

Chapter 6

Sanskrit compound type identifier

6.1 Type Identifier

After getting a constituency parse of a compound, the next task in the compound analysis is to assign an appropriate operator (tag) to each non-leaf node. This operator then operates on the leaf nodes to produce the associated meaning. For instance,

<<sumitrā-ānanda>-vardhanaḥ>

is tagged as

<<sumitrā-ānanda>T6-vardhanaḥ>T6

Tag specifies the relation between the components. This relation is semantic in nature and is not expressed by any morpheme. So it is only the semantics of the components involved that determines this relation. Hence unless one knows the पदार्थs (the meaning of a word) of the components, one can't determine the meaning of a compound. That means deciding the relation between the components is not easy. There are two basic questions one needs to answer.

1. What is the nature of the relations ?
2. Where are the clues that help in deciding the relations ?

Below we give two examples that help answer these questions. Consider the compounds राजपुरुषः, दशरथपुत्रः and वृक्षशाखा. In the first case the relation between राज and पुरुषः is that of servant-master (सेव्यसेवकभावः), in the second it is father-son (पिता-पुत्रसम्बन्धः) and in the third case it is part of (अवयव-अवयविसम्बन्धः). However, in all the three cases instead of specifying these deeper relations, the relation between the components is expressed by the genitive case suffix in the paraphrase of these compounds as राज्ञः पुरुषः, दशरथस्य पुत्रः and वृक्षस्य शाखा, and hence these are classified as षष्ठीतत्पुरुष. So, for deciding the relation between components, instead of deeper semantic analysis, we resort to the Paninian classification of compounds. At the same time it was observed earlier the coarse grain distinction of compounds into four major classes is not sufficient. We need the fine grain distinction, as specified in Appendix - B.

Now consider another compound रामेश्वरः, depending on the context this may mean

- a) रामश्चासौ ईश्वरश्च
- b) राम ईश्वरः यस्य सः
- c) रामस्य ईश्वरः

In the first case it is कर्मधारय, in the second case it is बहुव्रीहि and in the third it is षष्ठी तत्पुरुष. Though the semantic content of the components is the same in all the three cases, the relations between these components is determined by other words in the context.

Thus the compound type is guided by the paraphrase of a compound, the paraphrase depends on the context, and the present day computer technology is still not fully equipped to handle context. So on the face of it the problem seems to be difficult to handle. However a close look at the Pāṇini's sūtras dealing with compounds, provide us a lot of semantic clues. For example, look at the following sūtras

"द्वितीया श्रितातीतपतितगतात्यस्तप्राप्तापन्नैः" (2.1.24)

"तृतीया तत्कृतार्थेन गुणवचनेन" (2.1.30)

"अन्नेन व्यञ्जनम्" (2.1.34)

"पञ्चमी भयेन" (2.1.37)

Each of these gives a criterion either in terms of semantics or as a list of possible components in a compound of a particular type. This helps in deciding the compound types. However, such rules do not cover all the cases. Only for a few type of compounds, which is typically a closed list, the rules are provided. To decide the type in all other cases, we need to look at the semantics of the components involved.

There are two possible ways to handle the semantics. The first one is to use semantically tagged lexicon and build a rule-base for identification. Another approach is to use manually tagged corpus and machine learning algorithms to develop automatic identifier. In what follows we first look at all the relevent sūtra from Pāṇini and discuss the feasibility of their implementation. Next we discuss the building of automatic tagger using the manually tagged corpus and some simple statistical insights.

6.2 समासविधायकसूत्रs from Pāṇini's Aṣṭādhyāyī

In the whole Aṣṭādhyāyī the process of compound formation is optional and comes under the sūtra विभाषा "2-1-11" That is the speaker has an option to use either a paraphrase such as राज्ञः पुरुषः or a compounded form राजपुरुषः to express the same meaning. There are a few exceptions to it viz. यथाशक्ति.

Given a paraphrase Pāṇini describes the process of compound formation. To be precise, the process starts with the अलौकिक-विग्रह and not the लौकिक-विग्रह. Then the order of the components is decided followed by its type, the process then deals with the elision of the vibhakti of intermediate components, assigning the svaras, addition of the certain suffixes to change the gender if necessary and finally the compound form is generated.

The सूत्रs dealing with compounds then can be broadly classified into two types

- a) सूत्रs that deal with the semantic content to decide the type of a compound.
- b) सूत्रs that deal with the process of compound formation involving
 - (i) deciding the word order
 - (ii) deletion of विभक्ति
 - (iii) assigning a स्वर

The second type of सूत्रs are thus useful from generation point of view. They deal with the morphology and phonology. The first type of सूत्रs provide semantic clues for deciding the type of a compound. See the Appendix-C. We describe below whether each of these सूत्रs is useful in deciding the type of a compound, followed by the difficulties in using some of

them computationally. To explain the aphorisms below, we have referred the books "The Astadhyayi of Panini"¹ and "The Siddhanta Kaumudi of Bhattoji Dixita"².

6.2.1 अव्ययीभावः

1. सूत्र - "अव्ययं विभक्ति-समीप-समृद्धि-व्युद्घर्थाभावात्प्रया-सम्प्रति-शब्दप्रादुर्भाव..." - 2.1.6

Meaning - An indeclinable (is invariably compounded with a case-inflected word) conveying the meaning of (1) a case suffix, (2) nearness, (3) prosperity, (4) misfortune, (5) absense of a thing, (6) passing (of a particular time), (7) not (the proper time) now, (8) reputation of a name, (9) after, (10) yathā, (11) one coming after the other, (12) simultaneity, (13) resemblance, (14) fittingness, (15) totality, (16) end (the compound being called avyayībhāva).

Description - This is a vidhisūtra as well as a saññāsūtra. This aphorism provides a list of semantic clues to the initial component of a compound of type अव्यय-पूर्वपद-अव्ययीभाव. Below we give a list of examples.

- (1) विभक्ति (a case suffix) - अधिहरि is an example of विभक्तिवचन and हरौ इति is its paraphrase. Here अधि is an indeclinable word and it forces the word हरि to be into seventh case-suffix.
- (2) सामीप्य (nearness) - उपराजम् is an instance of सामीप्य. Here the word उप indicates the meaning 'nearness' to राजा and it also indicates that the second component will contain the sixth case-suffix. राज्ञः

¹Vol-V and VI, written by S. D. Joshi and JAF. Roodbergen, Published by Sahitya Academy

²Srisa Chandra Vasu, Published by Motilal Banarssidas publishers

समीपम् is the paraphrase of उपराजम्.

- (3) समृद्धि (prosperity) - मद्राणां समृद्धिः is the paraphrase of सुमद्रम् and it is derived in the sense of समृद्धि. Here सु is an indeclinable word.
- (4) व्यृद्धि (misfortune) - दुर्यवनम् and दुर्गवादिकम् are the instances of व्यृद्धि. Here दुर् is an indeclinable word and it is used in the sense of misfortune.
- (5) अभाव (absense of a thing) - निर् is an indeclinable which is mostly used in the sense of absense. For instance निर्माक्षिकम् means मक्षिकाणाम् अभावः (absense of flies).
- (6) अत्यय (passing (of a particular time)) - the word अत्यय means passing, departure or destruction of a particular time. For instance निर्हिमम् means हिमस्य अत्ययः (the snows have passed).
- (7) असम्प्रति (not (the proper time) now) - The word असम्प्रति suggests a time which is not proper to do. For instance अतितैसुकम्, derived in the sense of उपभोगस्य वर्तमानकालप्रतिषेधः 'the rejection of the present for use'. The काशिकावृत्ति explains that तैसुक means अच्छादन 'cloak'. So the compound means that now is not the proper time to wear a cloak.
- (8) शब्दप्रादुर्भाव (reputation of a name) - इतिपाणिनि derived in the sense of प्रकाशता शब्दस्य (the fame of a name). Here the word पाणिनि is a reputed name.
- (9) पश्चात् (after) - In the sense of after indeclinables such as अनु are used. For instance अनुविष्णु means विष्णोः पश्चात् "after the Vishnu".
- (10) यथार्थ (yathā) - The word यथा has various meanings such as योग्यता(Capability), वीप्सा(pervasion), पदार्थानतिवृत्ति(not going beyond

the word's meaning) and सादृश्यम्(Similarity). For instances अनु-
रूपम्, प्रत्यर्थम्, यथाशक्ति etc.

- (11) आनुपूर्व्यं (one coming after the other) - As अनुज्येष्ठं प्रविशन्तु भवन्तः (let your honors enter in the order of seniority)
- (12) यौगपद्य (simultaneity) - The word सचक्रम् is derived in the sense of चक्रेण युगपत् or सहचक्रेण.
- (13) सादृश्य (resemblance) - ससखि, derived in the sense of सदृशः किरव्यः 'one similar to a monkey'. Why mention सादृश्य as a separate category, if सादृश्य has already been stated as one of the meaning of यथा? The idea is that by this separate statement of सादृश्य compound-formation can also be justified when the notion of सादृश्य is not the main one, but the subordinate one.
- (14) सम्पत्ति (fittingness) - सब्रह्म 'in a way fitting for a Brahmin'.
- (15) साकल्य (totality) - सतृणम् अभ्यवहरति 'He eats right down to the grass'.
- (16) अन्त (end) - साग्नि अधीते 'he studies up to the end of (the section on) अग्नि'.

Clue for the type identification - Though it is not possible to decide the type of compounds in all the cases listed above, still a few can be classified automatically. If the initial component is an indeclinable such as प्र, परा, अप, सम्, अनु etc. and the final component is in neuter gender then the compound may be an अव्यय-पूर्वपद-अव्ययीभाव.

2. सूत्र - "यावदवधारणे" (2-1-8)

Meaning - (the indeclinable) यावत् in the sense of अवधारण

(specification) (is invariably compounded with a case-inflected word the compound being called अव्ययीभाव.)

Description - The current aphorism requires the context of अवधारण (limitations) to decide the type of an अव्ययीभाव compound. For instance यावच्छ्लोकम् (यावन्तः श्लोकाः तावन्तः अच्युतप्रणामाः).

Clue for the type identification - If the पूर्वपद is यावत् and उत्तरपद is in neuter gender then the compound may be an अव्यय-पूर्वपद-अव्ययीभाव compound, and such compounds typically carry the meaning अवधारण.

3. सूत्र - "सुप्रतिना मात्रार्थे" (2-1-9)

Meaning - A case inflected word (is invariably compounded with the indeclinable) prati in the sense of mātrā "a small quantity" (being called अव्ययीभाव).

