens to stick, which can be a pure accident. Sometimes, a new technology is introduced, and some brand names associated with it also creeps

in. In some other cases, when speakers of one language come into contact with speakers of another, such influx is common. If we talk about the technical vocabulary construction, it is usually driven by official language authorities, and many of these coinages may not have a long shelf-life.

2. Team Efforts, or the Workshop Method

It is at such moments that face-to-face or long-distance workshops methods can be used. Many a times, the 'Loneliness of a long-distance linguist' is of great use, as they join a community glossary program and do well in the process of taking important steps towards software localization. There is a lot to do in this area in our countries. Dictionaries tracing word attestations in texts and individual word-histories are very important activities. We need to bunch up all inclusions and innovations decade-wise and find out the logic behind the trends. Everyone may not be interested in this challenging area but there are readers and scholars who care. For instance, Michael Penn (2000) the University of Wisconsin news site tells us about someone like Frederic Cassidy and his team's heroic efforts to build the DARE project – the Dictionary of American Regional English under the title 'Romancing the word: The unfinished adventure of a dictionary maker's life' by giving the following description: "You and I may think we do, as readers and as writers, but not like Cassidy. Ours are tawdry liaisons, *use-'em-and-lose-'em* affairs that have us gallivanting with another word before the memory of the last has faded. We are word gigolos, compared to Cassidy, who is the king of lexical romance." In this methodology, the compiler would read fictions or plays or texts written in dialectal versions of English, and literally "meets with" these expressions. In the process, we may get "distracted by the contours of those unfamiliar words." But that is how, one must build resources that would help the ELT practitioners scattered over the vast English-speaking tracts of the world. Penn (2000) further elaborates:

> "Cassidy reads dictionaries - actually reads them, as opposed to flitting through them in a purpose-driven manner. ... Cassidy has spent much of his life doing just that, chasing down the rare bits of Americana that hide in backwaters and bayous. He and a small band of fellow logophiles on UW-Madison's campus are in the middle of an effort to publish the *Dictionary of American Regional English*, the most ambitious catalogue of the country's vernacular ever attempted. Some forty thousand nuggets of regional dialect, gathered from all fifty states, have been printed in the dictionary's first three volumes."

I have had some experience of looking into yet another study in India about such innovations and lexical coinages that had not found place in our most extensive dictionaries and was surprised to see that between 1947 and 1990, close to 7,000 new words and expressions had entered into our fictional writings in Bangla[1]. These could be eye-openers for any researcher, even though because of lack of a credible OCR technology, this is still time-intensive research.

In fact, as of now, the readers are expected to spend a long time browsing through dictionaries if the text is written in a dense manner or if it is a technical text. I would anticipate that there would be a lot of electronic help tools developed in Indian languages in near future. A variety of dictionaries floated as a part of CIIL's *Active Dictionaries* (to serve at the back-end tool under CIIL-NSOU's On-line Language Teaching program) may be mentioned here where the dictionary has the required visual as well as audio components, besides grammatical and semantic details that could be dynamically generated by firing queries. We also need other kinds of help-tools meant for children who are slow beginners in reading. Even as they grow up, since all linguistic products (from the epics and the classics to grammars and dictionaries) have abridged versions – meant for elementary level students or children, the product that comes closest to this kind of help-tool is a *picture dictionary* (in print or on-line) which contains words or entries, accompanied by *photographic* illustrations or drawings depicting what the words mean in as much as possible.

But as each of these new formats pose problem, we need to discuss them among professionals. At times, there are too much expectations from our dictionaries – especially because our languages in this continent are being modernized now, and many languages are still undergoing processes of standardization. For them, some ambitious dictionary-makers also take it upon themselves to prepare codes for creating and maintain standards. But we need to decide if our dictionaries are to be modeled after 'Road-maps' or are they to be treated as 'Etiquette-guides'.

We are all aware that dictionary-making is a very time-consuming process and needs utmost attention and caution. We all know that it took Sir James Murray 37 long years to edit the *Oxford English Dictionary* (1928), and it took another thirteen years to finish its fifteen volumes after he passed away in 1915. The editors of *Deutsches Wörterbuch* (1954, abbreviated as DWB), a German dictionary (the folklorist 'Brothers Grimm' were a part of the editorial team) took

---

[1] Personal communication, from the ongoing research findings by Suchita Singh under S.K.Banerjee at the University of Calcutta.

122 years on their project. Today's enterprises are much faster, and yet there is so much of thinking, contemplating, conceptualizing and working manually that are involved here. We need to pay attention to numerous areas in Lexicography these days – beginning from corpus-based lexicons, to concordances, special-purpose and region-specific dictionaries as well as text decoding, need assessment and structure of lexicons, dictionary tools and software in the context of learning strategies and progresses based on dictionaries, and training skills in dictionary-education and hands-on education in practical lexicography. This is absolutely necessary because creating appropriate lexicons or word approximators would be an essential tool for the language teachers these days. The lessons they would need to prepare will depend heavily on the kind of lexicon the students have in their knowledge system and the kind of word-power the teachers need to build for them so that they do not falter while speaking or writing in an alien tongue which they need to master only in a formal or unnatural setting of a school.

