KNOWLEDGE SHARING EVENT

MARCH 22 – 26, 2010

Linguistic Data Consortium for Indian Languages

OBJECTIVE

- While working on the POS tagging of Bangla, I have come across a whole lot of issues that this paper seeks to record in its limited range and scope.

- For the time being, discourse is restricted to the number and case markers of the said language and the way they affect the morpho-syntactic behavior of nouns leading to a major confusion in building up an appropriate tag set.

- These so called limitations of features captured in a single level like POS often appear to present a conflict among and across grammatical models which leaves a human tagger both mesmerized and perplexed.

CONTD..

- The aim of this paper is to show a brief and partial sketch of issues that human annotators might face while working out the Part of Speech (POS) tagging of Bangla.

- It is seen that the human taggers face immense difficulties in tagging the text corpus due to the confusion created by the words or lexical items in certain specific contexts.

Areas covered

Noun

1. Number
2. Case markers

NUMBER

- While tagging Bangla nouns, it was suggested that Common Nouns which are not capable of carrying any numeral markers can be tagged as NC.0 ; whereas ,Common Nouns which can carry numbers or denote numbers at any other context can be tagged as NC.sg… or NC.pl…

Example:-

সে চলে গেল এক মুহূর্তের \NC.sg মধ্যে।

se cale gela eka muhrtera madhye

S\he go go-3per-smp-pst one moment-gen within

'S\he went away within a moment.'

CONTD..

se cale gela   eka muhrtera   madhye

- In this sentence the difficulty is whether to tag common nouns such as 'মুহুর্তের' [muhrtera] as zero or singular or plural.
- Here I would prefer to tag it as singular instead of zero, because the word 'মুহুর্ত' [muhrta] is capable of forming a plural word with the addition of the suffix 'গুলো' [gulo] along with it, for example -

Example-

সেই মুহূর্তগুলোর\NC.pl কথা মনে পড়লে,

sei muhrtagulora kathA mane paD'ale

That moment-plu-gen statement remember do

আমার শরীর ঠান্ডা হয়ে যায়।
AmAra sharIra ThAnDA haYe yAYa

I-gen body cold become go-1per-smp-prs

'Remembering those moments, my body becomes cold.'

Example-

মিনিট\NC.0   সাতেকের\NC.pl   ভেতরে   মারুতি
miniTA   sAtekera   bhetare   mAruti
Minute   seven-gen   within   Maruti

বৃদ্ধাবাসে\NC.sg   পৌঁছে   যায়।

briddhAbAse   paum'che   yAYa

oldagehome-loc   reach   go-3per-smp-prs

'Within seven minutes Maruti reaches old age home.'

- Here, the difficulty is whether to tag common nouns such as 'মিনিট' [miniTA] and 'সাতেকের' [sAtekera] as zero or singular or plural. In IL-POSTS(Indian Language POS Tag Set) it has been suggested to tag the word as zero.

- But I prefer it to be tagged as plural by looking into the context of its occurrence and the semantic meaning we derive from the words.

- Evidently, 'minute' in English and [miniTA] in Bangla, both carry a sense of plurality.

'মিনিট' [miniTA] + x (plural marker)

'মিনিট' [miniTA] (x=0)

So the very word 'minute' stands as a unit to define the quality of something that has the numerical value 'seven' (quantity). Thus it stands like this-

Unit (minute)

Time <

Value(7)

Example-

এই প্রতাপপুরেই   জন্ম\NC.0   সুপ্রভার।

Ei pratApapu janma   suprabhAra

This Pratapur-loc-emp birth   Suprabha-gen

'Suprabha was born in this Pratappur.'

- Here the difficulty is whether to tag common noun 'জন্ম' [janma] as zero or singular.

- Here I prefer it to be tagged as zero instead of singular, because this word does not exactly carry number sense along with it.

- It can be stated that birth, death and other biological occurrences that happen once in a lifetime are most unlikely to carry a plural marker if no intentional stylistic device is involved.

- Even when we refer to the birth of more than one individual ,then also the event of birth does not share a 'quality-quantity' relation like the previous one(as 'জন্ম' [janma] does not refer to a unit).

- So birth is a single occurrence experienced by more than one individual.

What is Case?

- The case and the case markers play a crucial role especially in relation to Noun and Pronoun.

- Case is considered to be an inflectional category, basically of Nouns and Pronouns, which specifically marks their function in relation to other parts of the sentence.

