Lexical and Semantic Constraints in Compiling a Bilingual Dictionary from Tribal Language to English

Disha Bhatt, Minushree Pattnaik and Dr. Itishri Sarangi

Abstract--- Dictionary compiling is one of the most challenging tasks for Lexicographers. They have to focus on both logical and lexical semantics during the process of compiling the dictionary. The logical semantics includes sense, reference, presupposition and implication side of language whereas lexical semantics includes analysis of word meanings and relations between them. A good lexicographer strives to create a balance between logical and lexical semantics in order to compile a dictionary. The intensity of the challenge is mild in the case of monolingual dictionary but it shoots up when we have to compile a bilingual/trilingual dictionary. The reason behind this is that the differences in socio-cultural background come into the picture. The meaning of a word may differ because it varies from context to context, hence just jotting down literal meanings while compiling the dictionary may not solve the purpose. If we want to popularise tribal language then we have to produce a bilingual dictionary which will include English language. Compiling a dictionary from tribal language to English is a challenge because English is not the native language of the tribal community and in maximum cases the translator and lexicographer both are neither native speakers of that particular tribal language nor English. It is true that tribal language to English dictionary has been compiled but the credibility is at stake.

Keywords--- Dictionary Making, English Language Skills, Tribal Language.

I. INTRODUCTION

The state of Odisha has 62 tribes and 21 tribal languages. Out of these 21 languages only 7 languages have written scripts. The current government has weaved lot of schemes to preserve and revitalize tribal languages. Lot of proactive steps are being taken to document those languages so that their culture is kept vibrant and stored for the coming generations. Compiling dictionaries are an important form of documenting a language. The state government has encouraged both bilingual and trilingual dictionaries in order to bring tribal language to the global platform. But these things might sound good theoretically but implementing them in real life is a real challenge. Dictionary making is a pure academic activity and hence calls for different kind of writing skills compared to other types of academic writings. The lexicographer has to seek the help of a native tribal speaker who will work as a translator in order to compile a dictionary and may also know English but not necessarily will be a native/proficient speaker of English language. The reason behind this is that L1 is their mother tongue and English becomes L3 for the tribal learners. In maximum cases the transition between L1 and L3 is not smooth and hence the intervention of L2 is required which is Odia. But this pulls down their standard of English to a very great extent. Moreover, research says that people from different tribes have different socio cultural background which ultimately affects their

Disha Bhatt, Research Scholar, School of Humanities, KIIT Deemed to be University, Bhubaneswar.
Minushree Pattnaik, Research Scholar, School of Humanities, KIIT Deemed to be University, Bhubaneswar.
Dr. Itishri Sarangi, Associate Professor, School of Humanities, KIIT Deemed to be University, Bhubaneswar.

language skills. The transition between the languages is not smooth and hence the gaps can never be bridged. These gaps become the cause of constraints in the process of making dictionary and since there is no bilingual dictionary for tribal languages without written scripts then the process language learning becomes all the more difficult for the tribal population. Preserving, revitalising and documenting the tribal languages have become the need of the hour but not enough effort is being put in to achieve tangible results.

This paper deals with the study of the factors which pull back from producing a standard bilingual dictionary. It has been observed that dictionary skills demand good amount of lexical and semantic maturity and just adding a list of words along with their meanings is not enough and will not solve the purpose. The dictionary will be a futile from functionality angel. A good dictionary asks for equal amount of competence in both the languages but in this scenario the tribal learners are not competent in English and on top of it maximum tribal languages are oral in nature and hence don't have a written script. In the state of Odisha, Odia language (language of the State) is used to document tribal language. As a result of which the popularity of the tribal language remains confined to the state of Odisha only. It fails to reach the people staying outside of the state of Odisha. On the other hand dictionary making is considered a very good option to preserve and document tribal languages. The current study is based on Kalinga Institute of Social Sciences, KISS Deemed to be University a fully-free and fully-residential college for tribal students in the State of Odisha and nearby States like Bihar and North East. It is deemed to be the best option for the study because the students are native tribal speakers and are confident in Odia language which is used as a medium of documenting tribal language. However, the picture is not that perfect as it looks like. The students are imparted knowledge through English medium in higher education but still they are not confident in the language. The reasons is that the transition between L1 to L2 to L3 is not very smooth, the students are first generation learners in their community, socio-cultural and economic backgrounds, fear for the language, discomfort in using computers and technology, reluctance to come out of comfort zone so on and so forth. In a study conducted on Undergraduate tribal students lexical and semantic errors have been diagnosed to a very great extent.

Only a good trilingual dictionary can take tribal language to the world. It includes incorporating IPA, sense and meaning and putting in contextual meanings of the words rather than just a translated version of it. The dictionary needs to be more pictorial in nature. Changing trends in technology, apps and software need to see the light of the day in order to come up with an effective and efficient user friendly functional dictionary.

