

Muskan, Ph.D Research Scholar CALTS University of Hyderabad Hyderabad, Telangana India

Abstract

The present research work attempts to analyze Multi Word Expressions in Khorthia language. The main objective of the present paper is to analyze different types of MWEs present in the language. Khorthia is widely spoken in Hazaribagh, Dhanbad, Ranchi, Koderma, and Giridih districts of Jharkhand, a state in the Eastern part of India and it belongs to Indo Aryan language family. According to census (2011), it has 8.04 million native speakers and it has also got the status of official language of the state of Jharkhand in India. In this research paper, attempt has been made to identify and analyze different types of nouns, adjectives etc which are observable in the language. It also investigates how they are equally important for forming word formation process in Khorthia. The data for the analysis have been collected from Hazaribagh district of Jharkhand and it includes both primary and secondary data for the analysis. The findings of the study reveal the fact that modern Khorthia is highly influenced by different languages like Magahi, Maithili, Bengali and Hindi.

Keywords: MWE, Khorthia, Indo Aryan Language Family

INTRODUCTION

Jharkhand is a multi-ethnic, multi-cultural and multi-lingual state.

According to Census 2001, Khorthia is officially recognized as the second most populated language of the state Jharkhand after Hindi with a total number of 4725927 speakers using Khorthia as their first language or the mother tongue. The Census 2011, reports approximately

80 million speakers of Khorthia. The word Khorthia has itself several names like Khotta, Khattahi, Magahi, Magahia, Kurmali, Thar and many more. According to (Prasad and Shastri 1958:9) the name Khorthia itself refers to a corrupt or impure form of a language. Khorthia is spoken in the Indian state of Jharkhand, mainly in 16 districts of three divisions: North Chotanagpur, Santhal Pargana and Palamu. The 16 districts are Ramgarh, Hazaribag, Chatra, Giridih, Koderma, Bokaro, Dhanbad, Dumka, Godda, Deoghar, Pakur, Jamtara, Sahebganj, Latehar, Palamu, Gadhwara (Ohdar).

It is one of the four major Saadani (i.e., non-native) language varieties (i.e., Khorthia, Kurmali, Panchparganiya and Nagpuri). Saadani languages acts as a link language among the adivasi communities.

Khorthia speakers come in contact with the speakers of other communities and they show the features of multilingualism. A minute and careful observation of the language will show a number of linguistic influences on Khorthia. It can be assumed that the language is highly influenced by Bengali, Magahi, Maithili, Hindi.

Khorthia is written in Devanagari script.

MULTI-WORD EXPRESSIONS

My work is on MWEs in Khorthia which I have identified by interviewing people.

Multi Word Expression (MWE) is one of the key terms used in NLP and Computational Linguistics (CL) over the years. Since there is no generic definition given for MWE yet, but there are different views from many authors on this important term. The term MWE refers to the combination of two or more independent words which individually convey some meaning, but when used together convey a different meaning. So, based on this concept there are various types of MWEs, which include Compound nouns compound verbs, compound (Revised Manuscript Received on June 15, 2019) .(Rakhi Joon, Department of Computer

Science, University of Delhi, Delhi, India, Archana Singhal, Department of Computer Science, IP College for Women, University of Delhi, Delhi, India). .adverbs, noun+verb, named entities, idioms, phrases, collocations and phrasal verbs. Earlier, before 90s, MWEs was used and studied under the phraseological unit only, but later in NLP, MWEs have gained a lot of attention. In every language,

