

Hindi to Chhattisgarhi Translator

¹Shubham Kumar sahu, ²Anupriya dutta, ³Manish Kumar sinha, ⁴Sachin ther
Department of Information Technology, Bhilai Institute of technology, Durg, Chhattisgarh, India

Abstract: - A natural language translator translates one language into another with the help of certain logic, rules and algorithms. Our work will translate Hindi language to Chhattisgarhi and vice-versa. We will also need a POS tagger for this which we have already created earlier and for making this work correctly we have taken help from linguistic expert from a university in Chhattisgarh who teaches Chhattisgarhi there. We have used an online tool LingoJam to make possible our translator is completely online and tool based.

Keywords:-Translator, Chhattisgarhi, Hindi, lingoJam

I. INTRODUCTION

To make a translator first we need to understand the basic structure of sentence of both the languages, here the best part is the script of the both the languages is same that is Devanagari, and the structure of the sentence is also 70-80% is the same so we don't need a language parser for making a translator here, and only few parts are there where structure of the sentence is not same, there the lingoJam tools provide re-ordering and arranging of sentences but for that the words should be tagged and for that we can use a POS(parts of speech) tagger for Chhattisgarhi and Hindi language both. Our work is completely new and nothing has been done for Chhattisgarhi language

1.1 ABOUT LINGOJAM

It's an Open Source Toolkit for Statistical Machine Translation. It is developed by Australian company. Apart from providing an open-source toolkit for SMT, a further motivation for LingoJam is to extend phrase-based translation with factors and confusion network decoding

II. DETAILS OF TOOL AND USING METHOD

This tool can be easily found on the search engines and Google, the URL of the tool is <https://lingoJam.com> where anyone can go and create a translator

2.1 USING PHRASES OPTION

Phrases are exchanged first. Place phrases from language 1 (usually just plain English) into the first column, and then place what you want it to change into in the second column. Usually for Hindi to Chhattisgarhi translation basically where we need to replace one word with more than one words or vice-versa we use this option examples in the image below

रहता हु	रथो
खेल रहा	खेलत
तुमसे	तोर ले
करता हु	करथो
जा रहा हु	जाथो
नमस्ते	राम-राम
नाचता है	नाचथे

2.2 USING WORDS OPTION

These two lists will be the foundation of your translator. If your users type a word that's in the first column, it will be translated to the word in the second column. Remember: Words at the top of the list will be the first ones translated (after phrases). Press Enter to make a new line for each new word. Here's a list of common English words in case you need it. For Chhattisgarhi as I mentioned earlier for 70-80% of the language the structure is the same so what we need to do is just replace the words of Hindi for that exact same word existing in Chhattisgarhi for this words option of this tool is helpful examples in the image below

मेरा	मोर
शादी	बिहाव
समय	फुर्सत
मिला	मिलिस
तुझे	तोला
मेरे	मोर
लिए	बर
मिला	मिलिस
नाम	नाम
बहन	बहिनी
है	हरे

2.3 USING ORDERING OPTION

(This is a new feature, please report bugs and feel free to make suggestions!) Use this section to change the ordering of words based on their tagged word type. Tag your words by appending {{noun}}, {{verb}}, {{adjective}}, or any other tag you want (e.g. {{fruit}}, {{animal}}) to the end of words in your word list. Then you can switch the ordering from (for example) adjective->noun to noun->adjective by putting "adjective noun" in the left box and "noun adjective" in the right box (without the quotation marks). IMPORTANT: You should only add {{tags}} your language 2 words (the box on the right hand side in the 'words' tab) if you only want swapping to occur in the forward translations, otherwise tag both columns. For Hindi to Chhattisgarhi, Hindi tagger is available online and for Chhattisgarhi we have made an automatic POS tagger for Chhattisgarhi, which will help this to tag each word and this ordering and tagging, is the part on which we are still working to make it more accurate.

There are more other options on this tool but we are not using that we will only discuss on options which we are currently using. With all the options above discussed we can successfully make a Hindi to Chhattisgarhi Translator and the only extra thing needed is POS tagger and database to feed in to the tool which we are making, currently the translator is working but the size of database is small and we are increasing it day by day

III. WORKING OF TRANSLATOR

After knowing about the tools and its various options and how to use it and which option are we going to use which will fulfill our needs we just need to feed the data in the options discussed above and what type of data is to be feed is also shown in the images above. So basically what our translator does is 3 things and that also sequentially one by one. We can now them by menus that is (1) phrases (2) words (3) ordering

3.1 TRANSLATION METHOD

The phrases option works at the start it finds the words that are there in its databases and replaces it with words which are in the next column, then comes the time for words option this option finds the sentences for the words which are in its databases of first column and replaces it with the words which are in the second column, then the last option works the ordering option according to tags given to each language for both the languages the rule for reordering is to be set and then after replacing the words to make it grammatically correct the words need to be re-ordered, example of such a rule to be reordered is

NP NP V GF ⇒ NP V GF NP.

