A constraint based dependancy parser for Sanskrit

Amba Kulkarni

apksh@uohyd.ernet.in

Department of Sanskrit Studies

University of Hyderabad

Hyderabad

शाब्दबोध-प्रक्रियायाम् अन्वयबोधः एकः महत्त्वपूर्णः सोपानः वर्तते।

अन्वयः नाम किम्?

'एकस्य पदस्य अपर-पदेन सह कः सम्बन्धः' इति अन्वयः।

उदाहरणार्थम्,
रामः ग्रामं गच्छति।
अस्मिन् वाक्ये

'रामः' इति गमन-क्रियायाः कर्ता,
'ग्रामम्' इति गमन-क्रियायाः कर्म,
एतादृशः सम्बन्धः अन्वयः।

इदं सम्बन्धज्ञानं कथं जायते?

WHERE is the information about the relations coded?

पदस्य पदेन सह सम्बन्धः त्रिभिः प्रकारैः अभिधीयते।

- अभिहितत्वेन,
- विभक्त्या (उपपदेन, निपातेन च), एवं
- सामानाधिकरण्येन

वाक्य-संश्लेषण-समये विभक्ति-निर्धारणं सरलं भवति।

Analysis Process is non-deterministic
परन्तु वाक्य-विश्लेषण-समये नैक-विश्लेषणानि सम्भाव्यमानानि।

तस्य कारणं किम्?

तत्र कारणद्वयम्।

- विभक्ति-अर्थ-अनेकत्वम्,
- प्रत्येकस्य पदस्य नैक-विश्लेषणानि

a) विभक्ति-अर्थ-अनेकत्वम्

कर्तृकरणयोस्तृतीया (अनभिहिते) (पा.सू. 2.3.18)

अनेन सूत्रेण कर्मणि प्रयोगे कर्तरि एवं करणे द्वयोः अपि तृतीया विभक्तिः दरीदृश्यते।

तस्मात् विश्लेषण-समये, कर्मणि प्रयोगे, तृतीया विभक्तिम् आधारीकृत्य एतद् पदं कर्तृपदम् इति वक्तुं न शक्यते।

कर्मणि प्रयोगे तृतीयान्तं पदं कर्ता वा करणं वा भवितुम् अर्हति।

उदाहरणार्थम्,

रामेण बाणेन वालिः हन्यते।

अस्मिन् वाक्ये केवलं विभक्तिं दृष्ट्वा रामः कर्ता एवं बाणं करणं इति वक्तुं न शक्यते।

तत्र तादृश-ज्ञानबोधनाय योग्यता ज्ञानम् अपि आवश्यकम्।

b) पदस्य नैक-विश्लेषणानि

एकं वाक्यम् पश्यामः।

रामः वनं गच्छति।
अस्मिन् वाक्ये पद-विश्लेषणम् अनेन प्रकारेण भवितुम् अर्हति।

रामः = राम पुं 1 ए.व. / रा दाने अदादि लट् उ.पु. ब.व. परस्मैपदी
वनं = वन नपुं 1 ए.व. / वन नपुं 2 ए.व.
गच्छति = गच्छत् (गम्लृ शतृ) पुं 7 ए.व. / गच्छत् (गम्लृ शतृ) नपुं 7 ए.व. / गम्लृ लट् प्र.पु ए.व. परस्मैपदी

अत्र द्वौ अन्वयौ प्रतिपादयितुं शक्यते।

- रामः इति गमन-क्रियायाः कर्ता,
  वनं इति गमन-क्रियायाः कर्म।
- वयं इति रा-धातु-निर्दिष्ट-क्रियायाः कर्ता (अध्याहृत),
  वनं इति रा-धातु-निर्दिष्ट-क्रियायाः कर्म, एवं,
  गच्छति इत्यत्र सति सप्तमी

वस्तुतः द्वितीयान्वये पुनः तत्र द्वितीय-क्रियायाः आकाङ्क्षा वर्तते।
अर्थात् अस्मिन् उदाहरणे आकाङ्क्षापूर्त्यभावात् द्वितीयान्वयः त्यक्तुं शक्यते किन्तु साधारणतया त्यागनिर्णयः न सरलः।

Prioritizing Multiple Analysis

यदा नैक-अन्वयानि वर्तन्ते तदा किं तेषु अन्वयेषु पौर्वापर्यसम्बन्धः स्थापितुं शक्यते वा?