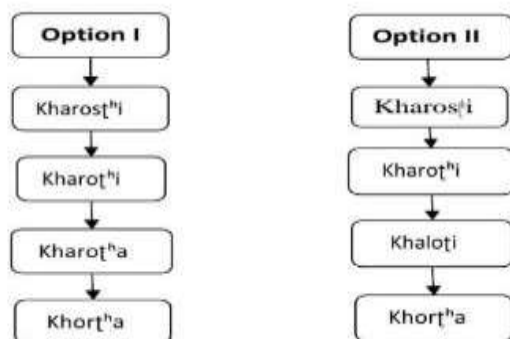
MWEs are used frequently in many practical applications of NLP and CL. In 2001, the project “Centre for the Study of Language and Information” (CSLI) at Stanford University was a milestone in the area of NLP as well as MWEs. The main focus of the paper was to encode different MWEs into their precision grammar. Later various dictionary resources and annotations tools were built for MWEs. And thus the journey of MWEs begins with such a historical background in the field of NLP. For a particular language, MWEs plays an important role as in formation of a sentence. The formation and categorization vary from language to language as the grammar rules are also considered for the formation of Multiwords. MWEs have gained popularity in many languages including English, Chinese, Hindi and many other Indian Languages. The language used for MWEs is Hindi, which is morphological rich language like most of other Indian languages and thus provides a good platform for MWEs. The text processing in a particular language depends on the grammatical constructs along with the linguistic and syntactic properties of the language. The formation of MWEs mainly deals with the combination of two or more meanings of the words rather than the words itself. These words can be any of the grammar constructs like Noun, Pronoun, Verb, Adverb, Adjectives, and so on. POS taggers mainly affect the representation of a particular construct in any language. The combinations of Hindi POS tags are considered for the experiments and results.

Examples of Hindi MWEs

Adj+Noun (साक्षात देवी - Saakshat Devi); Adj+Prep (बराबर वाला - Brabar vaala)
Adv+Adj (बिल्कुल ल गुबिया-सी - bilkul gudia si); Adv+Adv (शायद बिर - shyad fir) etc

ETYMOLOGY

According to A.K. Jha and Krishna Chandra Das Ala, Khortha has some relations with Indus Valley Civilization. The most common meanings of the Khortha according to Ohdar are regional, vernacular, rural and so on. According to A.K. Jha, Khortha has evolved from the ancient script Kharosthi. (Kharosthi > Kharothi > Kharotha > Khortha (written by Dangi)).



Evolution chart of the Kharosthi into 'Khortha' (written by Ohdar 2018).

Another important claim about the etymology of the word Khortha is made by Krishna Chandra Das Ala in his work Khortha Bhasha Grahane (Ohdar) where it is pointed out that the word Khortha has not evolved from Kharosthi, instead it has evolved from Khartha which means 'nature' in Khortha. Khartha and Khortha sounds more identical than Kharosthi and Khortha, Das took this option. The majorities of researchers including Ohdar agree with the claim made by Jha and relate Khortha to Kharoshti.

GEOGRAPHICAL LOCATION OF KHORTHAS

According to Dangi, the language is spoken in the fifteen districts that are on the northern side of the Tropic of Cancer that runs nearly through the middle of the state in north- south dimension (approximately 23°58'10" north longitude to 25°19'15" north latitude and 83°20'50" to 88°4'40" at east latitude). Dangi also claims that language is not used on the southern side of this imaginary geographical line.

Geographically, Khortha language covers above 31,000 sq. Km in its nucleus.(Sahu 2012;

Shivdayal Singh 'Shivdeep')

Language:- Khortha

Region:- Bihar, Jharkhand, West Bengal

Country:- India

No. of native speakers:- 4.7 million (2001 census report)
8.04 million (2011 census report)

The language is spread across 48758.37sq. Km in the fifteen districts of Jharkhand (Census 2011; Dangi). The total number of native speakers are 19136997. Nearly 61.16% of the total population living in these districts speak this language. (Dangi)

TYPES OF KHORTHAS AND THEIR SPOKEN AREAS (Ohdar)

1.SIKHARIYA :- Jainamora, Parasnath Pahad, Topchanchi, Giridih up to Eastern Bengal borders. Influence of Bengali is found.

2.GOLWARI:- South of Ramgarh area from Kasmar, Kherachatta, Gola, Rajrappa etc.

3.RAMGHARIA:- Ramgarh, Hazaribag, Bhurkunda, Eastern Patratu etc.

4.DESHWALI:- Places within the Patratu border to the western side of Palamu as in Daltanganj, Lohardaga etc.

