

Guidelines for tagging of Sanskrit Compounds

K V Ramkrishnamacharyulu, Amba Kulkarni, Tirumala Kulkarni and Anil Kumar

Final draft for circulation among the SHMT consortium members
dated 12/03/2012*

1 Background

With the advent of computers, and the advances in the field of NLP, the annotated corpora is gaining importance. Annotated corpora not only serves as an important resource for building Statistical tools for automatic annotation, but also provides useful insights for language teachers, language learners, and researchers working on various aspects of language.

Compound formation is very productive in Sanskrit. On an average every fourth word in a Sanskrit text is a compound. It is not practical to store all the compounds and their analysis/meaning. So for Sanskrit, one needs a very good compound identifier and a program to generate the paraphrase of the compound. Pāṇini has provided rules for compound formation and also semantic restrictions for many of the compound formations. But to implement these rules on machine, one requires a knowledge base. A good collection of tagged compounds will be useful in deciding the parameters for development of such a knowledge base. Similarly, a collection of tagged corpus will be useful for getting a frequency distribution of various compounds. Such a tagged corpus will also be useful for developing automatic compound identifiers using suitable machine learning algorithms.

2 Tag set for compounds

The Indian Grammatical Tradition has a vast literature on samāsa. Following the literature, Sanskrit compounds are classified into 5 major and 55

*The original draft has been modified after a series of workshops on samāsa tagging for Sanskrit. The details of the workshops are available on the SHMT portal viz: <http://sanskrit.uohyd.ernet.in/shmt/login.php>

minor categories. The major categories are

- अव्ययीभावः
- तत्पुरुषः
- कर्मधारयः
- बहुव्रीहिः
- द्वन्द्वः

Note here that there is a small deviation from the standard literature. कर्मधारयः has been added as a major type of compound, instead of sub-class of तत्पुरुषः. The reason is purely from the convenience of the tag-names. We give below the sub-classification along with the associated tags.

compound type:	अव्ययीभावः	compound type:	कर्मधारयः
अव्यय-पूर्वपदः	A1	विशेषण-पूर्वपद-कर्मधारयः	K1
अव्यय-उत्तरपदः	A2	विशेषण-उत्तरपद-कर्मधारयः	K2
तिष्ठद्भ्रुति	A3	विशेषण- उभयपद-कर्मधारयः	K3
संख्यापूर्वपद-नद्युत्तरपदः	A4	उपमान-पूर्वपद-कर्मधारयः	K4
नद्युत्तरपद-अन्यपदार्थं संज्ञायाम्	A5	उपमान-उत्तरपद-कर्मधारयः	K5
संख्यापूर्वपद-वंशयोत्तरपदः	A6	अवधारणापूर्वपद-कर्मधारयः	K6
पारे-मध्ये-पूर्वपद षष्ठ्युत्तरपदः	A7	सम्भावनापूर्वपद-कर्मधारयः	K7
		मध्यमपदलोपी-कर्मधारयः	Km

