ISSN 2250-3153

Morphology: Indian Languages and European Languages

Shweta Vikram

Computer Science, Banasthali University

Abstract- Natural Language Processing (NLP) is a very popular and research area of computer science. NLP is a part of Artificial Intelligent but NLP has combination of many fields such as Hindi, English, and Computer Science etc. This paper contains how verb work in Hindi and English languages and morphology of both languages. Morphological Analyzer and generator is a tool for analyzing the given word and generator for generating word given the stem and its features. There are many Indian languages and many European languages but generally Hindi language consider as an Indian language and English as a European language. Both languages have grammatical rules. In English language, we do not use verbs as gender identification but in Hindi we use verbs for gender identification.

Index Terms- European languages, Indian languages, morphology, and verb

I. INTRODUCTION

Natural Language Processing (NLP) is the basic interface between Human and Computer. It is although considered as a branch of Artificial Intelligence; it is the need for almost all IT applications at the front end at least. It is also considered as major localization problem. No software application can proliferate to all users unless it has utility to operate it with local language.

Morphology is the field of the linguistics that studies the internal structure of the words. Morphological Analysis is essential for Hindi it has a rich system of inflectional morphology as like other Indo-Aryan family languages. Main concern here is on the grammatical information of words and this grammatical information like gender, number, person etc. is marked through the inflectional suffixes. A variety of scripts, and hundreds of dialects Eighth Schedule, lists twenty two Scheduled Languages - Assamese, Bengali, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Malayalam, Manipuri, Marathi, Nepali, Oriya, Punjabi, Sanskrit, Sindhi, Tamil, Telugu, Urdu, Bodo, Dogri, Maithili and Santhali Hindi is spoken by 43% population of India followed by Bengali, Telugu, Marathi and others.

II. HISTORY

Artificial Intelligence (AI) goal initially was to give computer the ability to parse natural language sentences similar to sentence diagrams that grade-school children learn. One of the first such systems was developed in 1963 by Susumu Kuno of Harvard. The system revealed the depth of ambiguity in English language. Natural languages have very little inflectional morphology to distinguish between parts of speech; additional information is implied by emphasizing them by speech. The goal of NLP evaluation is to measure one or more qualities of an algorithm or a system, in order to determine whether and to what extent the system answers the goals of its designers, or meets the needs of its users. Research in NLP evaluation has received considerable attention.

III. NATURAL LANGUAGE PROCESSING

Natural language processing is a field of computer science and Computational linguistics and is associated with human – Machine interaction having two major components:

1) Natural language generation systems to convert information from computer to natural language and
2) Natural language understanding systems to convert reverse way.

As such NLP is the basic interface between Human and Computer. Many NLP problems fall under both generation and understanding NLP has significant overlap with the field of computational linguistics, and is often considered a sub-field of artificial intelligence.

Stages of language processing:

- Phonetics and phonology
- Morphology
- Lexical Analysis
- Syntactic Analysis
- Semantic Analysis
- Pragmatics
- Discourse

In this paper we discuss morphology in detail.

Word formation rules from root words

- Nouns: Plural (boy-boys); Gender marking (czar-czarina)
- Verbs: Tense (stretch-stretched); Aspect (e.g. perfective sit-had sat); Modality (e.g. request khaanaa→ khaaiie)
- First crucial first step in NLP
- Languages rich in morphology: e.g., Dravidian, Hungarian, Turkish
- Languages poor in morphology: Chinese, English
- Languages with rich morphology have the advantage of easier processing at higher stages of processing

A task of interest to computer science: Finite State Machines for Word Morphology.

IV. MORPHOLOGY

Morphology is the study of the way words are built up from smaller meaning bearing units, morphemes. European languages have both regular noun and irregular noun but Hindi language have only regular noun. English verbal inflection is complicated then nominal inflection.

English has three kinds of verbs:

- Main verbs (eat, sleep, impeach)
- Model verbs (can, will, should)
- Primary verb (be, have, do)

Morphology is often useful to distinguish two broad classes of morphemes: stems and affixes. The exact details of distinction vary from language to language but intuitively, the steam is the main morpheme of the word, supplying the main meaning, while the affixes add additional meanings of various kinds.

Affixes are further divided into prefixes, suffixes and circumfuse. Prefixes precede the stem, suffixes follow the stem, circumfuse do both, and infixes are inserted inside the stem.