5.PARNADIYA:- The northern side of Damodar River in and around Nai Sarai River situated opposite to Ramgarh, Hazaribag and Chatra. Also known as Hazaribagiya. Influence of Magahi is found.

6.SANTHAL PARGANIYA:- Deoghar, Madhupur, Dumka and other Northeastern areas of Jharkhand. Influence of Maithili and Santhali are found.

SOME RELATED WORKS ON KHORTHAS

Since the early days of independence few works have been written in the language and published. Grierson's Linguistic Survey of India too did not mention Khortha language. 1950 was the beginning year where one can observe publications on the language. Panuri's Meghadutam (1950) was the first work that concentrates on translation. He translated Kalidasa's Meghadutam into Khortha which created an impact on the native speakers to be creative in writing love poems in Khortha. After a stipulated period of time Mahto (2015) tried to create a script for Khortha language and succeeded in creation of Khortha script which includes 36 alphabets. In the same year Dangi (2015) wrote a monograph titled Khortha bhasha ek parichay. The text deals with the basic introduction of Khortha, its etymology and geographical location. Subsequently, Ohdar (2018) wrote another text titled Sampurna Khortha Bhasha avam Sahitya. This text deals with the introduction to Khortha language and few grammatical aspects. In the same year Priya and Singh (2018) wrote another text from a sociolinguistic perspective. The title of the text is "Khortha, a Dying Language and Urgency to Retain its Pure Variety". Recently, Dash (2020) wrote a theoretical paper on 'Language Attitude of Khortha speakers in Giridih: A Survey Report'. The paper focuses on language shift and maintenance. In the same year Aman, et.al (2020) published a paper on Designing a Linguistic profile of Khortha: A less resourced language spoken in the state of Jharkhand. It deals with the linguistic outline of Khortha language. Netra et.al. (2021) is another work titled 'How one language became four: the impact of different contact- scenarios between "Sadani" and the tribal languages of Jharkhand'. This paper deals with the different linguistic varieties spoken in Jharkhand.

DATA COLLECTION

According to the data elicited I got to know about the special types of words incorporated in the folktales, stories also with the pronunciation of the folktales enabled me to look into the Vowels and Consonants as well as the Person, Number, Gender of the Khortha language being spoken.

I asked the informants questions like *tum, mai, tmhara, hamara ko kya bolte*

Khortha me?, Kitne swar hai?

Kitne vyanjan?

Vakeel kaise bolte?, Udaaharan ko kya bolte?

Lokgeet ka matlab kya hota?etc

MULTI-WORD EXPRESSIONS EXAMPLES

- **पटतार:- उदाहरण (udaaharan)**
patataar
पट:- मौर ; तार:- उतारना
pat:- maur taar:- utarna
पट:- बिहा में बोर पट पिनहो हइ।
biha me bor pat pinaho hai.
शादी में वर मौर पहनता है।
shaadi me var maur pehnta hai.
तार :- पट के सम्मान के साथे तार जा हइ।
pat ke samman ke saathe taar ja hai.
मौर को सम्मान के साथ उतारा जाता है
maur ko samman ke sath utara jata hai.
E.g.:- राम रावण के युद्ध समाजिक लेल एक बेहतरीन पटतार हइ।
Ram Ravan ke yudh samajek lel ek behatareen patataar hai.
राम रावण का युद्ध समाज के लिए एक बेहतरीन उदाहरण है।
Ram Ravan ka yudh samaj ke liye ek behtareen udaaharan hai.
- **घरघुसना (gharghusana):- डरपोक (darpok)**
घर:- घर ; घुसना:- छिप जाना
ghar ghar ghusana chhip jana
घर:- हाम घरा जाइ रहल हियो
ham ghara jai rahal hiyo.
हम घर जा रहे है।
hum ghar ja rhe hai.
घुसना:- नयकी बहुरिया बोड़ के देख के घरे घुस गई।
nayaki bahuriya bod ke dekh kar ghare ghus gayi.
नयी बहू बड़ो को देखकर घर में छिप गई।
nayi bahu bado ko dekh kar ghar me chip gayi.
E.g.:- इ देखइत रहलय घरघुसना के ।
I dekhit rahlay gharghusana ke.
ये देख रहा था वो डरपोक को।
Ye dekh raha tha vo darpok ko.
- **अड़वाअर/ अड़वाइर :- वैसी जगह जहां जानवर डाड़ खेलते हैं।**
adwair:- waisi jagh jahan janwar daad khelte hain.
अड़:- एक जगह पे डटे रहना/ अड़े रहना ; वाइर:- अलग होना/भागना
ad ek jagh pe date rehna/ ade rehna wair alag hona/ bhagna
अड़:- ऊ अपन जगह पर अड़ल हइ।
Oo apan jagh par adal hai.
वह अपने जगह पर अड़े हैं।
vah apane jagh pe ade hai.
वाइर:- गरू इधर उधर वाइर गेलथी।
garu idhr udhr wair gelathi.
जानवर इधर उधर भागने लगे।
janwar idhr udhr bhagne lage.
E.g.:- गरू- डांगर के जमा होवे वाला सथल के लेल अड़वाइर बनल।
garu dangar ke jama hove wala sathal ke lel adwair banal.