- 'Case' is used in the analysis of word classes (or their associated phrases) to identify the syntactic relationship between the words in a sentence.

CASE SYSTEM IN BANGLA

- In Bangla, the Nouns having Nominative case and Vocative case are considered to be in Direct case

- Nouns having other case markers, or in situations where followed by postpositions are regarded to be in Oblique case.

- A single set of case markers can denote various cases depending on the contexts they are posed in.

NOMINATIVE CASE

- The nominative case is to mark the subject in a sentence .

- In Bangla 'Nominative' case markers are 'এ' [-e], 'য়' [-ya] , and generally null(0).

Example-

- লোকে\NC.pl.nom   বলে।

    (Lok-e)   bale

    People-nom   say-3per-smp-prs

    'People   say.'

Example-

- বেড়ালে\NC.sg.nom   দুধ   খায়।

  (beD'Al-e)   dudha   khAYa

  Cat-nom   milk   drink-3per-smp- prs

    'Cat drinks milk.'

Example-

- বাবা\NC.sg.nom   এ কাজ   করেছেন।

   (bAbA-0)   e   kAja   karechena

    Father   this work   do-3per-prf-prs-hon

       'Father has done this work.'

ISSUE

- But the fact is that whenever we find 'এ' [-e] as nominative case marker, at that time still a confusion remains whether to tag it as nominative, instrumental or locative case marker.

ACCUSATIVE CASE

- A noun is said to be in the accusative case when it is the direct object of a verb. The accusative case markers is [–ke] , but 0 and [–e], [-eye] and [-ya] can also be used as accusative markers.

Example-

আকাশ গীতাকে\NP.sg.acc খাবার   দেবে।

AkAsha   (gItA-ke)   khAbAra   debe

Akash   Gita-acc   food   give-3per-
   smp-fut

'Akash will give food to Gita.'

DATIVE CASE

- The basic role of dative case is to distinguish the recipient of something given, transferred, etc.

- It usually expresses the indirect object relationship as well as a range of meaning, which is similar to the meaning expressed by English 'to' or 'for'.

- The dative markers for Bangla are '- কে'[-ke], '-রে' [-re],'-এরে'[-ere],'-এ'[-e],'-য়'[-ya].

Example-

কথাটা রামকে\NP.sg.dat বলো ।

kathATA (rAma-ke) balo

Statement-cla Ram-dat say-2per-smp-prs

'Tell the statement to Ram'

GENITIVE CASE

- Genitive case indicates the possessive relationship of nouns. It basically reflects a possessor – possessed relationship.
- The genitive case markers in Bangla are [–r] and [–er].

Example-

রামের\NC.gen   বই।

(rAm-er)   bai

Ram-gen   book

'Ram's book.'

Example-

তার হাতের\NC.gen চুড়িও নকল ছিল।

tAra hAt-era cuD'io nakala chila

S\he-gen hand-gen bangle also false be-3per-smp-pst

'The bangle (worn in her hand) was also false.'

LOCATIVE CASE

- The locative case generally marks the location with both the spatial and temporal references.

- This case indicates the location of the event/action identified by the verb.

- Apart from this it also used to serve to locate something. The locative case markers are [-ya],[-te]; etc., extended also to the instrumental and the nominative case.

Example-

রাত দশটায়\NC.sg.loc শেষ হবে।

rAta (dashaTA-Ya) sheSa habe

Night ten-def-loc finish be-3per-smp-fut

'It will finish at ten at night'

Example-

আমি   বাড়িতে\NC.sg.loc   যাবো।

 Ami   (bAD'i-te)   yAbo

I   home-loc   go-1per-smp-fut

    'I will go home.'

Example-

ঘড়ি   খুলে   ধীরে-সুস্থেই   ছেলেটার   হাতে\NC.loc

ghaD'I   khule dhIre-susthei   cheleTAra   hAt-e

Watch   open   slowly-patiently   boy-def-gen   hand-loc

দিলেন।

 dilena

give-3per-smp-pst-hon

'(S\he) slowly and patiently removed the watch and gave it in the hands of the boy.'

31

INSTRUMENTAL CASE

- The Bangla instrumental case indicates the instrument that is used in carrying out the action identified by the verb.

- It is mostly marked by the postpositions 'দিয়ে' [diya], 'দ্বারা' [dara], etc, which expresses the meaning "by means of".

Example-

হাতুড়ি   দিয়ে\PP.ins   মারো ।

hAtuD'i   diYe   mAro

Hammer   with   hit-2per-smp-prs

'(You) hit with hammer.'