Description - A word ending in a case-affix is compounded with the word प्रति in the context of mātrā (a small quantity) and the compound is called an अव्ययीभाव.

Clue for the type identification - If the उत्तरपद is प्रति the compound may be an अव्यय-उत्तरपद-अव्ययीभाव and the meaning of प्रति is a small quantity. These are very rare and the only instances we know of are सूपप्रति "a small quantity of soup" and शाकप्रति "a small quantity of vegetables".

4. सूत्र - "अक्षशलाकासंख्याः परिणा" (2-1-10)

Meaning - (The case inflected words) akṣa "(cubic) die", śalākā

"(oblong) die" and numerals (are invariably compounded) with pari. Like previous aphorism (2-1-9), this is also a special rule of अव्ययीभाव. The compound is confined to words denoting loss in gambling (being called अव्ययीभाव).

Examples - अक्षपरि (an unlucky throw of dice), शलाकापरि (an unlucky throw of oblong), एकपरि (an unlucky throw by one over).

Clue for the type identification - If the पूर्वपद contains the words such as अक्ष, शलाका or numerals like एक, द्वि etc. and the उत्तरपद contains an indeclinable word परि then the compound may be of type अव्यय-उत्तरपद-अव्ययीभाव.

5. सूत्र - "अपपरिबहिरञ्चवः पञ्चम्या" (2-1-12)

Meaning - (The case-inflected words) apa, pari, bahis and (those ending in) -अञ् are marginally (compounded with (a word ending in) an ablative case ending (being called अव्ययीभाव).

Examples - अपविष्णु संसारः or अप विष्णोः संसारः (The world is outside or away from Viṣṇu). So also बहिरग्रामम् or बहिरग्रामात् (outside the village) etc.

Clue for the type identification - If पूर्वपद contains अप, परि and बहि or any word ending in -अञ् and the उत्तरपद contains either fifth case-affix or is of neuter gender then the compound is called अव्यय-पूर्वपद-अव्ययीभाव.

Problem - Here, it is not easy to know whether the initial word ends in अञ् or not unless we have the derivational information of all the head words in the lexicon. However, if such a rich lexicon is

available, then automatic identification is possible.

6. सूत्र - "आङ् मर्यादाऽभिविध्योः" (2-1-13)

Meaning - (The case-inflected word) आङ् in the sense of मर्यादा (limit exclusive) or अभिविधि (limit inclusive) (is marginally compounded with a case-inflected word (being called अव्ययीभाव).

Examples - आपाटलिपुत्रं or आपाटलिपुत्रात् वृष्टो देवः (It rained up to Pāṭaliputra), आकुमारम् or आकुमारेभ्यः यशः पाणिनेः (The fame of Pāṇini extends even to the boys) etc.

Clue for the type identification - If the पूर्वपद contains the indeclinable आङ् and the उत्तरपद is either in neuter gender or fifth case-affix then the compound is an अव्यय-पूर्वपद-अव्ययीभाव.

7. सूत्र - "लक्षणेनाऽभिप्रती आभिमुख्ये" (2-1-14)

Meaning - (the case-inflected words) अभि and प्रति in the sense of आभिमुख्य "direction towards" (are marginally compounded) with (a case-inflected word indicating) the goal (the compound when formed, conveying the same meaning as the formally corresponding wordgroup and being called अव्ययीभाव).

Example - अभ्यग्नि शलभाः पतन्ति "the moths fall in the direction of fire" etc.

Clue for the type identification - If the first component is either अभि or प्रति, this is अव्यय-पूर्वपद-अव्ययीभाव. However, it is not possible to decide the meaning of अभि or प्रति.

8. सूत्र - "अनुर्यत्समया" (2-1-15)

Meaning - (the case-inflected word) अनु (is marginally compounded with a case-inflected word standing for an indicator) nearness to which (is conveyed, the compound when formed, conveying the same meaning as the formally corresponding wordgroup and being called अव्ययीभाव).

Example - अनुवनमशनिर्गतः "the thunder bolt fell near the forest" etc.

Clue for the type identification - As above, if the first component of the compound is अनु then it may be classified as अव्यय-पूर्वपद-अव्ययीभाव. However, it is not possible to determine the meaning of अनु mechanically.

9. सूत्र - "यस्य चायामः" (2-1-16)

Meaning - also (the case-inflected word अनु is marginally compounded with a case-inflected word standing for an indicator) whose length (is conveyed, the compound, when formed, being called अव्ययीभाव).

Examples - अनुगङ्गं वाराणसी "the city of Banaras extends alongside the river Ganges", अनुयमुनं मथुरा "Mathura alongside of the Yamuna, the length of Mathura being measured by that of the Yamuna".

Clue for the type identification - The clue, as in the above rule, is in the first component अनु and the type of the compound is अव्यय-पूर्वपद-अव्ययीभाव compound. However, meaning of अनु can not be determined mechanically.

10. सूत्र - "तिष्ठद्गुप्रभृतीनि च" (2-1-17)

Meaning - This is a विधिसूत्र which assigns the designation अव्ययीभाव to specifically enumerated words (being called अव्ययीभाव).

Description - By this aphorism Panini mentions all irregular forms of अव्ययीभाव compounds. All the words are listed in गणपाठ. For instance तिष्ठद्गु (at milking time) etc. These are all treated as special cases and handled separately.

11. सूत्र - "पारे मध्ये षष्ठ्या वा" (2-1-18)

Meaning - (the case-inflected words) pāre (across) and madhye (in) are preferably compounded with (a case-inflected word) ending in a genitive case ending (being called अव्ययीभाव).

Examples - पारेगङ्गात् or गङ्गापारात् (bring across the Ganges), मध्येगङ्गात् or गङ्गामध्यात् (middle of the Ganges) etc.

Clue for the type identification - If पूर्वपद contains the words पारे or मध्ये and the उत्तरपद contains the fifth case-affix then the compound is called पारे-मध्ये-पूर्वपद-षष्ठ्युत्तरपद-अव्ययीभाव. Similarly if second component contains either पारात् or मध्यात् then also the compound is called पारे-मध्ये-पूर्वपद-षष्ठ्युत्तरपद-अव्ययीभाव.

12. सूत्र - "संख्या वंशयेन" (2-1-19)

Meaning - A numeral is preferably compounded with a case-inflected word signifying one who belongs to a family (being called अव्ययीभाव).

Examples - त्रिमुनि, द्विमुनि etc.

Clue for the type identification - In this kind of compound, the first component contains a numeral and the second component is a वंश्यवाची. To implement this aphorism one has to provide the lexicon with वंश्यवाची tag. Such a compound will be tagged as संख्यापूर्वपद-वंश्योत्तरपद-अव्ययीभाव.

13. सूत्र - "नदीभिश्च" (2-1-20)

Meaning - Also (a numeral is preferably compounded with case-inflected words) signifying a river (being called अव्ययीभाव).

Examples - द्वियमुनम्, सप्तगङ्गम् etc.

Clue for the type identification - To implement this aphorism, one has to enrich the lexicon with the names of the rivers. If the first component of compound contains a numeral and the second component contains the नदीवाचक words, then it is of type संख्यापूर्वपद-नद्युत्तरपद-अव्ययीभाव compound.

6.2.2 तत्पुरुषः

14. सूत्र - "द्वितीया श्रितातीतपतितगतात्यस्तप्राप्तापन्नैः" (2-1-24)

Meaning - (a word ending in) the second case-termination (is preferably compounded) with (the case-inflected words) श्रित "who has fallen upon", अतीत "who has passed beyond", पतित "who has fallen into", गत "who has gone to", अत्यस्त "who has thrown beyond", प्राप्त "who has attained" and आपन्न "who has obtained" (being called तत्पुरुष).

Description - This is a विधिसूत्र prescribing the preferred formation of

an accusative तत्पुरुष compound with कान्त forms conveying an active meaning. The examples quoted by काशिकावृत्ति are :

1. कष्टश्रितः 'who has landed in trouble', derived in the sense of कष्टं श्रितः.
2. कान्तारातीतः 'who has passed beyond the forest', derived in the sense of कान्तारम् अतीतः.
3. नरकपतितः 'who has fallen into hell', derived in the sense of नरकं पतितः.
4. ग्रामगतः 'who has gone to his village', derived in the sense of ग्रामं गतः.
5. तरङ्गात्यस्तः 'who has thrown beyond the waves', derived in the sense of तरङ्गान् अत्यस्तः.
6. सुखप्राप्तः 'who has attained happiness', derived in the sense of सुखं प्राप्तः.
7. सुखापन्नः 'who has obtained happiness', derived in the sense of सुखम् आपन्नः.

Clue for the type identification - If the stem of the second component of a compound is from the list mentioned above viz. श्रित, अतीत, पतित etc. then this is an instance of द्वितीया-तत्पुरुष compound.

15. सूत्र - "स्वयं केन" (2-1-25)

Meaning - (the case-inflected word) स्वयं "(by) oneself" (is preferably compounded) with a case-inflected word) ending in क्त (being called तत्पुरुष).

Description - This is a विधिसूत्र prescribing तत्पुरुष compound-formation. The indeclinable word स्वयम् 'oneself', is compounded with a word ending in the affix क्त, and the resulting compound is

called तत्पुरुष. The anuvritti of the word द्वितीया being inappropriate does not take place; though, however it is needed in the next aphorism. Because the word स्वयम् being an indeclinable, cannot take any case-affix. As स्वयं कृतस्यापत्यं = स्वायङ्कृतिः.

Clue for the type identification - If the first component contains an indeclinable स्वयं or its variation स्वायम् and the second component contains an inflected क्तन्त word then it is an instance of तत्पुरुष compound.

16. सूत्र - "खद्वा क्षेपे" (2-1-26)

Meaning - (the case-inflected word) खद्वा "bed" (ending in the second case-termination is compounded with a case-inflected word ending in the suffix क्त) when (the sense of) abuse is to be conveyed (the compound being called तत्पुरुष).

Example - खद्वारूढः जाल्मः (literally lying on a bed) ; silly, stupid, going wrong or astray.

Clue for the type identification - If the first component contains the word खद्वा and the second component is an inflected क्तन्त then we may postulate that it is likely to be a द्वितीयातत्पुरुष compound. The censure context is necessary for confirmation, and mechanical detection of censure context being impossible, we can just mark this as a possibility and can not confirm the type.

17. सूत्र - "सामि" (2-1-27)

Meaning - (the case-inflected word) सामि "incompletely, half" (is

preferably compounded with a case-inflected word ending in क् (being called तत्पुरुष).