जानवर (गाय- बैल) के जमा होने के लिए अड़वायर बनाया गया है।
janwar (gaay bail) ke jama hone ke liye adwayar banaya gaya hai.

- जमदूत:- एक जगह पर अड़े रहना/ डटे रहना
Jamdut ek jagh par ade rehna/ date rehna
जम:- अशुभ होना ; दूत:- दूतावास
Jam:- ashubh hona dut:- dutawas
जम:- सोहराइ से जमे बाहर निकला हइ।
sohrai se jame bahar nikala hai.
दीवाली में जम का दिया निकलता है।
Diwali me jam ka diya nikalte hai.
दूत:- ई सिपाही के दूत हइ।
sipahi ke dut hai.
यह सिपाही का दूत है।
yah sipahi ka dut hai.
E.g.:- ई मने जमदूत बन हइ।
Ee mane jamdut banal hai
ये जमदूत के रूप में अड़ा है।
ye jamdut ke roop me ada hai.
- बरवाअड्डा:- वह स्थान जिस जगह अखारा लगता है।(पंचायत)
barwaadda vah sthan jis jagh akhara lagata hai. (panchayat)
बरवा:- जगह/स्थान अड्डा:- बैठकी/ जमावड़ा
barwa jagh/sthan adda baithaki/ jamaavada
बरवा:- कवि के जनम भूमि के बरवा नाम देल गेलइ।
kavi ke janam bhumi ke barwa naam del gelai.
कवि के जन्म भूमि को बरवा नाम दिया गया।
kavi ke janm bhumi ko barwa naam diya gaya
अड्डा:- रचनाकारे सभे दुवारे अड्डा लगेलथिन हइ।
rachnakare sabhe duware adda lagelthin hai.
सभी रचनाकार दरवाजे पर अड्डा लगाए है।
sabhi rachnakaar darwaze par adda lagaye hai.
E.g.:- श्रीनिवास पानुरी के समाजेक लेल बरवाअड्डा के आयोजन सथल बनवल गेल ।
Shriniwas Panuri ke samajek lel barwaadda ke ayojan sathal banwal gel.
श्रीनिवास पानुरी जी के सामाजिक काम के लिए बरवाअड्डा (पंचायत) आयोजन स्थल बनाया गया।
Shriniwas Panuriji ke samajik kaam ke lie barwaadda ayojan sthal banaya gaya.

CONCLUSION

Khortha is a tribal language and is not only spoken by the Sadaanas but is also used by the Adivasis as a link language. It is the most spoken languages of Jharkhand.

Multi-word expressions which were in use some time back in Khortha are not in use and less prevalent in the modern Khortha as well as Hindi has taken its place.

A grave necessity is felt for the preservation of the Khortha Language which has a majority stake in the languages of Jharkhand. The lesser prevalence of Khortha language can be ascribed to the migration of native people of Jharkhand to other states for jobs, marriages being held in different parts of the country of the Khortha people as well as lesser used medium in schools which is the birth place of one's linguistic repertoire.