अन्वयबोधार्थं आकाङ्क्षा-योग्यता-तात्पर्य-सन्निधि-ज्ञानम् आवश्यकम्।

एषु चतुर्षु आकाङ्क्षा एवं सन्निधिज्ञानं सङ्गणकं प्रयोक्तुं शक्यते।

यदा योग्यताज्ञानार्थं पदार्थस्य ज्ञानम् आवश्यकम्
तदैव तात्पर्यज्ञानार्थं व्याप्ति-अनुमानम्-संस्कृत्यादि
नैकविध-ज्ञानस्य आवश्यकता भवति।

तादृशं ज्ञानम् अस्मिन् समये तु सङ्गणके नास्ति।

यद्यपि योग्यता-तात्पर्य ज्ञानाभावात् कानिचन
सम्भाव्य-विश्लेषणानि त्यक्तुं न शक्यते अपि तु
आकाङ्क्षा-सन्निधिम् अधिकृत्य तेषाम् पौर्वापर्यं तु दर्शयितुं
शक्यते।

एकम् उदाहरणं पश्यामः।

शकटं यानं गच्छति।

अस्मिन् वाक्ये
एवंविधम् अस्ति पद-विश्लेषणं-

शकटम् = शकट 1 नपुं ए.व. / शकट 2 नपुं ए.व.
वनम् = वन 1 नपुं ए.व. / वन 2 नपुं ए.व.
गच्छति = गम्ऌ लट् कर्तरि प्र.पु. ए.व. / गच्छत्(गम्ऌ शतृ)
नपुं 7 ए.व. / गच्छत् (गम्ऌ शतृ) पुं 7 ए.व.

अत्र षड्विध-विश्लेषणानि सम्भवन्ति

1. शकटम् इति गमन-क्रियायाः कर्ता,
   वनम् इति गमन-क्रियायाः कर्म।
2. शकटम् इति गमन-क्रियायाः कर्म,
   वनम् इति गमन-क्रियायाः कर्ता।
3. वनम् इति गमन-क्रियायाः कर्ता,
   शकटम् इति वनस्य विशेषणम्।
4. वनम् इति गमन-क्रियायाः कर्म,
   शकटम् इति वनस्य विशेषणम्।

5. शकटम् इति गमन-क्रियायाः कर्ता,
   वनम् इति शकटस्य विशेषणम्।

6. शकटम् इति गमन-क्रियायाः कर्म,
   वनम् इति शकटस्य विशेषणम्।

अत्र आद्यौ द्वौ अन्वयौ कर्ता एवं कर्म द्वयोः अपि कारक-सम्बन्धौ दृश्यतः।

तृतीयातः आरभ्य षड्पर्यन्तं केवलं एकं एव कारक-सम्बन्धः अस्ति। अपर-कारकस्य अध्याहारं करणीयम्।

तत्र अपि पुनः पञ्चमे एवं षट्-अन्वये वनं विशेषण-रूपेण भासते।

संस्कृत-भाषायां यद्यपि शब्दस्य क्रमस्य न कोऽपि संकेतः वर्तते तथापि विशेषणानन्तरं विशेष्यम् इति क्रमः सामान्यरूपेण अनुभूयते।

तथैव सामान्यतः प्रथमं कर्तृ-कारकं तदनन्तरं कर्म कारकम् इत्यपि क्रमः प्रायः गद्यपाठेषु दृश्यते।

तथा आकाङ्क्षा एवं सामान्य-क्रमम् अनुसृत्य अन्वयस्य पौर्वापर्यस्य निर्णयं कर्तुं शक्यते।

HOW is the information coded (explicitly or implicitly)?

कीदृशः सम्बन्धः अन्वये दर्शनीयः।

यः सम्बन्धः प्रत्ययेन साक्षात् उच्यते स एव सम्बन्धः स्वीकरणीयः। न तु परम्परया।

उदाहरणार्थम्,

रामः दुग्धं पीत्वा शालां गच्छति।
अस्मिन् वाक्ये,

रामः इति पिब्-धातु-निर्दिष्ट-क्रियायाः कर्ता एतद् ज्ञानम् 'समानकर्तृकयोः पूर्वकाले' इति सूत्राधारेण भवति, न केनापि प्रत्ययेन।
अत एव अस्मिन् सन्दर्भे रामः इति पदस्य पीत्वा अनेन पदेन सह सम्बन्धः प्रथम सोपाने न दृष्टव्यः।

Another criterion for prioritizing Multiple Analysis

अस्मिन् उदाहरणे अन्या एका समस्या वर्तते।

दुग्धम् एवं शालां द्वौ अपि द्वितीया-विभक्तौ स्तः।
तदा द्वौ अपि पिब्-धातु-निर्दिष्ट-क्रियायाः एवं गमन-क्रियायाः कर्म भवितुम् अर्हति।
तस्मात् द्वौ सम्भावित-उत्तरे सम्भवतः

- दुग्धं – पिब्-धातु-निर्दिष्ट-क्रियायाः कर्म
  शालां – गमन-क्रियायाः कर्म
- दुग्धं – गमन-क्रियायाः कर्म
  शालां – पिब्-धातु-निर्दिष्ट-क्रियायाः कर्म

अत्र अपि सङ्गणके योग्यता ज्ञानस्य अभावात् द्वितीय-अन्वयं त्यक्तुं न शक्यते।

सन्निधि-आश्रयेण अन्वये पौर्वापर्यं प्राप्तुं शक्यते।

Taking into account all the above consideration, now, we propose a mathematical model to get the अन्वय of a sanskrit sentence mechanically.

Let each word in a sentence be represented as a node in a graph. The relations between nodes are represented by the directional labelled edges.

Given a graph G with n nodes representing n words of a sentence, the task is to find a subgraph of G which is a directed tree.[a]

[a] A Tree is a graph in which any two vertices are connected by exactly one simple path.