Description - This is a विधिसूत्र prescribing a तत्पुरुष compound-formation. The indeclinable word सामि meaning 'half' is compounded with a word ending in the affix क्, and the resulting compound would be तत्पुरुष. As सामिकृतम् 'half-done'.

Clue for the type identification - If the first component contains an indeclinable सामि and the second component contains a क्कान्त word then it is a तत्पुरुष compound.

18. सूत्र - "कालाः" (2-1-28)

Meaning - (a case-inflected word ending in the second case-termination) signifying (a period of) time (is preferably compounded with a case-inflected word ending in क् (being called तत्पुरुष).

Description - This is विधिसूत्र prescribing तत्पुरुष compound formation. The words denoting the time, (but not the duration thereof) being in the accusative case, are optionally compounded with a word ending in the suffix क्, and the resulting compound is तत्पुरुष.

Example - मासप्रमितश्चन्द्रमाः 'the new moon' (literally, the moon that has begun to measure the month).

Clue for the type identification - For handling this type of compound, the list of the words which denote the time is required. If such a word is the first component and the second component contains the क्कान्त word and it may be classified as तत्पुरुष.

19. सूत्र - "पूर्वसदृशसमोनार्थकलहनिपुणमिश्रश्लक्ष्णैः" (2-1-31)

Meaning - (a case-inflected word ending in the third case-termination is preferably compound) with (case-inflected words) conveying the sense of पूर्व "previous", सदृश "like", सम "same", ऊन "less", and with (the case-inflected words) कलह "fight", निपुण "skilful", मिश्र "mixed" and श्लक्ष्ण "smooth" (being called तत्पुरुष).

Examples - From this aphorism we learn incidentally that the words पूर्व etc. govern the instrumental case. The examples quoted by the काशिकावृत्ति are :-

- (a) मासपूर्वः derived in the sense of मासेन पूर्वः, who is earlier by (one) month.
- (b) मातृसदृशः 'resembling his mother', पितृसदृशः 'resembling his father'.
- (c) मातृसमः 'equal to his mother'.
- (d) माषोनम् 'what is short by (one) माष.
- (e) असिकलहः 'a fight with swords', वाक्कलहः 'a fight with words'.
- (f) वान्निपुणः 'skilful with words', आचारनिपुणः 'skilful in behaviour'.
- (g) गुडमिश्रः 'mixed with molasses'.
- (h) आचारश्लक्ष्णः 'gentle in behaviour'.

Clue for the type identification - The current aphorism gives the list of the possible second components and in such cases it may be तृतीयातत्पुरुष compound.

20. सूत्र - "चतुर्थी तदर्थाथबलिहितसुखरक्षितैः" (2-1-36)

Meaning - (a case-inflected word) ending in the fourth case-termination (is preferably compounded) with (a case-inflected word

signifying) a thing for the sake of that (meaning expressed by the word in the fourth case) and with (the case-inflected words अर्थ "thing", बलि "food-offering", हित "beneficial", सुख "pleasant" and रक्षित "reserved" (being called तत्पुरुष).

Description - By the expression तदर्थ 'for the purpose there of', the special relation of a material and its modification alone is, held to be intended.³ This we infer from the ज्ञापक of the words बलि and रक्षित used in the aphorism. As यूपाय दारुः = यूपदारुः 'wood for stake' (that is wood which by modification will be changed into a stake). But not so here. रन्धनाय स्थाली 'pot for cooking'. So also or अवहननायोलुखलम् 'the wood mortar for threshing'.

Form this aphorism we may also infer that the Dative case conveys also the sense of 'for the purpose there of'. The word तदर्थार्थ is taken as one word by some, meaning 'a thing serviceable there to'.

Examples - भूतबलिः 'a sacrifice for भूत', गोहितम् 'what is good for cows', गोसुखम् 'what is pleasant for cows' and गोरक्षितम् 'what is kept for cows'.

Clue for the type identification - The current aphorism gives the list of words for second components. If we find any one of the word from the list as a second component then we may assume that the given compound is a चतुर्थीतत्पुरुष compound.

21. सूत्र - "पञ्चमी भयेन" (2-1-37)

Meaning - (a case-inflected word) ending in the fifth case ending (is preferably compounded) with (the case-inflected word) भय "fear"

³चतुर्थी तदर्थमात्रेण चेत्सर्वप्रसङ्गः । सर्वस्य चतुर्थ्यन्तस्य तदर्थमात्रेण सह समासः प्राप्नोति । अनेनापि प्राप्नोति - रन्धनाय स्थाली, अवहननायोलुखलमिति । किं कारणम् ? अविशेषात् । न हि कश्चिद्विशेष उपादीयते - एवंजातीयकस्य चतुर्थ्यन्तस्य तदर्थेन सह समासो भवति इति । अनुपादीयमाने विशेषे सर्वप्रसङ्गः ।

(being called तत्पुरुष).

Examples - चोरात् भयम् = चोरभयम् 'fear from thieves'.

Clue for the type identification - The current aphorism gives the information regarding the second component. In the compound, if the second component contains the words भय, भीति or भी then the compound type can be identified as a पञ्चमीतत्पुरुष.

22. सूत्र - "अपेतापोढमुक्तपतितापत्रस्तैरल्पशः" (2-1-38)

Meaning - 'in a few instances a case-inflected word ending in the fifth case-termination is preferably compounded) with (the case-inflected words) अपेत "parted from", अपोढ "caried off", मुक्त "loosened from", पतित "fallen from" and अपत्र "recoiling from" (being called तत्पुरुष).

Examples - सुखापेतः 'gone away from pleasure', कल्पनापोढः 'carried away beyond imagination', चक्रमुक्तः 'freed from the wheel', स्वर्गपतितः 'fallen from heaven', तरङ्गापत्रस्तः 'afraid of the waves'.

Why is the word अल्पशः 'when the action is gradual' used? It shows the limited range of this kind of compounds. Not every ablative word can be so compounded. Hence, there is no compounding at all in the following case :- प्रसादात् पतितः, 'fallen from the mansion'. For here, the fall, is violent and sudden and not gradual and slight.

Clue for the type identification - The current aphorism gives the information regarding the second component of a तत्पुरुष compound. In a compound, if we get one of these words then the compound type can be identified as a पञ्चमीतत्पुरुष.

23. सूत्र - "स्तोकान्तिकदूरार्थकृच्छ्राणि केन" (2-1-39)

Meaning - (case-inflected words ending in the fifth case-termination) conveying the sense of स्तोक 'a little', अन्तिक 'near' or दूर 'far' and (the case-inflected word ending in the fifth case-termination) कृच्छ्र 'difficulty' (are preferably compounded) with (a case-inflected word derived with the suffix) क्त (being called तत्पुरुष).

Examples - स्तोकान् मुक्त 'loosed from a little distance', अन्तिकादागतः 'come from near', अभ्याशादागतः 'come from near', दूरादागतः 'come from far', विप्रकृष्टादागतः 'come from a distance', कृच्छ्रादागतः 'come with difficulty'.

Clue for the type identification - The current aphorism explains a अलुक्समास and gives a list of first components in a तत्पुरुष compound. In a compound, if we get one of these words on fifth case of as a first component and the second component is a क्त ending कृदन्त then the compound type can be identified as a पञ्चमीतत्पुरुष.

24. सूत्र - "सप्तमी शौण्डैः" (2-1-40)

Meaning - (a case-inflected word) ending in the seventh case-termination (is preferably compounded) with (the case-inflected words) शौण्ड 'cunning', etc (being called तत्पुरुष).

Example - अक्षशौण्डः 'skilled in dice'.

The current aphorism explains a सप्तमीतत्पुरुष compound and gives a list of second component. In the aphorism the word शौण्डैः suggests a शौण्डादि-गण(group) from गणपाठ and the words are :- 1) शौण्ड 2) धूर्त 3) कित 4) व्याड 5) प्रवीण 6) संवीत 7) अन्तर 8) अधि 9) पटु 10) पण्डित 11) कुशल 12)

चपल 13) निपुण 14) संव्याड 15) भन्य and 16) समीर.

Clue for the type identification - In a compound, if the second component is from शौण्डादिगण then the compound type may be identified as a सप्तमीतत्पुरुष.

25. सूत्र - "सिद्धशुष्कपक्वबन्धैश्च" (2-1-41)

Meaning - 'also (a case-inflected word ending in the seventh case-termination is preferably compounded) with (the case-inflected words) सिद्ध 'well-known', शुष्क 'dried', पक्व 'cooked' and बन्ध 'binding'(being called तत्पुरुष).

Examples - साङ्काश्यसिद्धः 'perfect in साङ्काश्य', आतपशुष्कः 'dried in sun', स्थालीपक्वः 'cooked in pot' and चक्रबन्धः 'bound on the wheel'.

Clue for the type identification - The current aphorism gives a list of second components. In a compound, if the second component contains one of these words then the compound type may be identified as सप्तमीतत्पुरुष.

26. सूत्र - "ध्वाङ्गेण क्षेपे" (2-1-42)

Meaning - (a case-inflected word ending in the seventh case-termination is compounded) with (the case inflected word) ध्वाङ्ग 'crow' when (the compound conveys) the sense of abuse (the compound being called तत्पुरुष).

Example - तीर्थध्वाङ्गः 'a crow at the scared bathing place'.

Clue for the type identification - In a compound if the second component carries the word ध्वाङ्ग in the meaning of 'censure' then the

compound type may be identified as सप्तमीतत्पुरुष.

27. सूत्र - "केनाहोरात्रावयवाः" (2-1-45)

Meaning - (a case-inflected word ending in the seventh case-termination signifying) parts of the day or night (is preferably compounded) with (a case-inflected word) derived with (the suffix) क्त (being called तत्पुरुष).

Examples - पूर्वाह्नकृतम् 'done in the morning', अपररात्रकृतम् 'done in the last part of the night'.

Clue for the type identification - In a compound, if the first component contains words relating with name of divisions of day or night and the second component contains a word with ending with क्त suffix the the compound type can be identified as सप्तमीतत्पुरुष.

28. सूत्र - "तत्र" (2-1-46)

Meaning - (the case-inflected word) तत्र "there" (ending in a seventh case-termination is preferably compounded with a case-inflected word derived by means of the suffix क्त (being called तत्पुरुष).

Example - तत्रभुक्तम् 'eaten there'.

Clue for the type identification - In a compound, if the first component contains the word तत्र and the second component contains the word which ends with क्त suffix then the compound type can be identified as सप्तमीतत्पुरुष.