Language travels from one generation to other but it's not fully applied to Khortha it doesn't reach to the children and it is losing its identity because nowadays parents don't communicate with their children in Khortha.

Khortha speakers come in contact with the speakers of other communities and they show the features of multilingualism. There are number of linguistic influences on Khortha. The language is highly influenced by Bengali, Magahi, Maithili, Hindi.

There are lots of linguistic variation that can be observed in Khortha like pidgins, creoles, code switching, code mixing etc.

Measures suggested for preservation:

Some measures should be taken to safeguard Khortha like motivational teaching aid should be developed for children as well as for elders to learn Khortha. Nowadays elders are learning Khortha for the purpose of competitive exams.

More measures should be taken by the government to preserve the language by offering job opportunities in Khortha because it's the mother tongue of 8.04 million people. Students should be taught Khortha in schools to safeguard the language.

Future scope:

There is much future scope of linguistic work in this language as works can be done on different varieties of Khortha and also showing similarities and differences with each other in different aspects like morphology, syntax, semantics etc.

REFERENCES

Books

- Dangi A. K. (2012). Khortha Bhasha: Ek Parichay, Jamshedpur: Spardha Publication.
- Kachru B.B., Kachru Y., Sridhar S.N. (2016). Language in South Asia. Cambridge: Cambridge University Press.
- Mahto, N. (2015). Khortha lipi aur vyakarana. Bokaro: Nav Jagriti manch Publication
- Ohdar, B.N. (2007). Sampurna Khortha; Bhasha Evam Sahitya, Ranchi, Jamshedpur: Margdarshan Publication.
- Prabhakar G.M. (2020). Khortha Bhasha evam Sahitya. Ranchi: Safal Publication.
- Sahu, C. (2012). Encyclopedia of Indian Tribes. New Delhi: Sarup Book Publishers.

Articles

- Aman A., Dash N.S., Chakraborty J. & Gope A. (2021). Consonants in Khortha. In: Tariq Khan (ed.) Alternative Horizons in Linguistics: A Festschrift for Prof. Panchanan Mohanty. Chapter 23, pp: 311-329, Germany: Lincom Europa Press.
- Aman, A, Dash, N. S. and Chakrabarty, J. (2020) Designing a Linguistic Profile of Khortha: A Less-Resourced Language Spoken in the State of Jharkhand, India. Dialectologia, Vol. 25, No. 2. Pp: 25-43.
- Bhowmik, R. (2019), Language contact in the numerals of the Malda variety of Khortha. International Bilingual Journal of Culture, Anthropology and Linguistics. Pp. 75-84.
- Dash, N. S. (2016). Language Attitude of Khortha Speakers in Giridih: A Survey Report, Journal of Advanced Linguistic Studies. Vol. 5. No. 1-2. Pp. 147-173.
- Dash, N. S. and Aman, A. (2015). Generation of a Dialect Corpus in Khortha used in Jharkhand, India: Some Empirical Observations and Theoretical Postulations, Journal of Advanced Linguistic Studies. Vol. 4. No. 1-2. Pp. 151-171.
- Joon, R, A Comparative Analysis of Hindi Multi Word Expressions using Relevance Measure- RMMWE.
- Kumari, A. (2021), Folklore and Language: Ergativity and Agreement of Khortha Language in Jharkhand. International Journal of Humanities, Engineering, Science and Management (IJHESM), Volume II, Issue – I, pp. 35-44.
- Priya, S. & Singh, R. (2016), Khortha or Khotta: An Endangered Language of India and the Urgency to Retain its Pure Variety. International Journal of Humanities and Cultural Studies. Vol.3, No. 3, pp. 190-201.
- Singhal, A, A Comparative Analysis of Hindi Multi Word Expressions. Using Relevance Measure- RMMWE- 3437. BEIER No. H6869068819 & Sciences Publication.

- Web pages

Census (2011) Office of the Registrar General and Census Commissioner, India.

<https://censusindia.gov.in/2011-common/censusdata2011.html>