3. Further problems

As we know, globalization has intensified the code-mixing and code-switching activities among bilingual and multi-cultural learners and speakers of English plus other-tongues or mother-tongues. The presence of several linguistic groups in a social setting has become the norm rather than the exception not just in multilingual societies but throughout the world (UNESCO, 2003). This has happened due to large-scale dislocations, migration and contact situations among between speakers of different languages with the advent of globalization (Jasper, 2020).

Today, many people are forced to move from one place to another in search of better opportunities. They learn new tools and technologies to be able to survive in the new setting, and building word power and speaking skills in English is one such tool. This helps them to interact with speakers of different languages which contribute to making their boundaries porous (Blommaert, 2010). This has also raised the debate on whether plurality of languages could be taken as a resource or it is still a challenge (Cummins, 2007; Garcia, 2009a; Cook, 2009; Agnihotri, 2009). This fluidity of conversational situations – as in the countries like India, can be best utilized to the advantages of the teaching community if their task could be supplemented by appropriate age-specific, context-sensitive, and task-oriented lexicons. Such projects and programmes need to be undertaken by the language teaching experts along with dictionary-makers.

In fact, many scholars talk about something called 'Multicompetence' in today's multicultural context. Here local practices and usage of mother-tongue expressions play a major role while using a global language such as English. This then opens up the conversation for a negotiation and throws it open for interpretation by the fellow speakers who may be bilingual in other languages with knowledge of English. In such cases, as Canagarajah (2001: 401) says, "Competence doesn"t consist of separate competencies for each language, but a multicompetence that functions symbiotically for the different languages in one's repertoire integrated systems." So, the challenge for the ELT specialists and teachers of Global English becomes that of a test to help these speakers use a kind of English with a common minimum lexicon that each of those linguistic competencies would have internalized. This approach will stop short of allowing a free-flowing 'translanguaging,' i.e. an "intentional and systematic use of two languages in educational contexts" (Williams, 1996; Lewis et al, 2012).

We are all aware of the Chomskyan position on the issue of 'Ideal Speaker-listener.' Chomsky (1965: 4) had argued that his "Linguistic theory is concerned with an ideal speaker-listener in a completely homogeneous speech community." But as Skutnab-Kangas (1981) would show as a sociolinguist, the native monolingual speakers who are "accustomed" to use a single named language could be thought to be using a bounded and discrete system of semiotic resources, although the reality is otherwise. In reality, the speakers today use languages at different layers with different degrees of competence, and at times, make deliberate mixtures and "mistakes" to sound like their interlocutors only to secure a greater trust and friendship through a mutually agreed upon style of speech. In today's pluricultural contexts, if "meaning making is not confined to the use of languages as discrete, enumerable, bounded sets of linguistic resources" and if "signs are available for meaning making in communicative repertoires" that extend "across languages and varieties that have hitherto been associated with particular national, territorial and social group" (Rymes, 2010: 21), then the whole strategy of ELT will be required to be re-conceptualized and freshly planned to be meaningful in a fluid multilingual context. There, preparation of topic-specific, genre-sensitive and region-oriented lexical resources would be absolutely essential to help the language teachers.

The purpose of this paper was to raise certain issues in ELT or LT in general in the context of need for a fresh set of strategies for multi-cultural nation-states, where the lexicographers will have a special role to play to aid the teachers as well as make the interpreting mechanisms work well for Global English speakers from other parts of the globe. Many native English speakers

complain of great difficulties in reading through our tabloids, daily newspapers and news-feed on OTT platforms. To them, this appears to be largely un-English but as Braj B. Kachru had floated the concept of '*World Englishes*' (with a plural marking) long ago (Cf. Ferit Kilickaya 2009), there are fresh tasks and challenges for the ELT community and applied linguists.

References

Agnihotri, R. K. 2009. Multilinguality and a new world order. In A. K. Mohanty, M. Panda, R. Phillipson, and T. Skutnabb-Kangas, (Eds.) *Multilingual Education for Social Justice: Globalizing the Local,* 169-186. New Delhi: Orient BlackSwan.

Bandyopadhyay, Haricharan. 1933-1947. *Bangiya Sabdakosa* in 104 parts. Visva-Bharati. 1966 & 2004/Sahitya Akademi – a consolidated two-volume edition.

Basu, Rajshekhar. 1937. *Calantika.* Kolkata: M.C. Sircar & Sons Pvt Ltd.

Blommaert, J. 2010. *The sociolinguistics of globalization.* Cambridge, England: Cambridge University Press.

Cook, P. 2009. Multilingual Universal Grammar as the norm. In I. Leung (Ed.) *Third Language Acquisition and Universal Grammar,* 55-70. Bristol: Multilingual Matters

Cummins, J. 2007. Rethinking monolingual instructional strategies in multilingual classrooms. *Canadian Journal of Applied Linguistics,* 10(2), 221–240.