compound type:	तत्पुरुषः	compound type:	बहुव्रीहिः
प्रथमातत्पुरुषः	T1		
द्वितीयातत्पुरुषः	T2	द्वितीयार्थ-बहुव्रीहिः(समानाधिकरणः)	Bs2
तृतीयातत्पुरुषः	T3	तृतीयार्थ-बहुव्रीहिः(समानाधिकरणः)	Bs3
चतुर्थीतत्पुरुषः	T4	चतुर्थ्यर्थ-बहुव्रीहिः(समानाधिकरणः)	Bs4
पञ्चमीतत्पुरुषः	T5	पञ्चम्यर्थ-बहुव्रीहिः(समानाधिकरणः)	Bs5
षष्ठीतत्पुरुषः	T6	षष्ठ्यर्थ-बहुव्रीहिः(समानाधिकरणः)	Bs6
सप्तमीतत्पुरुषः	T7	सप्तम्यर्थ-बहुव्रीहिः(समानाधिकरणः)	Bs7
नञ्त्तत्पुरुषः	Tn	दिग्वाचक-बहुव्रीहिः(समानाधिकरणः)	Bsd
समाहार-द्विगुः	Tds	प्रहरणविषयक-बहुव्रीहिः(समानाधिकरणः)	Bsp
तद्धितार्थद्विगुः	Tdt	ग्रहणविषयक-बहुव्रीहिः(समानाधिकरणः)	Bsg
उत्तरपदद्विगुः	Tdu	अस्त्यर्थ-मध्यमपदलोपी(नञ्)-बहुव्रीहिः	Bsmn
गतिसमासः	Tg	प्रादि-बहुव्रीहिः	Bvp
कुसमासः	Tk	संख्योभयपद-बहुव्रीहिः(समानाधिकरणः)	Bss
प्रादिसमासः	Tp	उपमानपूर्वपद-बहुव्रीहिः(समानाधिकरणः)	Bsu
मयूरव्यंसकादिः	Tm	व्यधिकरण-बहुव्रीहिः	Bv
तत्पुरुषः बहुपदः	Tb	संख्योत्तपदः व्यधिकरण-बहुव्रीहिः	Bvs
तत्पुरुषः उपपदः	U	सहपूर्वपद-व्यधिकरण-बहुव्रीहिः	BvS
		उपमानपूर्वपद-व्यधिकरण-बहुव्रीहिः	BvU
		बहुपद-बहुव्रीहिः	Bb

compound type:	द्वन्द्वः
इतरेतरयोग-द्वन्द्वः	Di
समाहार-द्वन्द्वः	Ds
एकशेषः	E
compound type:	केवल
केवलसमासः	S
compound type:	द्विरुक्तिः
द्विरुक्तिः	d

3 General Guidelines

- The tagging of a compound involves the following
 - separating the constituent padas by ‘-’s,
 - Undoing the sandhi,
 - changing the samāsa purvapadas to their pratipadikas,
 - assigning a tag.

Below we give some examples of samasta padas and their tagging.

राजपुरुषः	<राजन्-पुरुषः>T6
लक्ष्मीच्छाया	<लक्ष्मी-छाया>T6
धनुर्विद्या	<धनुष्-विद्या>T6
षण्मासेन	<षट्-मासेन>Tds

- Do not split the padas in following cases
 - तिङन्त or कृदन्त with upa-padas as in आगत्य, प्रहारेण etc.
 - If the words have रूढ-अर्थः as in मण्डपः, प्रातराशः, उपाहारः etc.
- In case either the पूर्वपद or the उत्तरपद or both in turn are समस्तपदs, then they are also to be tagged. E.g.
 तपस्स्वाध्यायनिरतम् should be tagged as <<तपस्-स्वाध्याय>Di-निरतम् >T7
 Note the use of ‘<’ and ‘>’ to indicate the words constituting the compound.
- If a pada is क्रिया-विशेषण in a given context, then it should be marked as Bvs only.

- In case there are more than one possible tags, then show both the tags separated by '|'. E.g.
वृक इव अविरतचिन्तः पशुनः, the word अविरत is ambiguous between Tn and T7. Hence it should be tagged as:
<अ-विरत>Tn-चिन्तः >Bs6 | <अवि-रत>T7-चिन्तः>Bs6
- Handling taddhita constituents:
If a constituent of a compound is a taddhita formed from a compound pratipadika, then, the taddhita suffix is to be added after indicating the samāsa tag. E.g.
a) हेममालिकीर्तिः is to be tagged as <<<हेम-माला>T6 ^ इ>-कीर्तिः>T6.
Note the use of 'i' suffix, which indicates a taddhita pratyaya.
b) एकार्णवीकृते = <एक-अर्णव>K1 ^ ई>-कृते
c) रामकृष्णवत् = <राम-कृष्ण>Di ^ वत्
d) मोहननामक = <मोहन-नाम ^ क>Bs6
- In case the vighraha vākya must specify the number, then we suggest you to specify the number information while tagging the compound, as in
नृपतिः = <नृ{3}-पतिः>T6
धर्माभिरामप्रियः = <<धर्म-अभिराम>Bs6{3}-प्रियः>T6
- Similarly, if the लिङ्ग information is also required, it may also be specified as
सकृपम् = <स-कृपम् {P}>Bs, if कृपम् refers to कृपाचार्य, or
= <स-कृपम् {S}>Bs, if the pratipadika of कृपम् is कृपा.