29. सूत्र - "पात्रेसमितादयश्च" (2-1-48)

Meaning - also (the words) पात्रेसमिता etc (when the sense of abuse is to be conveyed, the compounds being called तत्पुरुष).

Examples - All these are irregular compounds. Some of the words contained in this list are compounds formed with the past-participle (क्त). Thus the following is the list of the words :- पात्रेसमिताः । पात्रेबहुलाः । उदुम्बरमशकाः । उदरकृमिः । कूपकच्छपः । कूपचूर्णकः । अवटकच्छपः । कूपमण्डूकः । कुम्भमण्डूकः । उदपानमःडूकः । नगरकाकः । नगरवायसः । मातरिषुरुषः । पिण्डीषूरः । पितरिषूरः । गेहेशूरः । गेहेनर्दी । गेहेक्ष्वेडी । गेहेविजिती । गेहेव्याडः । गेहेमेही । गेहेदाही । हेहेदसः । गेहेधृष्टः । गर्भेतृप्तः । आखनिकबकः । गोष्ठेशूरः । गोष्ठे विजिती । गोष्ठेक्ष्वेडी । गोष्ठेपटुः । गोष्ठेपण्डितः । गोष्ठेप्रगल्भः । कर्णेतिट्टिभः । कर्णेतिरिटिरा । कर्णेचुरचुरा । The use of च in aphorism is to restrict it to these very words. Therefore we cannot compound परमाः पात्रेसमिताः, and it cannot be part of another compound.

30. सूत्र - "पूर्वकालैकसर्वजरत्पुराणनवकेवलाः समानाधिकरणेन" (2-1-49)

Meaning - (a case inflected word denoting an action) which precedes in time and (the case-inflected words) एक "single", सर्व "all", जरत् "new", पुराण "ancient", नव "new" and केवल "only" (only compounded) with a syntactically agreeing (case-inflected word, the compound being called तत्पुरुष).

Examples - स्नातानुलिप्तः "bathed and perfumed", कृष्टसमीकृतम् "ploughed and levelled", एकनाथः "havin one master", सर्वयाज्ञिकाः "all the members of a sacrifice", सर्वमनुष्याः "all men", जरन्नैयायिकाः "old logician", पुराणमीमांसकाः "old मीमांसकs" etc.

Clue for the type identification - The aphorism provides a list of

first components. In a compound, if first component contains one of these words then the compound type may be identified as a कर्मधारय.

31. सूत्र - "पापाणके कुत्सितैः" (2-1-54)

Meaning - (the case-inflected words) पाप "bad" and अणक "gossipy" (are compounded) with (syntactically agreeing, case-inflected words signifying) objects of contempt (the compound being called तत्पुरुष).

Description - Both the words पाप and अणक are words of contempts(कुत्सन); by the last aphorism, they have stood as second members in the compound; the present aphorism, however is so framed with regard to aphorism 1-2-43 and 2-2-30 that they will stand as first.

Examples - पापनापितः or अणकनापितः 'a contemptible barber', पाप or अणक-कुलालः 'a contemptible potter'.

Clue for the type identification - The current aphorism explains a तत्पुरुष compound. In a compound, if the first component constains the words पाप or अणक then the compound type may be identified as a उभय-पद-विशेषण-कर्मधारय. To implement this aphorism, one needs to prepare a list of कुत्सित words.

32. सूत्र - "उपमितं व्याघ्रादिभिः सामान्याप्रयोगे" (2-1-56)

Meaning - (a case-inflected word signifying) the object compared (is preferably compounded) with (the syntactically agreeing, case-inflected words) व्याघ्र "tiger", etc. provided that (a word signifying) the common property is not used (being called तत्पुरुष).

Description - This is a modification of last aphorism by which the substantive (उपमान) would have stood first, by the present, the उपमेय or the qualified stands first.

Example - पुरुषोऽयं व्याघ्र इव = पुरुषव्याघ्रः 'a person-tiger' (in strength). In previous aphorism उपमानानि सामान्यवचनैः" (2-1-55), the compounding was between the उपमान and the उपमेय. In the present, the compounding is between the उपमित and certain उपमान words but never with सामान्यवचन. In the current aphorism, the word व्याघ्रादिभिः suggests the व्याघ्रादिगण and the words belong व्याघ्रादिगण are :- 1) व्याघ्र "A tiger", 2) सिंह "A lion", 3) ऋक्ष "A bear", 4) ऋषभ "A bull", 5) चन्दन "Sandal", 6) वृक "A wolf", 7) वृष "A bull", 8) वराह "A boar, hog", 9) हस्तिन् "A elephant", 10) तरु "A tree", 11) कुञ्जर "A elephant", 12) रुरु "A kind of deer", 13) पृषत "The spotted antelope", 14) पुण्डरीक "A lotus flower", 15) पलाश "A tree, Butea Frondosa" and 16) कितव "A rogue, cheat".

Clue for the type identification - Any compound whose second component is from any of the above list, it is possibly an उपमान-उत्तरपद-कर्मधारय compound.

33. सूत्र - "विशेषणं विशेष्येण बहुलम्" (2-1-57)

Meaning - (a case-inflected word signifying) a qualifier (is) variously (compounded) with (a syntactically agreeing case-inflected word signifying) the item qualified (the compound conveying the same meaning as the formally corresponding wordgroup and being called तत्पुरुष).

Example - नीलोत्पलम् 'a blue lotus'.

Clue for the type identification - To implement this aphorism one needs a list of विशेषणs, which are typically marked in the dictionaries. Any compound whose first compound is विशेषण-पूर्वपद-कर्मधारय compound.

34. सूत्र - "पूर्वापरप्रथमचरमजघन्यसमानमध्यमध्यमवीराश्च" (2-1-58)

Meaning - (the case-inflected words) पूर्व "previous", अपर "later", प्रथम "first", चरम "last", जघन्य "low", समान "same", मध्य "middle", मध्यम "middle" and वीर "brave" (are preferably compounded with a syntactically agreeing case-inflected word and the compound being called तत्पुरुष).

Examples - पूर्ववैयाकरणः 'Previous grammarian', अपरवैयाकरणः 'later grammarian' etc.

Clue for the type identification - If the compound has any of the words from the list पूर्व, अपर, प्रथम, चरम etc. as its first components then it may be a विशेषण-पूर्वपद-कर्मधारय compound.

35. सूत्र - "श्रेण्यादयः कृतादिभिः" (2-1-59)

Meaning - (the case-inflected words) श्रेणि "guild" etc. (are compounded) with (the syntactically agreeing, case-inflected words) कृत "made", etc. (the compound being called तत्पुरुष).

Example - श्रेणीकृतम् 'made into guild' etc.

Clue for the type identification - If the first component of the compound is from श्रेण्यादिगण and the second component from the कृतादिगण then it is a कर्मधारय compound.

36. सूत्र - "क्तेन नञ्विशिष्टेनानञ्" (2-1-60)

Meaning - (a case-inflected word) not containing (the negative particle) नञ् (is preferably compounded) with (a syntactically agreeing case-inflected word ending in the suffix) क्त which is distinguished (from the other constituent) by (the fact that it does contain) नञ् (being called तत्पुरुष).

Examples - कृताकृतम् "done and not done" i.e. partly done and partly not done.

Clue for the type identification - If a compound has both the components ending in क्त and one as a negation of the other, then it is of type विशेषण-उभयपद-कर्मधारय compound.

37. सूत्र - "सन्महत्परमोत्तमोत्कृष्टाः पूज्यमानैः" (2-1-61)

Meaning - (the case-inflected words सत् "good", महत् "great", परम "highest", उत्तम "best" and उत्कृष्ट "eminent" (are preferably compounded) with (syntactically agreeing case-inflected words signifying) objects which are being honoured/praised (being called तत्पुरुष).

Description - सद्द्वैद्यः "a good physician". The महत् becomes महा in compound such as महापुरुषः "A great man" परमपुरुषः "the highest person", उत्तमपुरुषः "the best person", उत्कृष्टपुरुषः "the excellent person".

Clue for the type identification - If the first component contains adjectives words like सत्, महत्, परम, उत्तम and उत्कृष्ट then the compound is of type विशेषण-पूर्वपद-कर्मधारय compound.

38. सूत्र - "वृन्दारकनागकुञ्जरैः पूज्यमानम्" (2-1-62)

Meaning - (A case-inflected word signifying) an object which is being honoured/praised (is preferably compounded) with the syntactically agreeing, case-inflected words) वृन्दारक (a kind of "deity", नाग "elephant" (being called तत्पुरुष).

Description - गोवृन्दारकः "An excellent bull or cow".

Clue for the type identification - If the second component contains adjectives words like वृन्दारक, नाग and कुञ्जर then the compound may be of type विशेषण-उत्तरपद-कर्मधारय compound.

39. सूत्र - "कतरकतमौ जातिपरिप्रश्ने" (2-1-63)

Meaning - (the case-inflected words) कतर "who (out of two)?", कतम "who (out of many)?" (are preferably compounded with a syntactically agreeing, case-inflected word, when a question (is asked) regarding (belonging to a particular) Vedic school (the compound being called तत्पुरुष).

Examples - कतरकठः "who (out of two) belongs to the कठ school", कतमकलापः "who (out of these many) belong to the कलाप school".

Clue for the type identification - If the first component contains the words कतर and कतम and the second component contains a जातिवाचक word then it is of type विशेषण-पूर्वपद-कर्मधारय.

40. सूत्र - "किं क्षेपे" (2-1-64)

Meaning - (the case-inflected word) किम् "who", which, what?" (is

compounded with a syntactically agreeing, case-inflected word) when (the compounded conveys the sense of) abuse (the compound being called तत्पुरुष).

Examples - किंराजा यो न रक्षति? "he is a bad king who does not protect his subjects.", किं सखा योऽभिद्रुहति "he is a bad friend who hates".

Clue for the type identification - If the first component contains the word किम् then it is of type विशेषण-पूर्वपद-कर्मधारय compound.

41. सूत्र - "पोटायुवतिस्तोककतिपयगृष्टिधेनुवशावेहद्वृक्षयणीप्रवक्तृश्रोत्रियाध्यापकधूर्तैर्जातिः"
(2-1-65)

Meaning - (a case-inflected word signifying) a जाति "class/generic notion" (is compounded) with the syntactically agreeing, case-inflected words) पोटा "a female with male characteristics", युवति "a young female", स्तोक "a little", कतिपय "some", गृष्टि "a cow which had just one calf", धेनु "a sucking cow", वशा "a barren cow", वेहत् "a cow that miscarries", वृक्षयणी "a cow having a young calf", प्रवक्तृ "one who expounds", श्रोत्रिय "a learned reciter of the Vedas", अध्यापक "a teacher of शास्त्रिक texts" and धूर्त "a shrewd one" (being called तत्पुरुष).