II. REVIEW OF LITERATURE

Research has constantly spoken about the gap between theory and praxis. There exists a relationship between theory and praxis. The terms lexical semantics or meanings of words talk about the interdisciplinary approach of the present study. All papers have few things in common and those are problems related to lexical semantic problems, theories, methodologies, and techniques. According to Lewis et al (2014) there exist 7,106 distinct languages in the world and out of which 2,303 are living languages which are currently being spoken in Asia, when compared to Europe where only 285 languages are spoken. They also stress on the fact that given the variety and diversity in Asian languages there is an urgent need of maintaining record, conducting more and more research on these issues and to preserve languages in the world. In addition to these points they also talked about collecting and preparing a

bank of words with the help of field work, compilation of dictionaries or preparing small glossaries of indigenous languages are the need of the hour. However, the challenge lies in the fact that indigenous culture is very oral in nature. Munday (2012) has opined that in case of bilingual and trilingual dictionaries skilled translators are required. Linguistic levels underline the presence of two levels form and meaning. Yong and Peng (2007) in their research have also talked about the importance and need of the study of form and meaning between the languages.

Translation studies have embraced recent computer technologies which will contribute in the use of parallel corpora, translation memory and lexical databases which is an important aspect of any professional translators' work Koehn (2009). In the context of tribal languages this holds lot of relevance because conserving and preserving has become very important and this is only possible through curating a platform which will address global problems of the lexicographers like how to preserve less-resourced and endangered languages and the use of newest technology to achieve the ends. These ends can be met when there is holistic, pluralistic and interdisciplinary dialogue with the changing world including computational linguistics, corpus linguistics, machine translation, translation memory, artificial intelligence, NLP and human language processing (text to speech and speech recognition), ICT are the ways and means to preserve and conserve indigenous languages and the right approach towards compiling a trilingual dictionary.

However, the above challenges are clubbed with language gap and rough transition between L1 to L2 to L3. Research on monolingual language acquisition brings to lime light that quantity and quality of language input in the child's surrounding influences and has an impact on the speed of language development. Hart and Risley (1995), identified a bunch of children with low language input as a result of low income of the family and low levels of parental education who at the age of 3 have heard 30 million lesser words than children their age from well-off families.

From the review of above literature it is very evident that factors like lexical and semantic errors, language gap, low vocabulary levels, low levels of parental education, lack of translators and interpreters hailing from non-tribal community, oral culture, no written script, socio-cultural differences contribute as pull-back factors from constructing a standard trilingual dictionary.

Survey

A survey was on the lexical and semantic errors committed by the tribal students pursuing undergraduate program (Arts, Science and Commerce) at Kalinga Institute of Social Sciences, Deemed to be University, Bhubaneswar. An error analysis was conducted of the students pursuing different streams (Arts/Science and Commerce). The questionnaire contained different sections for paragraph writing and narration skills. The sample size was 410 and the division on the basis of streams are – Science (boys – 21, girls – 78), Commerce (boys – 30, girls – 73) and Arts (boys – 124, girls – 84). However, out of 410 response sheets, only 344 response sheets were taken for analysis because they were found to be correctly filled in and were responded appropriately. Other 66 responses were not deemed fit for the study.

We observed that the English writing skills of Arts students is weak when compared to Science students whereas Commerce students have put an average performance. This study was conducted using Error analysis of content

Rosengren (1981). The rate of errors has been found maximum in Arts students and minimum in Science students. Maximum errors have been detected in the questionnaire in the case of Arts students whereas except for few grammatical errors least or minimum errors were observed in Science students. These errors were inclined towards lexical and semantic errors.

Table 2.1

TYPES OF ERRORS/ STREAMS		Arts	%	Science	%	Commerce	%
Lexical error	Structured	34	20.24	2	2.3	12	13.48
	Unstructured	4	2.38	0	0	4	4.49
Syntax error	Structured	73	43.45	38	43.68	44	49.44
	Unstructured	8	4.76	0	0	1	1.12
Lexical, Syntax error	Structured	23	13.69	19	21.84	20	22.47
	Unstructured	1	0.6	0	0	0	0
Syntax, lexical, Substance error	Structured	21	12.5	1	1.15	1	1.12
	Unstructured	0	0	0	0	0	0
Syntax, Substance error	Structured	4	2.38	25	28.74	6	6.74
	Unstructured	0	0	1	1.15	0	0
Grammatical error	Structured	0	0	1	1.15	0	0
	Unstructured	0	0	0	0	0	0
Substance, Lexical error	Structured	0	0	0	0	1	1.12
	Unstructured	0	0	0	0	0	0
TOTAL		168	100	87	100	89	100

Table 2.1 shows the frequency of different kinds of errors and numbers of structured and unstructured response sheets of students from different streams.

III. DISCUSSIONS AND ANALYSIS

The errors were divided into different categories. Five major types of errors were studied and were further divided into individual and combined errors. The individual errors included lexical and syntax errors, whereas the combined errors included lexical/syntax error, syntax/lexical/substance, syntax/substance, grammatical, substance/lexical. The study was clubbed with finding out numbers of structured and unstructured codes categories. The total sample size was 410 and 344 response sheets were taken for the study.

Keeping in mind the language and writing skills challenges faced by tribal students not much attention was paid to grammatical errors. The reason is that grammatical errors are bound to happen. Moreover, language learning theory says that just focussing on the grammar of a language, may make the language learning process all the more difficult. Groupings of errors were made instead of analysing single errors.