4 Rules for generating vighraha vākya

Though a substantial amount of literature on Sanskrit Compounds is available, for the benefit of annotators, we give below the vighraha vākya for each type of the compound, with an example. Further, for the benefit of the programmers, we also give rules to generate a vighraha vākya from the properly tagged compound.

Thus each of the example below consists of name of a compound, its major class, its tag, an example, paraphrase describing the meaning of the example compound(विग्रहवाक्य), and the rule to get the paraphrase mechanically (wherever possible) from its components.

4.1 Major class: अव्ययीभावः

1) compound type:	अव्ययीभावः
compound sub-type:	अव्यय-पूर्वपदः
tag:	A1
Example:	उपकृष्णम्
Example with tag:	<उप-कृष्णम् >A1
paraphrase:	कृष्णस्य समीपम्
paraphrase rule:	<x-y>A1 => y{6} f{x} where f maps x to a noun with same semantic content. A function f needs to be defined.

2) compound type:	अव्ययीभावः
compound sub-type:	अव्यय-उत्तरपदः
tag:	A2
Example:	अक्षपरि
Example with tag:	<अक्ष-परि>A2
paraphrase:	अक्षेण विपरीतम् वृत्तम्
paraphrase rule:	<x-y>A2 => x{3} विपरीतम् वृत्तम्

3) compound type:	अव्ययीभावः
compound sub-type:	तिष्ठदुप्रभृति
tag:	A3
Example:	तिष्ठदु
Example with tag:	<तिष्ठद्-गु>A2
paraphrase:	तिष्ठन्ति गावः यस्मिन् देशे
paraphrase rule:	List to be given; collect from पाणिनि's अष्टाध्यायी

4) compound type:	अव्ययीभावः
compound sub-type:	संख्यापूर्वपद-नद्युत्तरपदः
tag:	A4
Example:	सप्तगङ्गम्
Example with tag:	<सप्तन्-गङ्गम् >A4
paraphrase:	सप्तानाम् गङ्गानाम् समाहारः
paraphrase rule:	<x-y>A4 => x{6} y'{6} समाहारः y' is the प्रातिपदिक of y.

5) compound type:	अव्ययीभावः
compound sub-type:	नद्युत्तरपद-अन्यपदार्थे संज्ञायाम्
tag:	A5
Example:	उन्मत्तगङ्गम्
Example with tag:	<उन्मत्त-गङ्गम् >A5
paraphrase:	उन्मत्ता गङ्गा यस्मिन् देशे
paraphrase rule:	<x-y>A5 => x'{1} y{1} यस्मिन् देशे y' is the प्रातिपदिक of y. x' is derived from x by changing the gender to that of y'.

6) compound type:	अव्ययीभावः
compound sub-type:	संख्यापूर्वपद-वंशयोत्तरपदः
tag:	A6
Example:	त्रिमुनि
Example with tag:	<त्रि-मुनि>A6
paraphrase:	त्रयाणाम् मुनीनाम् समाहारः
paraphrase rule:	<x-y>A6 => x{6} y{6} समाहारः the number of x and y will be plural except when x = द्वि they are in द्विवचन

7) compound type:	अव्ययीभावः
compound sub-type:	पारे-मध्ये-पूर्वपद षष्ठ्युत्तरपदः
tag:	A7
Example:	पारेगङ्गम्
Example with tag:	<पारे-गङ्गम् >A7
paraphrase:	गङ्गायाः-पारे
paraphrase rule:	<x-y>A7 => y{6} x

4.2 Major class: तत्पुरुषः

8) compound type:	तत्पुरुषः
compound sub-type:	प्रथमातत्पुरुषः
tag:	T1
Example:	अर्धपिप्पली
Example with tag:	<अर्ध-पिप्पली>T1
paraphrase:	अर्धम् पिप्पल्याः
paraphrase rule:	<x-y>T1 => x{1} y{6}