	Regular Noun	Irregular noun	
singular	Dog	Mouse	English language
plural	Dogs	Mice	
singular	dqRrk	pwgk	Hindi language
plural	dqRras	pwgsa	

Table 1: Morphology

Morphemes have two broad classes:

1) Inflection is the combination of a word stem with a grammatical morpheme, usually resulting in a word of the same class as the original stem.

2) Derivation is the combination of word stem with a grammatical morpheme, usually resulting in a word of a different class.

Inflectional morphology is work for only noun, verbs and adjective.

Derivation in English is more complex than inflection.

Suffix	Base verb	Derived word
ation	Computerize	Computerization
able	Embraceable	Embraceable
er	Kill	Killer
ee	Appoint	appointee

Table 2: Derivational Inflection

In English language, we do not use verbs as gender identification but in Hindi we use verbs for gender identification.

For example:

Saanchi NLP padati hai.
(Sanchi reads NLP.)

Saachya NLP padtaa hai.
(Sachya reads NLP.)

Here we easily see the difference between genders. For female (Sanchi) padati hai and for male (Sachya) padata hai. But in English both gender have same verb read. In these examples we can easily understand one more thing that is in English we use verb in between of the sentences or after that noun or pronoun. But in Hindi we use verb at last of he sentences.

If we had an exhaustive lexicon which listed all the word forms of all the roots and along with each word from it listed its features values then clearly we do not need a morphological analyser.

Lexicon has many problems:

- It is extremely wasteful of memory space
- It does not show relationships among different roots that have similar word forms.
- Some languages have a rich and productive morphology.

The figure morphological analyzer input output shows input output in paninian grammar.

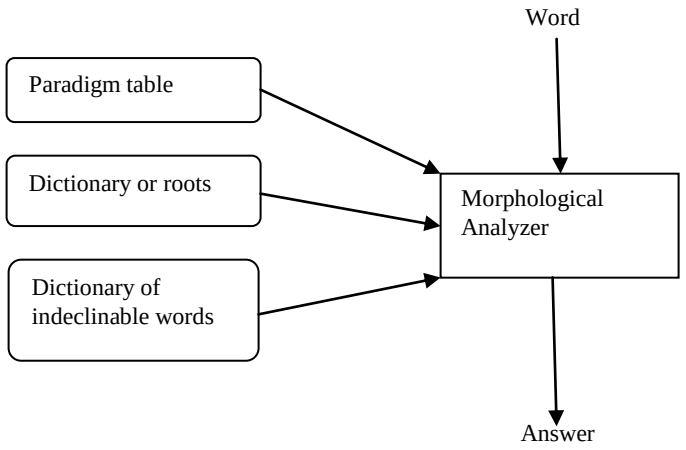

Figure.1: Morphological analyser input-output

A. Hindi Morphology

Morphology involves the study of inner structure of words and their forms in different uses and constructions. It can be mainly divided into two branches – derivational morphology and inflectional morphology. Derivational morphology involves the processes by which new lexemes are built from existing ones mainly through the addition of affixes. As an example in Hindi + e + esjk = eesjk (Pronoun to Adjective), like in English – go + at = goat (verb to noun) etc. Inflectional morphology involves the processes by which various inflectional forms are formed from a lexical stem. As an example in Hindi – inflectional forms of noun (guest) are vfrfFk (masculine-singular-direct), vfrfFk (masculine oblique-singular), vfrfFk (masculine-direct-plural), vfrfFk;ksa (masculine-oblique-plural). Hindi is very rich in inflectional morphology can be witnessed from the fact that in English usually there are maximum of 7-8 inflected word forms of noun but in Hindi it can be up to 40 and even more than that.

Grammatical information that are being used for Hindi:

Noun: Grammatical information required for Hindi Nouns are - gender, number and case. Gender can be masculine, feminine or both (as some nouns can be used both as masculine and as feminine). Number can be singular or plural. Case can be two types (in present work) – direct and oblique.

Pronoun: Grammatical information required is – number, case, person, and gender. Gender can be masculine, feminine or both. Number can be singular or plural. Case can be two types (in present work) – irect and oblique. Person can take first, second and third person.

Adjective: Grammatical information required for Hindi nouns is - gender, number and case. Gender can be masculine, feminine. Number can be singular or plural. Case can be two types – direct and oblique.

Verb: Grammatical information required is gender, number, person, Gender can be masculine, feminine. Number can be singular or plural. Person can take first, second and third person. Number can be singular and plural.