Analysis of Arts Students

In the case of Arts students it was observed that 20.24% was lexical errors but response sheets were structured, whereas 2.38% of the response sheets were unstructured. Syntax errors and structured response sheets were 43.45%, whereas unstructured lexical was 4.76%. Lexical/syntax errors and structured response sheets were 13.69% whereas lexical/syntax unstructured was 0.60%.

Syntax/lexical/substance errors with structured response sheets were 12.50% and no syntax/lexical/substance errors with unstructured responses. Syntax/substance errors with structured response sheets were 2.38%, whereas

0% syntax/substance errors with unstructured responses were observed. No substance/lexical errors were found in Arts students.

Analysis of Science Students

The lexical errors with structured response sheets were 2.30% and no lexical errors with unstructured response sheets were found in Science students. Syntax errors with structured responses were 43.68%, whereas no syntax errors with unstructured response sheets were found. Lexical/syntax errors with structured responses were 21.84%, whereas no lexical/syntax errors with unstructured response sheets were found. Syntax/lexical/substance errors with structured responses 1.15%, whereas no unstructured response sheets were found. Syntax/substance errors structured responses 28.74%, whereas unstructured with errors were 1.15%. Grammatical error 1.15% and no substance/lexical errors were observed.

Analysis of Commerce Students

In Commerce students 13.48% lexical errors with structured responses, whereas 4.49% errors with unstructured responses. 49.44% syntax errors with structured response sheets and 1.12% errors with unstructured response sheets were observed. Lexical/syntax errors with structured response sheets were 22.47% and no errors with unstructured response sheets were found. Syntax/lexical/substance errors with structured response sheets were 1.12% and no errors with unstructured response sheets were observed. Syntax/substance errors with structured response sheets 6.74% and no other errors were found except for substance/lexical errors with structured response sheets.

Significance of Errors

It has been found that improper use of rules of writing makes the text very unimpressive to read and on top of it speaks low about the language proficiency of the respondents, capitalization/punctuation and spelling are key to understanding the text and it helps in understanding the text with proper pauses used while writing sentences, avoiding spelling errors may narrow down the degree of confusion in similar sounding words, using capital alphabets distinguishes proper nouns, forming words and selecting words can bring in lot of difference in the sense and meaning of the text, sentence structure/ordering/subordination and coordination is the key to writing a correct or flawless sentences and all the above mentioned errors add up to semantic errors. The reason is that because of all the 4 errors mentioned above, the 5th error is generated. The rest of the errors lead to ambiguity in the text which is generated. It also leads to miscommunication. The writer fails to convey his or her thoughts. This is a clear indication of gaps in competence and performance. Knowledge of differences in the contexts which lead to changes in meanings of words is very important when it to comes to developing a trilingual dictionary. Differences in sociolinguistic aspects also play a considerable role in this regard.

The amount of lexical, semantic errors committed by the students is clear indicators of poor knowledge of words in the tribal community. As mentioned earlier tribal languages are known only to tribal people and very less percentage of other population speaks any of the tribal languages. This proves that only tribal can contribute to building of a trilingual dictionary by building bridges between tribal languages and English. It is their window to the world and making their language, history, culture and heritage to different places.

IV. SUGGESTIONS AND CONCLUSION

KISS provides mother tongue based multilingual education which is solves the purpose of transiting from mother tongue to Odia but not from Odia to English. This program is implemented only at early childhood level but fails to address the next level of difficulty faced by the students at University level who need English for different purposes. A similar program should be designed to address the issue until them constructing a trilingual dictionary will be a far-fetched dream in the state of Odisha. Many programs have been launched like English Access Program run by US Embassy but an exclusive vocabulary building program is the need of the hour. A pictorial dictionary in digital form and requisite training catering to different tribal languages are needed. It is only then the tribal language can be brought to the forefront using English as a medium.

REFERENCES

[1]   Mieszkowska, K.,Luniewska, M., Kolak, J.,Kacprzak, A.,Wodniecka, Z., & Ewa Haman, E. (2017).Home Language Will Not Take Care of Itself: Vocabulary Knowledge in Trilingual Children in the United Kingdom. *Frontiers in Psychology,* 8, 1358.
[2]   Koehn, P. (2009). Statistical Machine Translation. *Cambridge: Cambridge University Press.*
[3]   Munday, J. (2009). Introducing Translation Studies: Theories and Applications. *Routledge.*
[4]   Yong, H., & Peng, J. (2007). Bilingual Lexicography from a Communicative Perspective. *John Benjamins.*
[5]   Hart, B., & Risley, T.R. (1995). Meaningful differences in the everyday experience of young American children. *Paul H. Brookes Publishing.*
[6]   Rosengren, Karl Erik, ed. Advances in Content Analysis. Vol – 9. *SAGE Publications, Incorporated,* 1981.
[7]   Bhatt, D., & Sarangi, S., (2019). Difficulties Associated with Fostering Education to Tribal Population Through English Language Skills. Vol – 11, 05-Special Issue. *Journal of Advanced Research in Dynamical & Control Systems.* Pp. 1159-1170.