9) compound type:	तत्पुरुषः
compound sub-type:	द्वितीयातत्पुरुषः
tag:	T2
Example:	कृष्णाश्रितः
Example with tag:	<कृष्णा-श्रितः>T2
paraphrase:	कृष्णम् श्रितः
paraphrase rule:	<x-y>T2 => x{2} y

10) compound type:	तत्पुरुषः
compound sub-type:	तृतीयातत्पुरुषः
tag:	T3
Example:	शङ्कुलाखण्डः
Example with tag:	<शङ्कुला-खण्डः>T3
paraphrase:	शङ्कुलया खण्डः
paraphrase rule:	<x-y>T3 => x{3} y

11) compound type:	तत्पुरुषः
compound sub-type:	चतुर्थीतत्पुरुषः
tag:	T4
Example:	यूपदारुः
Example with tag:	<यूप-दारुः>T4
paraphrase:	यूपाय दारुः
paraphrase rule:	<x-y>T4 => x{4} y

12) compound type:	तत्पुरुषः
compound sub-type:	पञ्चमीतत्पुरुषः
tag:	T5
Example:	चोरभयम्
Example with tag:	<चोर-भयम् >T5
paraphrase:	चोरात् भयम्
paraphrase rule:	<x-y>T5 => x{5} y

13) compound type:	तत्पुरुषः
compound sub-type:	षष्ठीतत्पुरुषः
tag:	T6
Example:	दशरथपुत्रः
Example with tag:	<दशरथ-पुत्रः>T6
paraphrase:	दशरथस्य पुत्रः
paraphrase rule:	<x-y>T6 => x{6} y

14) compound type:	तत्पुरुषः
compound sub-type:	सप्तमीतत्पुरुषः
tag:	T7
Example:	अक्षशौण्डः
Example with tag:	<अक्ष-शौण्डः>T7
paraphrase:	अक्षेषु शौण्डः
paraphrase rule:	<x-y>T7 => x{7} y

15) compound type:	तत्पुरुषः - नञ्
compound sub-type:	नञ्तत्पुरुषः
tag:	Tn
Example:	अब्राह्मणः/अनश्वः
Example with tag:	<न-ब्राह्मणः>Tn / <न-अश्वः>Tn
paraphrase:	न ब्राह्मणः/ न अश्वः
paraphrase rule:	<x-y>Tn => न y

16) compound type:	तत्पुरुषः - द्विगुः
compound sub-type:	समाहार-द्विगुः
tag:	Tds
Example:	पञ्चगवम्
Example with tag:	<पञ्चन्-गवम् >Tds
paraphrase:	पञ्चानाम् गवाम् समाहारः
paraphrase rule:	<x-y>Tds => x{6;ba} y{6;ba} समाहारः

17) compound type:	तत्पुरुषः
compound sub-type:	तद्धितार्थद्विगुः
tag:	Tdt
Example:	अष्टकपालः
Example with tag:	<अष्टन्-कपालः>Tdt
paraphrase:	अष्टसु कपालेषु संसृतः
paraphrase rule:	No Specific Rule

18) compound type:	तत्पुरुषः
compound sub-type:	उत्तरपदद्विगुः
tag:	Tdu
Example:	पञ्चगवधनः
Example with tag:	<<पञ्चन्-गव>Tdu-धनः>>Bs
paraphrase:	
paraphrase rule:	

19) compound type:	तत्पुरुषः
compound sub-type:	गतिसमासः
tag:	Tg
Example:	संगृह्य
Example with tag:	<सम्-गृह्य>Tg
paraphrase:	No paraphrase, since it is a 'नित्य' compound
paraphrase rule:	

20) compound type:	तत्पुरुषः
compound sub-type:	कुसमासः
tag:	Tk
Example:	कुपुरुषः/कापुरुषः
Example with tag:	<कु-पुरुषः>Tk / <का-पुरुषः>Tk
paraphrase:	No paraphrase, since it is a 'नित्य' compound
paraphrase rule:	

21) compound type:	तत्पुरुषः
compound sub-type:	प्रादिसमासः
tag:	Tp
Example:	प्राचार्यः
Example with tag:	<प्र-आचार्यः>Tp
paraphrase:	प्रकृष्टः आचार्यः
paraphrase rule:	<x-y>Tp =<fx y>
	Meanings of upasargas (fx) need to be listed.