Here the postposition is indicating the instrumental case and the noun হাতুড়ি [hAtuD'i] contains null case marker for the instrumental case.

Example-

- এটা নিজহানীর নিজ হাতে\NC.ins লেখা লম্বা চিঠি।

eTa nijahAnIra nija hAt-e   lekhA   lambA ciThi

This nijahani's   self   hand-gen   write   long   letter

'This letter is written by Nijohanir's own hand.'

Example-

- কুটুমবাড়িতে হাতে\NC.ins করে নিয়ে যাবেন।

kuTumabAD'ite hAt-e kare niYe yAbena

In-laws-home-loc hand-ins with take go-3per-smp-fut-hon

'S\he will take that thing to her\his in-law house with her\his own hand.'

ISSUE

- The locative and instrumental cases were confused through the possession of the common affix 'এ' [-e] and in some cases the form could be equally either locative or instrumental.

- This resulted in the employment of the postpositional locative affix [-ta] in the instrumental as well. In New Bengali [–ta], [-te], [-ete] became established as the instrumental case marker.

- For the instrumental, and for words ending in vowels –te has become the characteristic of instrumental affix.

ABLATIVE CASE

- The ablative case indicates movement away from some location.

- It specifies the point in space to where something is transferred at the culmination of the action, identified by the verb.

- Bangla does not possess an organic affix for ablative. The use of locative form in [–ta], [-te], for ablative exists.

CONTD..

- There is a confusion or interchangeability among the four cases, instrumental, ablative, locative and genitive case markers.

- Since ablative does not have a specific case marker so instrumental, genitive and locative case markers are sometimes employed as ablative case markers.

Example-

চোখ দিয়ে\PP.abl আগুন বের হচ্ছে ।

cokha   diYe   Aguna   bera   hacche

Eye   from   fire   (come out) happen-3per-smp-prs

'Fire is coming out from eye.'

The postposition দিয়ে [diYe] mostly denotes instrumental

case, but here in example, it refers to the semantic sense

of ablative case.

বাজারের\NC.sg.abl   কেনা   দই

  (bAjAra-er)   kenA   dai

Market   buy   curd

   '(This) curd is bought from market.'

Here the genitive case marker is employed as an ablative case marker. But in this case the actual sense is that shown in following example

বাজার   থেকে\PP.ins   কেনা   দই।

bAjAra   theke   kenA   dai

Market   from   buy   curd

   '(This) curd is bought from market'

Some grammarians stated that from locative case marker,

[- e], we get ablative sense. Example as follows-

দুধে \NC.abl   হয়   দই।

(dudh-e)   haYa   dai

 Milk-abl   happen-3per- smp-prs   curd

          'Curd is made from milk.'

Here the embedded sense is that shown in example, that is as
     follows-

দুধ   থেকে\PP.abl   হয়   দই।

dudha   theke   haYa   dai

 Milk   from   happen-3per-smp-prs   curd

          'Curd is made from milk.'

There can also be constructions where there is no visible ablative sense present. For example-

তুমি রোজ স্কুল পালাও |

tumi roja skula pAlAo

You everyday school skip -2per-smp-prs

'Every day you skip from school.'

Carrying the sense as shown in example as follows-

তুমি রোজ স্কুল থেকে\PP.abl পালাও |

tumi roja skula theke pAlAo

You everyday school from skip-2per-smp-prs

'Everyday you skip from school.'

CONCLUSION

- Thus to identify a grammatical conception like case becomes a challenge for the computer.

CONTD..

- More so, when it comes to multi-layered applications of general linguistic categories, even if there are no number markers still the words can assume zero, singular or plural number depending on the contexts of its occurrence.

CONTD..

- It is also observed that there are certain abstract nouns, which assume markers so they should be tagged accordingly, although they are generally regarded as uncountable.

CONTD..

- The next concern of the paper was the case system of Bangla. As far as the case markers are concerned, we have to contemplate on both c-selection and s-selection of Bangla sentences.

- The consideration of the scheme reveals different rules which these markers claim in relation to different constituent of the sentence.

CONTD..

- This paper seeked to show that a marker cannot define as marking a particular case without reference to its context.

- The syntactic categories and semantic dimensions together assign a case to these markers.

CONTD..

- In brief, Bangla markers have multi-faceted exposure to different case relations and once used in the sentence, they assume particular inflectional features depending on the interaction among the concerned constituents

THANK YOU

☺

Questions and Suggestions