Examples - इभपोटा "a young female elephant", इभयुवतिः "a female elephant", अग्निस्तोकः "a little fire", उद्भिश्चत् कतिपयम् "a little butter milk", गोगृष्टिः, गोधेनुः, गोवशाः, गोवेहत्, गोवृक्षयणी, कठप्रवक्ता "an expounder of कठ", कठश्रोत्रियः "a ब्राह्मण who has mastered the कठ branch of the यजुर्वेद", कठाध्यापकः "a teacher of the कठ branch of the यजुर्वेद".

Clue for the type identification - If the second component contains पोटा, युवति, स्तोक, कतिपय, गृष्टि, धेनु, वशा, वेहद्, वृक्षयणी, प्रवक्तृ, श्रोत्रिय, अध्यापक

and धूर्त then it is of type विशेषण-उत्तरपद-कर्मधारय.

42. सूत्र - "प्रशंसावचनैश्च" (2-1-66)

Meaning - also (a case-inflected word signifying a जाति "class/generic notion" (is preferably compounded) with (syntactically agreeing case-inflected words) signifying praise (being called तत्पुरुष).

Description - गोमतल्लिका "an excellent cow", गोमचर्चिका "an excellent cow" etc.

Clue for the type identification - If the first component of the compound is a जाति and the second component is a word used in the sense of praising such as मतल्लिका, मचर्चिका, उद्ध, तल्लज etc.⁴ then the compound may be of type विशेषण-उत्तरपद-कर्मधारय.

43. सूत्र - "युवा खलतिपलितवलिनजरतीभिः" (2-1-67)

Meaning - (the case-inflected word) युवन् "young" (is preferably compounded) with (the syntactically agreeing, case-inflected words) खलति "bald", पतित "grey-haired", वलिन "wrinkled" and जरति "old woman" (being called तत्पुरुष).

Description - युवखलति: "bald in youth", युवपलित: "grey-headed in youth", युववलिन: "wrinkled in youth" and युवजरती "appearing old in youth" etc.

Clue for the type identification - In a compound, if the first component contains the word युवन् and the second component contains any one of the word from the list such as खलति, पलित, वलिन

⁴अमरकोषः

etc. then the compound type may be indentified as विशेषण-उत्तरपद-कर्मधारय.

44. सूत्र - "वर्णो वर्णेन" (2-1-69)

Meaning - (a case-inflected word signifying) a colour (is preferably compounded) with (a syntactically agreeing, case-inflected word signifying) a colour (being called तत्पुरुष).

Example - कृष्णसारङ्गः "spotted antelope".

Clue for the type identification - If both of the components of a compound are वर्णवाची then it is of type of विशेषण-उभयपद-कर्मधारय. To implement this aphorism one has to prepare the list of वर्णवाची words.

45. सूत्र - "चतुष्पादो गर्भिण्या" (2-1-71)

Meaning - (case-inflected words signifying) a quadruped (are preferably compounded) with (the syntactically agreeing, case-inflected word) गर्भिणी "pregnant" (being called तत्पुरुष).

Examples - गोगर्भिणी "a pregnant cow", अजागर्भिणी "a pregnant she-goat" etc.

Clue for the type identification - The current aphorism explains a तत्पुरुष compound where first component contains a word in the sense of quadruped and the second component contains the word गर्भिणी and the compound type can be indentified as a विशेषण-उत्तरपद-कर्मधारय compound. To implement this one needs a list of quadruped animals.

46. सूत्र - "मयूरव्यंसकादयश्च" (2-1-72)

Meaning - also (the compound) मयूरव्यंसक 'a cunning peacock' etc are called तत्पुरुष.

Examples - All these are irregular compounds. Thus the following is the list of the words :- मयूरव्यंसकः । छात्रव्यंसकः । काम्बोजमुण्डः । यवनमुण्डः । छन्दसि हस्तेगृह्य । पादेगृह्य । लाङ्गलेगृह्य । पुनर्दाय । एहीडादयोऽन्यपदार्थे एहीडम् । एहियवं वर्तते । एहिवाणिजाक्रिया । अपेहिवाणिजा । प्रेहिवाणिजा । एहिस्वागता । अपोहिस्वागता । प्रेहिस्वागता । एहिवितीया । अपेहिवितीया । इहवितर्का । प्रोहकटा । अपोहकटा । प्रोहकर्दमा । अपोहकर्दमा । उद्धरचूडा । आहरचेला । आहरवसना । आहरवनिता । कृन्तविचक्षणा । उद्धरोत्सृजा । उद्धमविधमा । उत्पचिविपचा । उत्पतनिपता । उच्चावचम् । उच्चनीचम् । अचितोपचितम् । अवचितपराचितम् । निश्चप्रचम् । अकिञ्चनः । स्नात्वाकालकः । पीत्वास्थिरकः । भुक्त्वासुहितः । प्रोष्यपापीयान् । उत्पत्यपाकला । निपत्यरोहिणी । निषण्णाश्यामा । अपेहिप्रसवा । इहपञ्चमी । इहद्वितीया । जहिकर्मणा बहुलमाभीक्ष्ये कर्तारं चाभिदधाति जहजोडः । उज्जहिजोडः । जहिस्तम्बः । उज्जहिस्तम्बः । आख्यातमाख्यातेन क्रियासातत्ये अश्रीतपिबता । पचतभृजता । खादतमोदता । खादतवमता । खादताचमता । आहरनिवपा । आवपनिष्क्रिा । उत्पचिपचा । भिन्धिलवना । छिन्धिविचक्षणा । पचलवना । पचप्रकूटा । The use of च in aphorism is to restrict it to these very words.

47. सूत्र - "अर्धं नपुंसकम्" (2-2-2)

Meaning - (the case-inflected) neuter (word) अर्धम् "a half" (is preferably compounded with a case inflected word signifying a whole, provided that the whole mentioned is one single entity, the compound conveying the same meaning as the formally corresponding wordgroup and being called तत्पुरुष).

Example - अर्धपिप्पली "a half of the पिप्पली" etc.

48. सूत्र - "द्वितीयतृतीयचतुर्थतुर्याण्यन्यतरस्याम्" (2-2-3)

Meaning - (the case-inflected words) द्वितीय "second", तृतीय "third", चतुर्थ "fourth" and तुर्य "fourth" (are) indifferently (compounded with a case-inflected word signifying a whole, provided that the whole mentioned is one single entity (being called तत्पुरुष).

Example - द्वितीयभिक्षा "second begging" etc.

Reason - The clue exist in the first part of the component, viz. that it refers to signified portion, and thus is a पूरण-संख्या.

49. सूत्र - "नञ्" (2-2-6)

Meaning - (the case-inflected words) नञ् (is preferably compounded with a case-inflected word (being called तत्पुरुष).

Examples - अब्राह्मणः "not a Brahmana or Non-Brahmana", अनश्वः "Not a horse" etc.

Clue for the type identification - If the first component contains the word 'अ' or 'अन्' then the compound type may be indentified as नञ्-तत्पुरुष compound.

50. सूत्र - "ईषदकृता" (2-2-7)

Meaning - (the case-inflected word) ईषत् "to some extent, slightly" (is preferably compounded with a case-inflected word) not (ending in a कृत्(-suffix)(being called तत्पुरुष).

Example - ईषत्पिङ्गलः "a little brownish".

Reason - If the first component of the compound is ईषत् then it is likely that the compound is a तत्पुरुष. However for confirmation,

one needs to know whether the word ईषत् is used in the sense of slightly/to some extent. This can be done only manually.

51. सूत्र - "याजकादिभिश्च" (2-2-9)

Meaning - also (a case-inflected word ending in the sixth case termination is preferably compounded) with (the case-inflected words) याजक "one who worships/pays honour to" etc. (being called तत्पुरुष).

Examples - ब्राह्मणयाजकः "a Brahmana's sacrificer" and the words which comes under the याजकादिगण are :- 1) याजक, 2) पूजक, 3) परिचारक, 4) परिवेषक, 5) परिषेचक, 6) स्नापक or स्नातक, 7) अध्यापक, 8) उत्साहक or उत्सादक, 9) उद्वर्तक, 10) होतु, 11) भर्तु, 12) रथगणक, 13) पत्तिगणक, 14) पोतु, 15) हर्तु and 16) वर्तक.

Reason - The current aphorism explains a तत्पुरुष compound and provides a list of some irregular compounds. We treat these as special compounds and do not analyse them further.

52. सूत्र - "कुगतिप्रादयः" (2-2-18)

Meaning - (the case-inflected word) कु "evil", (the case inflected words called) गति and (the case-inflected words mentioned as) प्रादयः "प्र,परा,अप... etc." (are always compounded with a case-inflected word the compound, being called तत्पुरुष).

Description - कुपुरुषः "a sinful man", उररीकृतम् "having asserted", दुष्पुरुषः "a bad man".

The current aphorism explains a तत्पुरुष compound and provides us

three types of compounds 1) कुसमास, 2) गतिसमास and 3) प्रादिसमास.

- (a) कुसमास is used in the sense of निन्दा "censure". For example कुपति: "a bad husband". By looking the first component the type of this kind of compound can be identified easily and can be implemented.
- (b) गतिसमास is found in many senses and need so many informations for implementation such as the second compound always should be a क्रिया not the तिङ्, first component may contain ऊर्यादिगण words or ending with the चि or डाच् suffix etc. For examples उररीकृत्य, शुक्कीकृत्य, पटपटाकृत्य etc.
- (c) प्रादिसमास is found in many senses and can be implemented easily for type identification because the first component always contains a prefix प्र, परा, अति etc. and the second component will be a case-inflected word. For instance प्राचार्यः.

6.2.3 बहुव्रीहि

53. सूत्र - "संख्ययाव्ययासन्नदूराधिकसंख्याः संख्येये" (2-2-25)

Meaning - (case-inflected words called) अव्यय "indeclinable", (the case-inflected words) आसन्न "near", अदूर "not far", अधिक "additional", and (the case-inflected words called) संख्या "numeral" (are compound) with (a case-inflected word called संख्या "numeral", when (the compound) conveys the sense of संख्येय "item to be counted" (the compound being called बहुव्रीहि).