22) compound type:	तत्पुरुषः
compound sub-type:	मयूरव्यंसकादिः
tag:	Tm
Example:	राजान्तरम्
Example with tag:	<राजन्-अन्तरम् >Tm
paraphrase:	
paraphrase rule:	<x-y>Tm => ?? गणपाठ is there. So no rule for making विग्रहवाक्य is required. The list should be given

23) compound type:	तत्पुरुषः
compound sub-type:	तत्पुरुषः बहुपदः
tag:	Tb
Example:	बृहत्सेनम्
Example with tag:	<द्वि-अहि-सेनम् >Tb
paraphrase:	
paraphrase rule:	<x-y-z>Tb = x{1} y{1} z{1}

24) compound type:	तत्पुरुषः
compound sub-type:	तत्पुरुषः उपपदः
tag:	U
Example:	कुम्भकारः
Example with tag:	<कुम्भ-कारः>U
paraphrase:	कुम्भम् करोति
paraphrase rule:	<x-y>U => x{2} y' Here y' is the prathama puruṣa ekavacana rūpa of the verb whose kṛdanta

4.3 Major class: कर्मधारयः

25) compound type:	कर्मधारयः
compound sub-type:	विशेषण-पूर्वपद-कर्मधारयः
tag:	K1
Example:	नीलोत्पलम्
Example with tag:	<नील-उत्पलम् >K1
paraphrase:	नीलम् तत् उत्पलम् च
paraphrase rule:	<x-y>K1 => x{1} तत् y{1} च

26) compound type:	कर्मधारयः
compound sub-type:	विशेषण-उत्तरपद-कर्मधारयः
tag:	K2
Example:	ब्राह्मणबालः
Example with tag:	<ब्राह्मण-बालः >K2
paraphrase:	ब्राह्मणः च बालः च
paraphrase rule:	<x-y>K2 => x{1} च y{1} च

27) compound type:	कर्मधारयः
compound sub-type:	विशेषण-उभयपद-कर्मधारयः
tag:	K3
Example:	मन्दशीतलः
Example with tag:	<मन्द-शीतलः >K3
paraphrase:	मन्दः च असौ शीतलः च
paraphrase rule:	<x-y>K3 => x{1} च असौ y{1} च

28) compound type:	कर्मधारयः
compound sub-type:	उपमान-पूर्वपद-कर्मधारयः
tag:	K4
Example:	मेघश्यामः
Example with tag:	<मेघ-श्यामः >K4
paraphrase:	मेघः इव श्यामः
paraphrase rule:	<x-y>K4 => x{1} इव y{1}

29) compound type:	कर्मधारयः
compound sub-type:	उपमान-उत्तरपद-कर्मधारयः
tag:	K5
Example:	पुरुषव्याघ्रः
Example with tag:	<पुरुष-व्याघ्रः>K5
paraphrase:	पुरुषः व्याघ्रः इव
paraphrase rule:	<x-y>K5 => x{1} y{1} इव

30) compound type:	कर्मधारयः
compound sub-type:	अवधारणा-पूर्वपदः
tag:	K6
Example:	गुरुदेवः
Example with tag:	<गुरु-देवः>K6
paraphrase:	गुरुः एव देवः
paraphrase rule:	<x-y>K6 => x{1} एव y{1}

31) compound type:	कर्मधारयः
compound sub-type:	संभावना-पूर्वपदः
tag:	K7
Example:	अयोध्यानगरी
Example with tag:	<अयोध्या-नगरी>K7
paraphrase:	अयोध्या इति नगरी
paraphrase rule:	<x-y>K7 => x{1} इति y{1}

32) compound type:	कर्मधारयः
compound sub-type:	मध्यमपदलोपी
tag:	Km
Example:	शाकपार्थिवः
Example with tag:	<शाक-पार्थिवः>Km
paraphrase:	शाकप्रियः पार्थिवः
paraphrase rule:	<x-y>Km => x{1} z* y{1} z* is a missing madhyama pada.

