

PARAPHRASE GENERATION

(CL2 Project)

By- Pranav Goyal



What's a paraphrase

- Two sentences that have common proposition
- Two sentences that entail each other
- Two sentence that have identical set of entailments
- “Mean the same”

Dataset Used


'Hindi/Urdu Dependency Treebank' (Bhatt et al., 2009) (http://ltrc.iiit.ac.in/hutb_release/)

- SSF Format (I used an SSF-api)
- Divided the sentence in chunk/nodes
- Major classes:
 - SSFInfo (sentenceList, globalWordList)
 - Sentence (chunkList, wordNumList, nodeDict)
 - Node (nodeName, nodeRelation, nodeRelation, childList, chunkNum)
 - Chunk (chunkTag, chunkNum, nodeName, wordNumList)
 - Word (wordTag, word, featureSet)

Main Approach

I separately created paraphrases using the three general strategies:

- Synonym/Antonym replacement
- Re-arrangement
- Active to Passive voice



Synonym/Antonym Replacement

Main Approach

- Paryayvachi and Vilom Shabd list scratched from hindistudent.com
- Basically replacing a word with its synonyms/antonyms if found in the scratched dictionary
- Got 190 set of paryayvachi shabds and 750 vilom shabds
- Replaced word should not be a Proper Noun (अनिल कुंबले)
- Replaced word should not be part of a compound word (Word tag should not end in C)

82	जीवन	जिदगी, पराण, जीविका, वायु, पुत्र, जल, जान, हयात
83	झंडा	ध्वज, ध्वजा, पताका, केतन, निशान, केतु, परचम
84	डाली	डाल, शाखा, टहनी, पक्ष
85	डर	भय, भीति, त्रास, खौफ, साध्वस, अंदेशा, आतंक, विभीषिका
86	चोर	आखु, खनक, तस्तर, रजनीचर, चोरटा, चोरकट, चोरक
88	ठाकुर	हजाम, ब्राह्मण, परमेश्वर, देव
89	तरंग	लहर, लहरी, कल्लोल, ऊर्मि, वीचि, हिल्लोल
90	तलवार	असि, करवाल, कृपाण, खंजर, खड्ग, चंद्रहास
91	तमाशा	खेल, कौतुक, कौतूहल, कुतुक, कुतूहल

26	अघ	अनघ
27	अमावस्या	पूर्णिमा
28	अचल	चल
29	अमृत	विष
30	अर्जन	वर्जन
31	अर्थ	अनर्थ
32	अल्पसंख्यक	बहुसंख्यक
33	अति	अल्प
34	अर्थ	अनर्थ
35	अलम्य	लम्य

Negation of VM in antonym replacement

- Assuming antonym replacement result in a proposition which is not of original proposition.
- Negation of head VM to get back the original proposition.
- No CCOF should be present on the way to VM
- “नहीं” added before VM if not present, else deleted

Ex- “राम अच्छा बच्चा है” -> “राम बुरा बच्चा नहीं है”

Inflected forms of replacement

- The scraped list only contains words in root form
- Made two function : morph and try_apply_morph
- Morph results in the minimum addition and deletion required to convert an original word to inflected form
- Try_apply_morph tries if possible to apply the same deletion and addition to another word
- So, the inflections of original words were applied on the replaced antonyms and synonyms

Limitation

- Multiple senses of word -> may not match the sense referred in the synonym/antonym list
- Antonym replacement and verb negation often doesn't result in original proposition



Re-arrangement

Main Approach

- Using the fact that Hindi is a Free word order language.
- Made a tree where the children could freely permute
 - Head position will not change. For ex- A B & C -> C B & A
 - Made the symbol the initial head and the verb a child of it
- Fragof and pof were first combined with their head
 - Ex - “अखिल भारतीय कांग्रेस कमेटी (एआईसीसी) के”
 - Ex - “विचार - विमर्श किया जाएगा”
- ‘कि’ was deleted
- Finally generated random permutations of the tree



Active to Passive

(कर्तृवाच्य से कर्मवाच्य में वाच्य परिवर्तन)



Main Approach

- Rules Followed (from [here](#)) :
 - कर्तृवाच्य के कर्ता के साथ यदि कोई विभक्ति लगी हो, तो उसे हटाकर 'के' अथवा 'के द्वारा' परसर्ग का प्रयोग किया जाता है।
 - कर्म के साथ कोई परसर्ग हो तो उसे हटा दिया जाता है।
 - कर्तृवाच्य की मुख्य क्रिया को सामान्य भूतकाल में परिवर्तित किया जाता है।
 - परिवर्तित क्रिया के साथ 'जाना' क्रिया का काल, पुरुष, वचन और लिंग के अनुसार जो रूप हो, उसे जोड़कर साधारण क्रिया को संयुक्त क्रिया में बदला जाता है।

Implementation Pipeline

- Find sentences which can be converted to passive voice
 - Non-copula verbs with distinct k1 and k2
 - Ex - “राम अच्छा है” can NOT be converted to passive voice
- Removed post-positions from K1 and K2
- Added “द्वारा” to the last noun/pronoun of k1 (RP aren't affected)
- If pronoun in k1, replace by appropriate form
- In the verb, add appropriate form of “जा” just before the auxiliaries
 - The inflection of “जा” is based on the inflection of the verb
- Verb is converted to samanya bhootkaal form
- K2's head's gender applied to the verb



THANK YOU