Examples - उपदशाः "those who are neat to ten" or उपविंशः "nineteen or twenty-one".

Clue for the type identification - If the first component contains the indeclinable words and the words like आसन्न, अदूर, अधिक and second component contains the words in sense of the numeral then the compound type may be identified as संख्योत्तरपद-व्यधिकरण-बहुव्रीहि compound.

54. सूत्र - "दिङ्गामान्यन्तराले" (2-2-26)

Meaning - (case-inflected words which are) names of directions (are invariably compounded with each other, the compound) conveying the sense of अन्तराल "intermediate region" and being called बहुव्रीहि.

Examples - दक्षिणपूर्वादिक् "the direction midway between south and east", पूर्वोत्तरा "north-east"

Clue for the type identification - To implement this rule, one needs to prepare the list of दिङ्गाम "names of directions" and has to see whether both of the components contain the words from दिङ्गाम list then the compound type may be identified as दिग्वाचक-बहुव्रीहि compound.

55. सूत्र - "तेन सहेति तुल्ययोगे" (2-2-28)

Meaning - (the case-inflected word) सह "together with" (is always compounded with a case-inflected word represented by the word-pattern) तेन (the compound) conveying the sense of तुल्ययोग "equal connection" (with an action), and being called बहुव्रीहि.

Example - सपुत्रः "with son" etc.

Clue for the type identification - If the first component contains

the word स then it is possible that the compound is of type सह-पूर्वपद-व्यधिकरण-बहुव्रीहि. In case the gender of the second component is different from its default gender, then the compound is definitely of this type. However, if the gender of the second component is same as its regular gender, than there is also a possibility of this compound being an अव्ययीभाव⁵.

Below we summarise the conditions and extra semantic information needed for automatic detection of compound type.

S.No.	Sūtra Reffr.	Conditions		Extra-Info	Type
		Initial word	Final word		
1	2-1-6	All the prefixes or indeclinables like प्र, परा, अप, यथा etc.	neuter gender	-	A1
2	2-1-8	यावत्	Neuter gender	-	A1
3	2-1-9	सु, प्रति	Neuter gender	-	A1
4	2-1-10	अक्ष, शलाका or संख्या: like एक, द्वि etc.	परि	-	A2
5	2-1-12	अप, परि, बहिस् or word ending with अच्	Neuter gender or fifth-case ending word	-	A1

⁵पा०सू० - अव्ययं विभक्तिसमीप... "2-1-6"

6	2-1-13	आङ्	Neuter gender or fifth-case ending word	-	A1
7	2-1-14	अभि, प्रति	Neuter gender or fifth-case ending word	-	A1
8	2-1-15	अनु	Neuter gender	-	A1
9	2-1-16	अनु	Neuter gender	-	A1
10	2-1-17	तिष्ठदुप्रभृतीनि च	तिष्ठद्, etc.	-	A3
11	2-1-18	पारे, मध्ये	fifth-case affix ending word	-	A7
12	2-1-19	The numerals like एक, द्वि	वंश्यवाची words	वंश्यवाची list	A6
13	2-1-20	The numerals like एक, द्वि	नदीवाची words	नदीवाची list	A6
14	2-1-24	-	श्रित, अतीत, पतित, गत, अत्यस्त, प्राप्त, आपन्न	-	T2
15	2-1-25	स्वयम् or variation of स्वयम्	कान्त word	-	T2
16	2-1-26	खद्वा	कान्त word	-	T2
17	2-1-27	सामि	कान्त word	-	T2

18	2-1-28	the list of the words which denote the time	कान्त word	list of words denoting time	T2
19	2-1-31	-	पूर्व, सदृश, सम, ऊ- नार्थ, कलह, निपुण, मिश्र, श्लक्ष्ण	-	T3
20	2-1-36	-	तदर्थ, अर्थ, बलि, हित, सुख, रक्षित	-	T4
21	2-1-37	-	भय, भीति, भी	-	T5
22	2-1-38	-	अपेत, अपोद, मुक्त, पतित, अपत्र	-	T5
23	2-1-39	स्तोका,अन्तिक, दूरार्थ, कृच्छ्र	कान्त word	-	T5
24	2-1-40	सप्तमी शौण्डैः	अक्षशौण्डः	-	T7
25	2-1-41	-	सिद्ध, शुष्क, पक्व, बन्ध	-	T7
26	2-1-42	-	ध्वाङ्ग	-	T7
27	2-1-45	The words relating with name of divisions of day or night	कान्त word	list of words denoting names of divisions of days and night	T7

28	2-1-46	तत्र	क्लान्त word	-	T7
29	2-1-48	पात्रेसमितादयश्च	पात्रेसमिताः, पात्रेव- हुलाः etc.	-	T7
30	2-1-49	पूर्व, काल, एक, सर्व, जरत, पुराण, नव, केवल	-	-	K1
31	2-1-54	पाप, अणक	word related to कुत्सित	list of कुत्सित words	K3
32	2-1-56	उपमितं व्याघ्रादिभिः सामान्याप्रयोगे	पुरुषव्याघ्रः etc.	-	K5
33	2-1-57	any विशेषण	-	list of विशेषणs	K1
34	2-1-58	पूर्व, अपर, प्रथम, चरम, जघन्य, समान, मध्य, मध्यम, वीर	-	-	K1
35	2-1-59	श्रेण्यादयः कृतादिभिः	श्रेणीकृतम् etc.	-	K
36	2-1-60	क्लान्त word	क्लान्त word with negation अ	-	K2
37	2-1-61	सत्, महत्, परम, उत्तम, उत्कृष्ट	-	-	K1
38	2-1-62	-	वृन्दारक, नाग, कुञ्जर	-	K2
39	2-1-63	-	कतर, कतम	-	K1
40	2-1-64	किम्	-	-	K1

41	2-1-65	जातिवाचक word	पोटा, युवति, स्तोक, कतिपय, गृष्टि, धेनु, वशा, वेहत, बष्कयणी, प्रवक्तु, श्रोत्रिय, अध्यापक, धूर्त	list of जातिवाचक words	K2
42	2-1-66	जातिवाचक word	words signifying praise such as मतल्लिका, मचर्चिका, उद्ध, तल्लज etc.	list of जा- तिवाचक and प्रशसावाचक words	K2
43	2-1-67	युवन्	खलति, पलित, वलिन, जरति	-	K2
44	2-1-69	वर्णवाचक word	वर्णवाचक word	list of वर्णवाचक words	K3
45	2-1-71	चतुष्पादवाचक word	गर्भिणी	list of वर्णवाचक words	K2
46	2-1-72	मयूरव्यंसकादि	मयूरव्यंसकः etc.	-	K
47	2-2-2	अर्ध नपुंसकम्	अर्धपिप्पली	-	T1
48	2-2-3	द्वितीयतृतीयचतुर्थ...	द्वितीयभिक्षा etc.	-	K
49	2-2-6	अ/अन्	-	-	Tn
50	2-2-7	ईषत्	-	-	T
51	2-2-9	याजकादिभिश्च	ब्राह्मणयाजकः etc.	-	K
52	2-2-18	कु	-	-	Tk

52.1	2-2-18	प्रादय - प्र, परा, निस्, निर् etc.	-	-	Tg/Tp
53	2-2-25	अव्ययs, आसन्न, दूर, अधिक, संख्या	संख्या in the sense of संख्येय like दश, विंश	list of सं- ख्यावाचक words with special gender ⁶	Bvs
54	2-2-26	names of directions	names of directions	list of दिङ्म "names of directions"	Bsd
55	2-2-28	स	words with special gender	-	BvS
55.1	2-2-28	स	words	-	A1

Till now we have explained the aphorisms which we have implemented to identify the types of various compounds. Now we are going to explain the aphorisms which we could not implement and the reason(s) thereof.

1. सूत्र - "यथाऽसादृश्ये" (2-1-7)

Meaning - (the indeclinable) यथा (is invariably compounded with a case-inflected word) in a meaning other than सादृश्य "similarity" (the compound being called अव्ययीभाव).

⁶The second components in बहुव्रीहि compounds contain the special gender like in द्वित्राः, the stem त्रि contains the masculine gender त्रिः etc.

Example - यथावृद्धम् = ये ये वृद्धाः (every old person).

Reason - The current aphorism explains an अव्ययीभाव compound and its an exceptional rule of 2-1-6. Usally the word यथा is used in four types of meanings :- 1) योग्यता "ability", 2) वीप्सा 3)पदार्थानतिवृत्ति and 4) सादृश्य "similarity". Except the fourth one सादृश्य, one of the meaning of rest of three should be there in the compound then only the compound may be identified as an अव्ययीभाव. It is not possible to check whether the meaning is in the desired sense or not without doing any semantic analysis of the context. Hence it is not possible to decide the type of the given compound.

2. सूत्र - "अत्यन्तसंयोगे च" (2-1-29)

Meaning - also (case-inflected words ending in the second case-termination, signifying a period of time, are compounded with a case-inflected word) when (the sense of) अत्यन्तसंयोग "invariable, uninterrupted connection" is to be conveyed (the compound being called तत्पुरुष).

Example - मुहूर्तसुखम् "a momentary pleasure".

Reason - It is not possible to decide just by looking at the word whether there is अत्यन्तसंयोग or not and hence it is not possible to classify such compounds automatically.

3. सूत्र - "तृतीया तत्कृतार्थेन गुणवचनेन" (2-1-30)

Meaning - (a case-inflected word) ending in the third case-termination (is preferably compounded) with (a case-inflected word) standing for a quality which is an entity caused/brought

about by (what is referred to) by that (being called तत्पुरुष).

Examples - शङ्कुलाखण्डः "cut by nipper", धान्यार्थः "wealth acquired by grain".

Reason - To classify such compounds, it is necessary to decide semantic relation between the components, and this requires the knowledge of योग्यता - semantic compatibility between the word meanings. Therefore it is not possible to classify such compounds purely on the basis of syntax or lexicon.

4. सूत्र - "कर्तृकरणे कृता बहुलम्" (2-1-32)

Meaning - (a case-inflected word ending in the third case-termination) which conveys the sense of कर्तृ "agent" or करण "instrument" (is) variously (compounded) with (a case-inflected word) ending in a कृत-suffix (being called तत्पुरुष).

Examples - अहिना हतः = अहिहतः "killed by a snake", नखभिन्नः "divided by the nails" etc.

Reason - To implement this aphorism, the knowledge of योग्यता is required.