4.4 Major class: बहुव्रीहिः

33) compound type:	बहुव्रीहिः
compound sub-type:	द्वितीयार्थ-बहुव्रीहिः(समानाधिकरणः)
tag:	Bs2
Example:	प्राप्तोदकः
Example with tag:	<प्राप्त-उदकः>Bs2
paraphrase:	प्राप्तं उदकं यम्
paraphrase rule:	<x-y>Bs2 => x{1} y{1} यत् {g}{2} where 'g' is the gender of the given compound.

34) compound type:	बहुव्रीहिः
compound sub-type:	तृतीयार्थ-बहुव्रीहिः(समानाधिकरणः)
tag:	Bs3
Example:	ऊढरथः
Example with tag:	<ऊढ-रथः>Bs3
paraphrase:	ऊढः रथः येन
paraphrase rule:	<x-y>Bs3 => x{1} y{1} येन/यया/येन

35) compound type:	बहुव्रीहिः
compound sub-type:	चतुर्थ्यर्थ-बहुव्रीहिः(समानाधिकरणः)
tag:	Bs4
Example:	दत्तवस्त्रा
Example with tag:	<दत्त-वस्त्रा>Bs4
paraphrase:	दत्तम् वस्त्रम् यस्यै
paraphrase rule:	<x-y>Bs4 => x{1} y{1} यस्मै/यस्यै/यस्मै

36) compound type:	बहुव्रीहिः
compound sub-type:	पञ्चम्यर्थ-बहुव्रीहिः(समानाधिकरणः)
tag:	Bs5
Example:	अपगतजीवः
Example with tag:	<अपगत-जीवः>Bs5
paraphrase:	अपगतः जीवः यस्मात्
paraphrase rule:	<x-y>Bs5 => x{1} y{1} यस्मात्/यस्याः/यस्मात्

37) compound type:	बहुव्रीहिः
compound sub-type:	षष्ठ्यर्थ-बहुव्रीहिः(समानाधिकरणः)
tag:	Bs6
Example:	पीताम्बरः
Example with tag:	<पीत-अम्बरः>Bs6
paraphrase:	पीतम् अम्बरम् यस्य
paraphrase rule:	<x-y>Bs6 => x{1} y{1} यस्य/यस्याः/यस्य

38) compound type:	बहुव्रीहिः
compound sub-type:	सप्तम्यर्थ-बहुव्रीहिः(समानाधिकरणः)
tag:	Bs7
Example:	निक्षिप्तविश्वासः
Example with tag:	<निक्षिप्त-विश्वासः>Bs7
paraphrase:	निक्षिप्तः विश्वासः यस्मिन्
paraphrase rule:	<x-y>Bs7 => x{1} y{1} यस्मिन्/यस्याम्/यस्मिन्

39) compound type:	बहुव्रीहिः
compound sub-type:	दिग्वाचक-बहुव्रीहिः(समानाधिकरणः)
tag:	Bsd
Example:	पूर्वोत्तरा
Example with tag:	<पूर्व-उत्तरा>Bsd
paraphrase:	पूर्वस्याः च उत्तरस्याः च यदन्तरालम्
paraphrase rule:	<x-y>Bsd => x{6} च y{6} च यदन्तरालम्

40) compound type:	बहुव्रीहिः
compound sub-type:	प्रहरणविषयक-बहुव्रीहिः(समानाधिकरणः)
tag:	Bsp
Example:	दण्डादण्डि
Example with tag:	<दण्डा-दण्डि>Bsp
paraphrase:	दण्डैः च दण्डैः च प्रहत्य इदम् युद्धम् प्रवृत्तम्
paraphrase rule:	<x-y>Bsp => x{3} च y{3} च प्रहत्य इदम् युद्धम् प्रवृत्तम्