Adverb: There are two classes of adverbs, inflected and uninflected. Inflected adverbs behave like nouns so no separate paradigms were required for these. Grammatical information required for inflected adverbs will be same as required for nouns and for uninflected adverbs no grammatical information is to be stored.

Sharisthi Pronoun: Grammatical information required is – number, case, person, and gender, parsarg. Gender can be masculine, feminine or both. Number can be singular or plural. Case can be two types (in present work) – direct and oblique. Person can take first, second and third person. Parsarg will be shashthi.

V. NDIAN LANGUAGE PROPERTY

A scripts used are phonetic in Nature Better Articulatory discipline Systematic manner of production Five or Six distinct places of Articulation Various types of Flaps/Taps or Trills Fewer fricatives compared to English / European languages Presence of retroflex consonants A significant amount of vocabulary in Sanskrit with Dravidian or Austroasiatic origin gives indications of mutual borrowing and counter influences.

A. Some Exceptions

- In Tamil language, all plosives of a given place of articulation are represented by a single grapheme. The pronunciation of such graphemes depends on the context.
- More fricative consonants are present in Hindi, Punjabi, Sindhi, Kashmiri and Urdu due to influence of Perso-Arabic & English.
- च and ज are dental-alveolar in Marathi only, while these are alveolar in Hindi.
- ड़ and ढ़ are present in Hindi, Urdu, Sindhi, Punjabi & Oriya.
- Fricative स is श or ष in Oriya.
- व and ब are pronounced as ब in Bengali.
- व and ब are pronounced as भ in Oriya.
- Punjabi is a tonal language mostly in aspirated voiced sounds.
- Sindhi language has implosives.
- Native words of the Dravidian languages do not contain aspirated sounds.
- ण sound is more frequently used in Gujarati and Marathi.
- त and द are pronounced mostly as ट and ड in Assamese and Gujarati languages.
- य and स are pronounced as ज and ह in Assamese.

VI. DIFFERENCES

There are many differences between Hindi and English languages.

A Difference in Origin

The Hindi language could be traced back to the Khariboli dialect used in Delhi. It belongs to the Indo-European language family under the Western Hindi. English meanwhile is of the Germanic language family. Its roots could be traced to as far back as the year 400's from the tongues of the Anglo-Saxon kingdoms in England. Though both Hindi and English are under the Indo-European language family, they have major differences.

B Difference in Alphabet

The English language is written using the English alphabet consisting of 26 letters. Written Hindi employs the Devangari script which contains 10 vowels and 40 consonants. Bars on top of the symbols make them distinct. Since Hindi is a phonetic language, the word is pronounced according to its spelling. This makes the language easier to pronounce since it follows the written form all the time. This is unlike English where pronunciations of certain words do not strictly follow the written form.

C Difference in Grammar

English and Hindi share the same verb tenses, i.e. simple present, past, future, and so on. In Hindi, however, there is a deficiency in associating them to properly convey different meanings. Thus, Hindi speakers often find themselves using present continuous instead of the simple present.

Hindi does not have the equivalent of "do". So instead of structuring the words to form queries, intonation is used to convey a question. Moreover, in conditional sentences, Hindi makes use of the future tense in the independent clause.

In English, polite requests are usually expressed in the form of questions. In Hindi, subjunctives are employed and the sentence structure changed to ask for something. While the English language has definite articles, Hindi does not. Moreover, the number "one" is used instead of the indefinite article.

In terms of sentence structure, the subject-object-verb is used in Hindi, whereas English uses the subject-verb-object word order.

Prepositions in English come after the pronoun or noun they qualify. For Hindi, prepositions succeed the noun or pronouns. Hindi speakers typically have difficulty using the correct prepositions as is common to most non-native speakers learning English.

D Difference in Vocabulary

Hindi has adapted many English words. The pronunciations of these incorporated words, though, have been modified.

Though differences between Hindi and English languages are clear, especially with the Devangari script it uses, Hindi is not too hard to master.