5. सूत्र - "कृत्यैरधिकार्थवचने" (2-1-33)

Meaning - (a case-inflected word ending in the third case-termination conveying the sense of कर्तृ "agent" or of करण "instrument" is compounded) with (a case-inflected word) derived by means of a कृत्य(-suffix), provided that (the compound) conveys the sense of exaggeration (being called तत्पुरुष).

Examples - वातच्छेद्यं तृणं 'thin grass' (so fragile that it can be cut by wind) in the sense of praise, denoting softness, or it may be the reverse ; denoting weakness.) So also काकपेया नदी 'a full river' (so full that a crow may dip his beak into it, and drink while sitting on the bank. In this sense it is praise. It may be censure also, and it will mean a shallow river that a crow may dip his beak into it and touch the bottom and drink.)

Reason - Since this aphorism relies heavily on the context, and the semantics it is not possible to implement this rule computationally.

6. सूत्र - "अन्नेन व्यञ्जनम्" (2-1-34)

Meaning - (a case-inflected word ending in the third case-termination) signifying a flavouring ingredient (is preferably compounded) with (a case-inflected signifying) an article of food (being called तत्पुरुष).

Description - That which is to be prepared is called अन्न and that which prepares is व्यञ्जन as दध्योदनः 'rice prepared or made relishable with curd.' The words 'food' and 'condiment' as represented in the above compounds, are connected in the sense by a verb understood.

Reason - To identify such compounds as तृतीयातत्पुरुष, one has to provide a list of words of अन्नवाची and corresponding व्यञ्जनवाची.

7. सूत्र - "भक्ष्येण मिश्रीकरणम्" (2-1-35)

Meaning - (a case-inflected word ending in the third case-termination signifying) an admixture (is preferably compounds)

with (a case-inflected word signifying) something edible (being called तत्पुरुष).

Example - गुडधाना 'barley prepared with jaggery'.

Anything eatable, whether hard or soft, is called भक्ष्य, its refinement is called मिश्रिकरण.

Reason - As in the previous aphorism, to identify these type of compounds a तृतीयातत्पुरुष, one needs a lexicon of भक्ष्य and means of मिश्रिकरण for the corresponding भक्ष्यs. In the absence of such lexicon, it is not possible to identify such compounds.

8. सूत्र - "कृत्यैर्ऋणे" (2-1-43)

Meaning - (a case-inflected word ending in the seventh case-termination is compounded) with (a case-inflected word) ending in a कृत्य-suffix, when (the compounded conveys) the sense of debt (the compounded being called तत्पुरुष).

Example - मासदेयम् "a debt repayable within a month".

The word ऋण indicates by implication any appointed time in general, and not merely a time for the payment of a debt. Therefore we get compounds in the following case also - पूर्वाह्णगेयं साम "the साम that would be sung in the morning".

Reason - Since this involves semantic interpretation, it is not easy to detect the type of such compounds automatically.

9. सूत्र - "संज्ञायाम्" (2-1-44)

Meaning - (a case-inflected word ending in the seventh case-

termination is compounded with a case-inflected word) provided that (the compound formed) is a name (the compound being called तत्पुरुष).

Description - A संज्ञा is expressed by the complete word, hence it is an invariable (nitya) compound ; for we cannot express an appellative by a sentence. For instance अरण्येतिलकाः "wild sesamum" yielding no oil ; anything which does not answer to one's expectation.

Reason - This aphorism also needs the knowledge of the real world to interpret the relation between the words involved and thus is out of scope of mechanical identification.

10. सूत्र - "क्षेपे" (2-1-47)

Meaning - (a case-inflected word ending in the seventh case-termination is compounded with a case-inflected word ending in the suffix क्त) when (the compound conveys) the sense of abuse (the compound being called तत्पुरुष).

Example - अवतप्तेनकुलस्थितं त एतत् "thy this work is as if an ichneumon (नकुल) standing on hot ground" etc.

Reason - Since this involves semantic interpretation, it is not easy to detect the type of such compounds automatically.

11. सूत्र - "दिक्संख्ये संज्ञायाम्" (2-1-50)

Meaning - (a case-inflected word signifying) a direction or a number (is compounded with a syntactically agreeing case-inflected word) provided that (the compound formed is) a name (the compound

being called तत्पुरुष).

Examples - पूर्वेषुकामशमी "the town of इषुकामशमी in the East", सप्तर्षयः "seven sages" (the constellation of Great Bear) etc.

Reason - Since this involves semantic interpretation, it is not easy to detect the type of such compounds automatically.

12. सूत्र - "तद्धितार्थोत्तरपदसमाहारे च" (2-1-51)

Meaning - also (a case-inflected word signifying a direction or a number is compounded with a syntactically agreeing case-inflected word) provided that (the form derived) conveys the meaning of a तद्धित(-suffix), or that a final compound-constituent follows, or that (the form derived) conveys the sense of समाहार "group" (the compound being called तत्पुरुष).

Examples - पौर्वशालः, आपरशालः, अपरशालप्रियः etc.

The current aphorism explains a तत्पुरुष compound. To implement this aphorism, it needs the whole तद्धित environment and also the context.

13. सूत्र - "संख्यापूर्वो द्विगुः" (2-1-52)

Meaning - (a compound conditioned by 2-1-51) whose first constituent is (a word signifying) number (is called तत्पुरुष and) द्विगु.

Description - In a case where the sense is that of a तद्धित affix or when an additional member comes after the compound (उत्तरपदे) or when an aggregate (समाहारे) is to be expressed, the compound, the first member of which is numeral, is called द्विगु.

Example - पञ्चगवम् "an aggregate of five cows".

Reason - The anuvritti of तद्धित and समाहारे is coming to this aphorism. To implement it, we need तद्धित environment and the context of aggregate. It is not possible to decide type mechanically.

14. सूत्र - "कुत्सितानि कुत्सनैः" (2-1-53)

Meaning - (case-inflected words signifying) objects of contempt (are compounded) with (syntactically agreeing case-inflected words signifying) abuse (the compound being called तत्पुरुष).

Examples - वैयाकरणखसूचिः 'A bad grammarian, who does not know grammar, but contemplates the heaven (ख) when asked any questions, मीमांसकदुर्दुरूढः 'an evil conducted मीमांसक'.

Reason - In a compound, if the first component contains an adjective and the second component contains also an adjective in sense of contempt then the compound type can be identified as a उभयपदविशेषण-कर्मधारय compound. Since it not easy to decide whether the second component such as खसूचिः is used in the sense of abuse or not, it is not possible to mark such compound types automatically.

15. सूत्र - "उपमानानि सामान्यवचनैः" (2-1-55)

Meaning - (case-inflected words signifying) the standard of comparison (are compounded) with (syntactically agreeing, case-inflected words) signifying a common property (the compound being called तत्पुरुष).

Description - That by or to which a thing is compared is 'उपमान' and

the उपमेय (the thing compared) is called सामान्य or common. Thus घनश्यामः कृष्णः 'cloud-black Krishna' (Krishna black as a cloud). Here श्याम is a quality common to Krishna and cloud, therefore 'cloud' which is the उपमान is compounded with it. So also कुमुदशयेनी 'lily-white', हंसगद्गदा 'Swan sounding', न्यग्रोध-पारिमण्डला 'globular as न्यग्रोध tree'.

In analysing the above compound घनश्यामः we must use the word इव, as घन इव श्यामः, and this shows that it is merely a simile or metaphorical use of the word. This aphorism is made in order to declare a niyama rule, so that the उपमान word should stand first.

Reason - To implement this aphorism, one needs a list of उपमानवाची words and the corresponding उपमेयवाची words.

16. सूत्र - "कृत्यतुल्याख्या अजात्या" (2-1-68)

Meaning - (case-inflected words ending in a कृत्य(-suffix) and synonyms of तुल्य "equal" (are preferably compounded) with (a syntactically agreeing case-inflected word signifying) something other than a जाति "class/generic notion" (being called तत्पुरुष).

Examples - भोज्योष्णम् "hot food", तुल्यश्वेतः "equally white", सदृशश्वेतः "equally white" etc.

Reason - To implement this aphorism, the second component should not refer to a class. This is possible only if the lexicon is rich with this kind of information.

17. सूत्र - "पूर्वापराधरोत्तरमेकदेशिनैकाधिकरणे" (2-2-1)

Meaning - (the case-inflected words) पूर्व "fore (part)", अपर "hind

(part)", अधर "lower (part)" and उत्तर "upper (part)" are preferably compounded with (a case inflected word) signifying (a whole) having parts, provided that (the whole mentioned) is one single item, (being called तत्पुरुष).

Examples - पूर्वकायः "the front of the body", अपरकायः "the black of the body".

Reason - The clue viz. second word indicating it refers to a body with parts is semantic. So unless a lexicon rich with such an information is available, such compound types can't be identified mechanically.

18. सूत्र - "नित्यं क्रीडाजीविकयोः" (2-2-17)

Meaning - (a case-inflected word ending in the sixth case-termination, when the genitive case ending conveys the sense of कर्मन् "direct object" is) always compounded with a case-inflected word ending in the suffixes तृच् or अक्) when (the compound) conveys the sense of क्रीडा "a game" or जीविका "a profession" (the compound being called तत्पुरुष).

Examples - उद्दालकपुष्पभञ्जिका "a sort of game played by the people in the eastern districts in which उद्दालक flowers are broken or crushed", दन्तलेखकः "one who earns his bread by painting or marking the teeth", नखलेखकः "a nail-painter by profession" etc.

Reason - The current aphorism explains a तत्पुरुष compound. To implement this rule one needs a list of words indicating either a क्रीडा or जीविका.

19. सूत्र - "उपपदमतिङ्" (2-2-19)

Meaning - an उपपद which (does) not (end in a suffix called) तिङ् (is always compounded the compound being called तत्पुरुष).

Examples - कुम्भकारः "one who makes pots", नगरकारः "one who makes the city" etc.

Reason - If the second component is a kind of कृदन्त word which is found only in compounds such as कार, ज्ञः, गः etc. then the compound is likely to be उपपद-तत्पुरुष compound. For this we need a special morph module that handles such forms.

20. सूत्र - "अनेकमन्यपदार्थे" (2-2-24)

Meaning - two or more (case-inflected words are invariable compounded) when (the compound) conveys the sense of a word other (than the compound-constituents, the compound being called बहुव्रीहि).

Examples - प्राप्तमुदको ग्रामः "a water reached village", वीरपुरुषकः ग्रामः "a village possessed of heroic men" etc.

Reason - There is no clue whatsoever except the fact that this word refers to something which is different from the referents of the components involved. Thus it needs extra Linguistic knowledge to identify the type of such compounds.