41) compound type:	बहुव्रीहिः
compound sub-type:	ग्रहणविषयक-बहुव्रीहिः(समानाधिकरणः)
tag:	Bsg
Example:	केशाकेशि
Example with tag:	<केशा-केशि>Bsg
paraphrase:	केशेषु केशेषु गृहीत्वा इदम् युद्धम् प्रवृत्तम्
paraphrase rule:	<x-y>Bsg => x{7}-y{7} गृहीत्वा इदम् युद्धम् प्रवृत्तम्

42) compound type:	बहुव्रीहिः
compound sub-type:	अस्त्यर्थ-मध्यमपदलोपी-(नञ्)बहुव्रीहिः
tag:	Bsmn
Example:	अपुत्रः
Example with tag:	<अ-पुत्रः>Bsmn
paraphrase:	न विद्यते पुत्रः यस्य
paraphrase rule:	<x-y>Bsmn => न विद्यते-y{1} यस्य/यस्याः/यस्य

43) compound type:	बहुव्रीहिः
compound sub-type:	प्रादि-बहुव्रीहिः
tag:	Bvp
Example:	निर्दयः
Example with tag:	<निर्-दयः>Bvp
paraphrase:	निर्गता दया यस्मात्
paraphrase rule:	

44) compound type:	बहुव्रीहिः
compound sub-type:	संख्योभयपद-बहुव्रीहिः(समानाधिकरणः)
tag:	Bss
Example:	त्रिचतुरः
Example with tag:	<त्रि-चतुरः>Bss
paraphrase:	त्रयः वा चतुरः वा यस्य
paraphrase rule:	<x-y>Bss = > x{1} वा y{1} यस्य/यस्याः/यस्य

45) compound type:	बहुव्रीहिः
compound sub-type:	उपमान-पूर्वपद-बहुव्रीहिः(समानाधिकरणः)
tag:	Bsu
Example:	चन्द्रमुखी
Example with tag:	<चन्द्र-मुखी>Bsu
paraphrase:	चन्द्र इव मुखम् यस्याः
paraphrase rule:	<x-y>Bsu => x{1} इव y{1} यस्य/यस्याः/यस्य

46) compound type:	बहुव्रीहिः
compound sub-type:	व्यधिकरण-बहुव्रीहिः
tag:	Bv
Example:	कण्ठकालः/चन्द्रशेखरः
Example with tag:	<कण्ठ-कालः>Bv/<चन्द्र-शेखरः>Bv
paraphrase:	कण्ठे कालः यस्य/चन्द्रः शेखरे यस्य
paraphrase rule:	<x-y>Bv => x y{1} यस्य/यस्याः/यस्य

47) compound type:	बहुव्रीहिः
compound sub-type:	संख्योत्तरपद-व्यधिकरण-बहुव्रीहिः
tag:	Bvs
Example:	उपदशाः
Example with tag:	<उप-दशाः>Bvs
paraphrase:	दशानाम् समीपे ये सन्ति ते
paraphrase rule:	<x-y>Bvs => y{6} x' ये सन्ति ते

48) compound type:	बहुव्रीहिः
compound sub-type:	सहपूर्वपद-व्यधिकरण-बहुव्रीहिः
tag:	BvS
Example:	सपुत्रः
Example with tag:	<स-पुत्रः>BvS
paraphrase:	पुत्रेण सह
paraphrase rule:	<x-y>BvS => y{3} सह

49) compound type:	बहुव्रीहिः
compound sub-type:	उपमानपूर्वपद-व्यधिकरण-बहुव्रीहिः
tag:	BvU
Example:	उष्ट्रमुखः
Example with tag:	<उष्ट्र-मुखः>BvU
paraphrase:	उष्ट्रस्य इव मुखम् यस्य
paraphrase rule:	<x-y>BvU => x{6} इव y यस्य/यस्याः/यस्य

50) compound type:	बहुव्रीहिः
compound sub-type:	बहुपद-बहुव्रीहिः
tag:	Bb
Example:	
Example with tag:	
paraphrase:	
paraphrase rule:	

4.5 Major class: द्वन्द्वः

51) compound type:	द्वन्द्वः
compound sub-type:	इतरेतरयोग-द्वन्द्वः
tag:	Di
Example:	रामकृष्णौ
Example with tag:	<राम-कृष्णौ>Di
paraphrase:	रामः च कृष्णः च
paraphrase rule:	<x-y+>Di => x{1} च (y{1} च)+
	Here + indicates one or more occurrences.