Garcia, O. 2009. Bilingual *education in the 21st century: A global perspective.* UK: Wiley-Blackwell.

Green, Jonathan. 1996. *Chasing the Sun: Dictionary Makers and the Dictionaries They Made.* New York: Henry Holt and Co.

Grimm, Jacob & Wilhelm Grimm, eds. 1854. *Deutsches Wörterbuch.* Leipzig: Verlag Von S. Birzel.

Jaspers, J. 2020. Linguistic dilemmas and chronic ambivalence in the classroom: introduction, *Journal of Multilingual and Multicultural Development*, Retrieved from 10.1080/01434632.2020.1733586

Johnson, Samuel. 1755. *A Dictionary of the English Language (in which the words are deduced from their originals, and illustrated in their different significations by examples from the best writers.* London: W. Strahan.

Kachru, Braj B. 1986. *The Alchemy of English: The Spread, Functions and Models of Non-Native Englishes.* Oxford: Pergamon.

Kilickaya, Ferit. 2009. 'World Englishes, English as an International Language, and Applied Linguistics.' *English Language Teaching* 2.3: 35-38.

Lewis, G., Jones, B. & Baker C. (2012). Translanguaging: Origins and development from school to street and beyond. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 18(7), 641-654. http://dx.doi.org/10.1080/13803611.2012.718488

Murrey, James et al, comp & eds. 1928. *Oxford English Dictionary.* Oxford University Press.

Oi, Vincent. 2000. 'Asian or Western Realities: Collocations in Singaporean-Malaysian English' ; in J.M.Kirk, ed. *Corpora galore: Analysis and Techniques in Describing English: Papers from the 19th ACAME Conference*, Amsterdam, Rodopi, 73-89. Also reprinted in 2007: Teubert W. and R Krishnamurthy (eds) *Corpus Linguistics: Critical Concepts in Linguistics,* London: Routledge, pp. 194-212)

Penn, Michael. 2000. 'Romancing the word: The unfinished adventure of a dictionary maker's life' (https://news.wisc.edu/romancing-the-word-the-unfinished-adventure-of-a-dictionary-makers-life/) June 15.

Rymes, B. R. (2010).Classroom discourse analysis: A focus on communicative repertoires. In N. Hornberger & S. McKay (Eds.), *Sociolinguistics and Language Education.* Avon, UK: Multilingual Matters

Singh, Udaya Narayana 2011a. [Ed. & Comp. with Suchita Singh & Chandan Hazra]. *Longman-CIIL Basic English-English-Bengali Dictionary.* 2011b. [Ed. & Comp. with B.R.Pal & S.K.Singh]. *Longman-CIIL Basic English-English-Hindi Dictionary.* New Delhi: Pearson Education.

Singh, Udaya Narayana 2012a. [Ed. & Comp. with Pradhan Gurudatt & S.S.Yadurajan]. *Longman-CIIL Basic English-English-Kannada Dictionary.* 2012b. [Ed. & Comp. with Basanta Panda & B.N.Patnaik]. *Longman-CIIL Basic English-English-Oriya Dictionary.*2012c. [Ed. & Comp. with V.S.Nair & Sadashivan]. *Longman-CIIL Basic English-English-Malayalam Dictionary.* 2012d [Ed. & Comp. with Sam Mohanlal, L.Ramamoorthy & Vimala]. *Longman-CIIL Basic English-English-Tamil Dictionary.* New Delhi: Pearson Education.

Skutnabb-Kangas, T. (1981). *Bilingualism or Not: The education of minorities.* New Delhi: Orient Longman.

Trageser, Jim. 1996. Review of *Chasing the Sun,* titled 'History of dictionaries falls short of own goals', *American Reporter;* see http://trageser.com/books/bookreview-chasing.php

UNESCO. *Education in a multilingual world.* Retrieved from http://unesdoc.unesco.org/images/ 0012/ 001297/129728e.pdf

Williams, C. (1996). Secondary education: Teaching in the bilingual situation. In C. Williams, G. Lewis & C. Baker (Eds.), *The language policy: Takingstock*, 39-78. Llangefni, UK: CAI.

ABOUT THE AUTHOR

Born in 1951, Prof. (Dr.) Udaya Narayana Singh is a linguist, creative writer, translation theoretician, lexicographer & a researcher in culture studies. He held several administrative and academic positions in different institutions spread all over India. He was Professor, Rabindra Bhavana (2009-2016) and the first Pro-Vice-Chancellor (2010-12) of Visva-Bharati, Santiniketan. Earlier, he was the Director of India's premier institution in Linguistics - Central Institute of Indian Languages, Mysore (2000-09). Currently, he is Chair-Professor, Dean (Faculty of Arts), and Head (Amity Centre for Linguistic Studies) at Amity University Haryana, India.