This kind of rules is to be set to make it correct at the end and for various kinds of sentences various such rules is needed to be feed on the tool on which we are currently working.

This three options when works one by one in correct order that makes our translator works correctly and the order of their working is (1) phrases (2) words (3) ordering

IV. RESULTS AND DISCUSSION

URL of our Hindi to Chhattisgarhi translator is <https://lingojam.com/hinditochhattishgarhi>, currently our translator is 70-75% accurate we are making our database large to make it more accurate, see the results and translated examples in image

HINDI TO CHHATTISHGARHI

contact sksdbest@gmail.com for details

मेरा नाम शुभम है
मैं भिलाई में रहता हू

मोर नाम शुभम हरे
मे भिलाई में रथों

Currently we tested our translator with this following lines and it worked correctly

- 1 मेरा नाम शुभम है
- 2 मैं देल्ही मे रहता हू
- 3 मेरा पति सब्जी लेने बजार गया है
- 4 मुझे टमाटर की सब्जी बहुत पसन्द है
- 5 वह लड़का खेल रहा है
- 6 मैं कल आउगा
- 7 राम मेरा दोस्त है
- 8 मैं तुमसे बहुत प्यार करता हू
- 9 क्या खाना खाया तुमने
- 10 मैं निमन्त्रण देने जा रहा हू
- 11 नमस्ते चाचा
- 12 वह लड़का अच्छा नाचता है
- 13 ये मेरा किताब है
- 14 कल मेरी बहन का शादी है
- 15 अब समय मिला तुझे मेरे लिए



V. CONCLUSION

Our translator successfully translates the sentences which are feeded in the tool in its database and other sentences which are not feeded its unable to translate therefore we need a bigger corpus and a bigger database of replacing exact words, otherwise the tools and the set of rules and algorithm that we made are very much accurate to translate we made only one side rule and the other side translation is done by the tool its self.

VI. FUTURE SCOPE

Speech based translation system could be made in future using our translator and this translator is basically rule based translator to make it more accurate a translator can also be made using a language parser and using various many more natural language processing tools

VII. REFERENCES

- [1] [Avramidis2012] Avramidis, E. (2012). Quality estimation for machine translation output using linguistic analysis and decoding features. In Proceedings of the Seventh Workshop on Statistical Machine Translation, pages 84–90. Association for Computational Linguistics.
- [2] [Bharati et al.2006] Bharati, A., Sangal, R., Sharma, D. M., and Bai, L. (2006). Anncorra annotating corpora guidelines for pos and chunk annotation for indian languages. LTRC-TR31.
- [3] [Bojar et al.2013] Bojar, O., Buck, C., Callison-Burch, C., Federmann, C., Haddow, B., Koehn, P., Monz, C., Post, M., Soricut, R., and Specia, L. (2013). Findings of the 2013 workshop on statistical machine translation. In 8th Workshop on Statistical Machine Translation.
- [4] [Bojar et al.2014] Bojar, O., Buck, C., Federmann, C., Haddow, B., Koehn, P., Leveling, J., Monz, C., Pecina, P., Post, M., Saint-Amand, H., et al. (2014). Findings of the 2014 workshop on statistical machine translation. In Proceedings of the Ninth Workshop on Statistical Machine Translation, pages 12–58. Association for Computational Linguistics Baltimore, MD, USA.
- [5] [Callison-Burch et al.2011] Callison-Burch, C., Koehn, P., Monz, C., and Zaidan, O. (2011). Findings of the 2011 workshop on statistical machine translation. In Proceedings of the Sixth Workshop on Statistical Machine Translation, pages 22–64, Edinburgh, Scotland, July. Association for Computational Linguistics.
- [6] [Callison-Burch et al.2012] Callison-Burch, C., Koehn, P., Monz, C., Post, M., Soricut, R., and Specia, L. (2012). Findings of the 2012 workshop on statistical machine translation. In Proceedings of the Seventh Workshop on Statistical Machine Translation, pages 10–51, Montréal, Canada, June. Association for Computational Linguistics.
- [7] [Chiang2007] Chiang, D. (2007). Hierarchical phrase-based translation. *computational linguistics*, 33(2):201–228.
- [8] Gurpreet Singh Josan, Punjabi to Hindi machine translation system , COLING '08 22nd International Conference on Computational Linguistics: Demonstration Papers
Pages 157-160