52) compound type:	द्वन्द्वः
compound sub-type:	समाहार-द्वन्द्वः
tag:	Ds
Example:	संज्ञापरिभाषम्
Example with tag:	<संज्ञा-परिभाषम् >Ds
paraphrase:	संज्ञा च परिभाषा च एतयोः समाहारः
paraphrase rule:	<x-y+>Ds => x{1} च (y{1} च)+ एतत् n समाहारः Here + indicates one or more occurrences. n=2 if there are only two components. n=3 otherwise.

53) compound type:	एकशेषः
compound sub-type:	एकशेष-द्वन्द्वः
tag:	E
Example:	पितरौ
Example with tag:	<पितरौ>E
paraphrase:	माता च पिता च
paraphrase rule:	Give a list of exceptions with विग्रहवाक्यम् No common rule

54) compound type:	केवल
compound sub-type:	केवल
tag:	S
Example:	भूतपूर्वः
Example with tag:	<भूत-पूर्वः>S
paraphrase:	पूर्वम् भूतः
paraphrase rule:	<x-y>S => y{1} x{1}

55) compound type:	द्विरुक्तिः
compound sub-type:	द्विरुक्तिः
tag:	d
Example:	उपर्युपरि
Example with tag:	<उपरि-उपरि>d
paraphrase:	उपरि उपरि
paraphrase rule:	<x-y>d => x y

5 Examples of compound tagging from बालकाण्ड of वाल्मीकिरामायणम्

Sloka 1.1.1:

<<तपस्-स्वाध्याय>Di-निरतम् >T7 तपस्वी <वाग्-विदाम् >U वरम्
नारदम् परिपप्रच्छ वाल्मीकिः <मुनि-पुङ्गवम् >T7 ||1.1.1||

Sloka 1.1.8:

<<इक्ष्वाकु-वंश>T6-प्रभवः>Bs6 रामः नाम जनैः श्रुतः
<नियत-आत्मा>Bs6 <महा(महत)-वीर्यः>Bs6 द्युतिमान् धृतिमान् वशी ||1.1.8||

Sloka 1.1.14:

<<<वेद-<वेद-अङ्ग>T6>Di-तत्त्व>T6-ज्ञः>U <धनुर्-वेदे>T6 च निष्ठितः
<<<<सर्व-शास्त्र>K1-अर्थ>T6-तत्त्व>T6-ज्ञः>U स्मृतिमान् प्रतिभानवान् ||1.1.14||

6 Structure of Sanskrit Compounds

The Sanskrit compounds are binary in nature (with an exception of द्वन्द्व, and बहुपद-बहुव्रीहि). Hence they can be faithfully represented as binary trees as in Figure 1. The analysis shown in this figure may be represented in a linear notation as <A-<B-C>> . We add a tag to each of the compounds labeling its name. Thus the compound ABC after proper labeling will be <A-<B-C>tag1>tag2, where tag1 is the name of compound formed by the words B and C, and tag2 is the name of the compound formed by A and BC.

The grammar for validation of tagged compounds is given below.

7 Grammar of tagged compounds

```
compound: '<' component '-' component '>' tag
| '<' component '-' component '>' tag taddhita
| '<' component '-' component '>' tag number
| '<' component '-' component '>' tag gender
| '<' component '-' dvandvacomponents '>' dvandvatag
| '<' component '>' Etag
;
```

```
dvandvacomponents: dvandvacomponents '-' component
```

```
| component
```

```
;
```

```
component: pada
```

```
| compound
```

```
;
```

```
tag: A[1-7]
```

```
| Bs[2-7]
| Bs[dpgsu]
| Bsmn
| Bv[sSU]
| B[bv]
| K[1-7]
| Km
| T[1-7]
| T[bgkmnp]
| Td[stu]
| [ESUd]
;

dvandvatag: D[is]
;

Etag : E
;
pada: [a-zA-Z]+
;
```