Figure.2: European language country

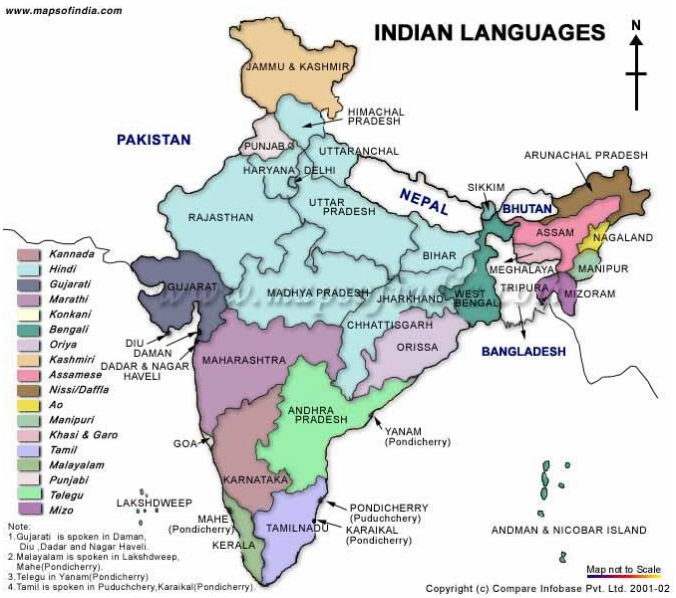

Figure.3: Indian language country

VII. ENGLISH LANGUAGE

Words are divided into different kinds or classes, called Parts of speech, according to their use; that is, according to they do in a sentence. The parts of speech are eight in number:-

- Noun: A noun is a word used as the name of a person place or thing.
- Pronoun: A pronoun is a word used instead of a noun.
- Verb: A verb is a word used to express an action or state.
- Adjective: An adjective as a word used to add something to the meaning of a noun.
- Adverb: An adverb is a word used to add something to the meaning of a verb, an adjective, or another verb.
- Preposition: A preposition is a word used with a noun or a pronoun to show how the person or thing denoted by the noun or pronoun stand in relation to something else.
- Conjunction: A conjunction is a word used to join word or sentences.
- Interjection: An interjection is a word which expresses some sudden feeling.

A verb is a word that tells or asserts something about a person or thing. Verb comes from the Latin verbum, a word. It is so called because it is the most important word in a sentence.

The verb, like the personal pronouns, has three person- the first, the second and third. Thus we say

1) I speak.
2) You speak.
3) He speaks.

This is because of the difference in person of the subjects, as all the three are subjects of the singular number.
In sentence 1, the subject is of the first person; therefore the verb is also of the first person.
In sentence 2, the subject is of the second person; therefore the verb is also of the second person.
In sentence 3, the subject is of the third person; therefore the verb is also of the third person.
We thus see that the verb takes the same person as its subject: or, that the verb agrees with its subject in person.

The verb like the noun and the pronoun has two numbers: the singular and the plural. Thus we say-

1) He speaks.
2) They speak.

This is because of the difference in number of the subjects.
In sentence 1, the subject is singular, therefore the verb is singular.

In sentence 2, the subject is plural, therefore the verb is plural.
We thus see that the verbs takes the same number as its subject; or that the verb agrees with its subject in number.

VIII. CONCLUSIONS

In this paper we see both languages: Indian and European. The growth of European languages is increase day by day and every most of the thing available in computer in your own languages. But in Indian language there are not same as European language in computer field. For example most o the thing available in Wikipedia in Turkish, but in Hindi reverses of this. So we are trying to convert all things in Hindi. This language problem most of the Indian people not connect with computer. If everything available in Hindi then every people connected with computer. With the help of Natural Language Processing not far that day when everything available in Hindi.

REFERENCES

[1] http://en.wikipedia.org/wiki/Morphology_(linguistics).

[2] http://clas.uiowa.edu/lingustics/hindi-verb-ptoject.

[3] http://www.academia.edu/25541/Hindi_Noun_Inflection_and_Distributed_Morphology.

[4] Speech and Language Processing- An introduction to natural Language Processing, computational Linguistics, and Speech Reorganization by Daniel Jurafsky and James H. Martin, University of Colorado, Boulder.

[5] First International Conference on Emerging Treads in Engineering and Technology, Hindi Morphological Analyzer and Generator By Vishal Goyal, Gurpreet Singh Lehal lectuter, Department of Computer Science, Punjabi University, Patiyala, Professor and head department of Computer Science, Punjabi University,Patiyala..

[6] Akshar Barti, Vineet Chaitanya and Rajeev Singhal, Natural Language Processing: A Paninian Perseptive, Prentice Hall of India,1995.

[7] Development of Resources and Techniques for Processing ofssome Indian Languages by Shyam S. Agrawal, Advisor,CDAC,Noida and Exicutive Director, KIIT,Gurgaon

AUTHOR

First Author – Shweta Vikram, M.tech (Computer science), Banasthali University Jaipur, India.
shwetavikram.2009@rediffmail.com