Each node consists of one or more cells, each cell representing one morphological analysis of the word represented by the node.
We break the problem into three parts:

- For a given sentence, draw all possible labelled directed edges between cells of one node to that of the other.
- Identify a subgraph T of G such that T is a directed Tree.
- Prioritize the solutions, in case there are more than one possible directed Trees.

Task 1:

Let a 5-tuple (i,j,R,l,m) represent an edge labelled R from the mth cell of the lth node to the jth cell of the ith node.

In other words this 5-tuple represents a relation R from the mth morphological analysis of the lth word to the jth morphological analysis of the ith word.

Let C[i,j,R,l,m] = 1, if such a relation exists, else = 0.

Based on the available information in a given sentence in the form of अभिहितत्व, विभक्ति, and सामानाधिकरण्य, the cost matrix is populated with 0s and 1s.

Task 2:

Following constraints are imposed on the matrix C

- A word can't satisfy its own expectancy. i.e. a word can't be linked to itself. Or in other words there can not be self loops.
  $\sum_{j,R,k} C[i,j,R,i,k] = 0$

- A word can fulfill only one expectancy. i.e. a node can have at the most one incoming arrow. This also ensures that a word can have only one valid analysis per solution.
  $\sum_{j,R,k,l} C[i,j,R,k,l] = 1$

- If a word has both an incoming arrow as well as an outgoing arrow, they should be through the same cell.
  $\forall i \forall j$
  $\sum_{R,l,n} C[i,j,R,l,n] + \sum_{a,b,R,k!=j} C[a,b,R,i,k] \leq 1$

- If there are more than one outgoing arrows through a node, then they should be through the same cell. i.e. there can be only one valid morph analysis per word, per analysis.

- There can not be more than one outgoing arrows with same label from the same cell, if the relation is a kaaraka relation. i.e. there can not be two words satisfying the same kaaraka role for the same verb.

- If all the nodes are plotted in a straight line, then they should not intersect each other. i.e.,
  if C[i,j,R,k,l]=1, then
  $\forall v \forall y C[u, v, w, x, y] = 0$,
  if i $<$ x $<$ k and u $<$ i or u $>$ k

Task 3:

The solutions are prioritised using the conditions specified below.

- For each of the analysis, the distance is measured as $d = \sum_{i,R,j} d_{iRj}$,
  where
  - $d_{iRj} = |j - i| * wt_R$,
    if C[i,a,R,j,b] = 1 for some a and b.
    $d_{iRj} = 0$ otherwise.
  - $wt_R = 100$, if R marks a relation with an upapada,
    = 10, if R marks a relation with a verb,
    = 1, otherwise.

This takes care of

- आकाङ्क्षा (कारक relation is preferred over the सामानाधिकरण्यम्)
- सन्निधिः

DEMO

Some real life problems:

गुहेन लक्ष्मणेन सीतया च सहितः रामः वनेन वनं गत्वा बहूदकाः नदीः तीर्त्वा भरद्वाजस्य शासनात् चित्रकूटम् अनुप्राप्य वने रम्यम् आवसथं कृत्वा देवगन्धर्वसङ्काशाः ते त्रयः रममाणाः सुखं न्यवसन्

Some real life problems:

गुहेन लक्ष्मणेन सीतया च सहितः रामः वनेन वनं गत्वा बहूदकाः नदीः तीर्त्वा भरद्वाजस्य शासनात् चित्रकूटम् अनुप्राप्य वने रम्यम् आवसथं कृत्वा देवगन्धर्वसङ्काशाः ते त्रयः रममाणाः सुखं न्यवसन्

के ते त्रयः?

Some real life problems:

गुहेन लक्ष्मणेन सीतया च सहितः रामः वनेन वनं गत्वा बहूदकाः नदीः तीर्त्वा भरद्वाजस्य शासनात् चित्रकूटम् अनुप्राप्य वने रम्यम् आवसथं कृत्वा देवगन्धर्वसङ्काशाः ते त्रयः रममाणाः सुखं न्यवसन्

रामः वनेन वनं गत्वा ... ते त्रयः न्यवसन्
Is it not violation of समानकर्तृकयोः पूर्वकाले?

Some real life problems:

गुहेन लक्ष्मणेन सीतया च सहितः रामः वनेन वनं गत्वा बहूदकाः नदीः तीर्त्वा भरद्वाजस्य शासनात् चित्रकूटम् अनुप्राप्य वने रम्यम् आवसथं कृत्वा देवगन्धर्वसङ्काशाः ते त्रयः रममाणाः सुखं न्यवसन्

Should the anvaya of गत्वा be with तीर्त्वा or न्यवसन्?

Some real life problems:

गुहेन लक्ष्मणेन सीतया च सहितः रामः वनेन वनं गत्वा बहूदकाः नदीः तीर्त्वा भरद्वाजस्य शासनात् चित्रकूटम् अनुप्राप्य वने रम्यम् आवसथं कृत्वा देवगन्धर्वसङ्काशाः ते त्रयः रममाणाः सुखं न्यवसन्

वनेन वनं अत्र कारक सम्बन्धः कः?

धन्यवादः