21. सूत्र - "चार्थे द्वन्द्वः" (2-2-29)

Meaning - (two or more case-inflected words are invariably

compounded) when (the compound) conveys the sense of च "and" (the compound being called) द्वन्द्व.

Examples - रामलक्ष्मणौ "Rama and Lakshman", धवखदिरौ "the Mimosea and the Grislea" etc.

Reason - Just by looking at the components, the type-identification of this kind of compounds, is very difficult because the decision involves the meaning of components.

Finally we discuss the rules which deal with special or exceptional cases of compounds, which can not be analysed further. We just list them as exceptional compounds. Here, we are giving the list of exceptional or special rules.

1. सूत्र - "अन्यपदार्थे च संज्ञायाम्" (2-1-21)

Meaning - Also, (a case-inflected word is compounded with a case-inflected word signifying a river) when the sense of अन्यपदार्थ (an additional meaning (not expressed by the compound constituents) is to be conveyed, provided the (the compound formed) is a name (the compound being called अव्ययीभाव).

Description - This is a विधिसूत्र prescribing compound formation. A word ending in a case affix is compounded with words denoting the names of rivers, when the compound word denotes a thing other than that expressed by the terms of the compound, and is an appellative ; the compound so formed being an अव्ययीभाव.

The anuvritti of the word संख्या does not extend to this aphorism. Though this rule is given in the subdivision relating to optional compounds, it is, however, a nitya-samāsa rule : for no name (संज्ञा)

can ever be expressed by a sentence, and that being so, these compounds can never be analysed. As उन्मत्तगङ्गम् 'the country called उन्मत्तगङ्गम्.'

Since such compounds can never be analysed, they are handled as special compounds, and are termed as नद्युत्तरपद-अन्यपदार्थे-अव्ययीभाव.

2. सूत्र - "कुमारः श्रमणादिभिः" (2-1-70)

Meaning - (the case-inflected word) कुमार "boy" (is preferably compounded) with (the syntactically agreeing, case-inflected words) श्रमणा "Buddhist/Jaina nun", etc. (being called तत्पुरुष).

Example - कुमारीश्रमणा or कुमारश्रमणा "a virgin ascetic".

3. सूत्र - "प्राप्तापन्ने च द्वितीयया" (2-2-4)

Meaning - also (the case-inflected words) प्राप्त "who/which has reached/obtained" and आपन्न "who/which has obtained" (are preferably compounded with a case-inflected word ending in the second case termination (being called तत्पुरुष).

Examples - प्राप्तजीविकः (प्राप्तो जीविकां) as well as जीविकाप्राप्तः 'obtained his livelihood', आपन्नजीविकः and जीविकापन्नः.

Reason - This is a special compound with only two possible words as second component. These are just treated as special cases and not analysed further.

4. सूत्र - "कालाः परिमाणिना" (2-2-5)

Meaning - (case-inflected words signifying a period of) time (are

preferably compounded) with (a case-inflected word signifying) an item to be measured (being called तत्पुरुष).

Examples - मासो जातस्य = मासजातः "a month old", संवत्सरजातः "a year old" etc.

The current aphorism explains a तत्पुरुष compound and it needs the information (of परिमाण) the words denoting the object whose duration is measured by the time. We treat it as exceptional rule.

5. सूत्र - "अमैवाव्ययेन" (2-2-20)

Meaning - (a case-inflected word called उपपद is always compounded) with (a case-inflected word called) अव्यय "indeclinable" (ending in the suffix) अम् "only" (the compound being called तत्पुरुष).

Examples - स्वादुङ्कारं भुङ्क्ते "he eats having made his food sweet", लवणंकारम् "having seasoned" etc.

Reason - If the component is a word ending in second case suffix and the second component is a कृदन्त used in उपपद-तत्पुरुष compounds, then the compound may be classified under उपपद-तत्पुरुष.

6. सूत्र - "तृतीयाप्रभृतीन्यन्यतरस्याम्" (2-2-21)

Meaning - (the case-inflected words called उपपद) starting from (those prescribed as ending in) the third case termination (are) indifferently (compounded with a case-inflected word called अव्यय "indeclinable", ending in the suffix अम् only (being called तत्पुरुष).

Example - मूलकोपदंशं भुङ्क्ते or मूलकेनोपदंशं भुङ्क्ते "he eats after having relished the food with radish".

Type of compound -

Reason - The current aphorism explains a उपपद-तत्पुरुष compound. These type of compounds can be analysed provided कृदन्त morphological analyser analyses special forms. In the absence of such morphological analyser, we treat all such compounds as special cases.

7. सूत्र - "तत्र तेनेदमिति सरूपे" (2-2-27)

Meaning - two (case-inflected words) alike in form (and represented by the word-pattern) तत्र (or) तेन (are compounded with each other, the compound) conveying the meaning इदम् "this" (and being called बहुव्रीहि).

Examples - केशाकेशि "hair to hair, fighting by pulling each other's hair", दण्डादण्डी "stick against stick, fight with and stave" etc.

Reason - The current aphorism explains a प्रहरण-विषयक-बहुव्रीहि compound. These type of compounds can be analysed provided the morphological analyser analyses special forms such as केशा and केशि etc. In the absence of such morphological analyser, we treat all such compounds as special cases.

List of exceptional or special rules				
No.	Sūtra reff.	Sūtra	Example	Type
1	2-1-21	अन्यपदार्थे च संज्ञायाम्	उन्मत्तगङ्गम्	A5

2	2-1-70	कुमारः श्रमणादिभिः	कूमारीश्रमणा, कु- मारश्रमणा etc.	K
3	2-2-4	प्राप्तापन्ने च द्वितीयया	प्राप्तजीविकाः etc.	K
4	2-2-5	कालाः परिमाणिना	मासजातः, संवत्सरजातः etc.	K
5	2-2-20	अमैवाव्ययेन	स्वादुङ्कारं भुङ्के ।	-
6	2-2-21	तृतीयाप्रभृतीन्यन्यतरस्याम्	मुलकोपदंशं भुङ्के ।	-
7	2-2-27	तत्र तेनेदमिति सरूपे	केशाकेशि, दण्डा- दण्डी	Bsp

6.3 Statistical Approach

In this section we describe the statistical approach. We use manually tagged corpus as a model for predicting the tags, given a pair of constituents of a compound. Manually tagged corpus has been generated by the SHMT consortium funded by DIT. The total corpus is around 600K words of which 80,155K words are compounds. These texts were tagged using the tagset given in appendix-B. All these compounds are thus tagged 'in context' and contain only one tag. This corpus formed the training data. Another small corpus with 400 tagged compounds from totally different texts, was kept aside for testing.

We observe that certain words have affinity for certain other words. For example घन has an affinity towards a word indicating the colour such as

श्याम or name of the colour has an affinity towards an object in the real world such as flower as in नीलोत्पलम् etc. Further from the manually tagged data we also know the likely tag for such compounds. The manually tagged data also reveals that certain words when used as the first component of a compound indicate a particular compound. For example the indeclinables such as यथा as the first component always indicates an अव्ययीभाव, which is also supported by the Pāṇini's rule (अव्ययं विभक्तिसमीपसमृद्धिवृद्धार्थाभावात्पया-सम्प्रति... "2-1-6").

6.3.1 Some features of the manually tagged data

1. Around 25% of the compounds were repeated.
2. The 38,594 tokens of compounds contain 14,203 types of left word and 24,391 types of right word.
3. The frequency distribution of highly frequent tags is shown in Table 6.10. To study the effect of fine-grained-ness we also merged the sub-types. Table 6.12 gives the frequency distribution of major tags, after merging the sub-types.

Coarse and fine grain tagset

Pāṇini has classified the compounds in four major classes however, for the purpose of generating a paraphrase, these four class are not sufficient. Hence fine grain tagset of 55 tags was evolved by the Sanskrit consortium. The performance of any classifier is inversely proportional to the number of tags. More the tags, less is the accuracy of a word getting correct tag. This is true with the compound type identification as well. Especially because of the associated words and their frequencies. There are more chances of the

data becoming sparse. Hence we decided to measure the performance of our type identifier using both the fine grained as well as coarse grained tagset. For the purpose of identification, we define our coarse grained tagset to be consisting of eight tags viz. तत्पुरुष, कर्मधारय, बहुव्रीहि, द्वन्द्व, उपपद, अव्ययीभाव.

Tag	% of words
T6	29.12
K1	12.05
Bs6	6.97
T3	5.64
Tn	4.50
Di	3.75
U	3.64

Table 6.4: Distribution of Fine-grain-
Tags

Tag	% of words
T	50.51
K	19.01
B	12.53
D	4.70
U	3.66
A	0.94
S	0.78

Table 6.6: Distribution of Coarse-
Tags

From the manually tagged data, we define the following probabilities

(a) $P(T/w_1-w_2)$ = probability that the compound w_1-w_2 has a tag T.

$$= \frac{\#(w_1 - w_2 \text{ of type } T)}{\#(w_1 - w_2)}$$

(b) $P(T/w_1-)$ = probability that a compound with w_1 as the initial component has tag T.

$$= \frac{\#of \text{ words with } w_1 \text{ as initial component with tag } T}{\#of \text{ words with } w_1 \text{ as an initial component}}$$

(c) $P(T/-w_2)$ = probability that a compound with w_2 as the final

component has tag T.

$$= \frac{\#(\text{of words with } w_2 \text{ as initial component with tag } T)}{\#(\text{of words with } w_2 \text{ as an initial component})}$$

6.3.2 Algorithm

Given w_1-w_2

- (i) If $P(T/w_1-w_2)$ exists, we choose the max $P(T_i/w_1-w_2)$
- (ii) else if $P(T_i/w_1-)$ and $P(T_i/-w_2)$ exist, we choose maximum $P(T_i/w_1-) \times P(T_i/-w_2)$
- (iii) else if only $P(T_i/w_1-)$ exists, we choose maximum $P(T_i/w_1-)$
- (iv) if only $P(T_i/-w_1)$ exists, we choose maximum $P(T_i/-w_1)$
- (v) and finally if both w_1- and $-w_2$ do not occur in the manually tagged corpus, we assign a tag with maximum probability.

6.3.3 Performance Evaluation

The test data of 400 words are tagged 'in context'. While our compound tagger does not see the context, and thus suggests more than one possible tag, and ranks them. Normally a tool is evaluated for its coverage and precision. In our case, the tool always produces tags with weights associated with them. Hence, the coverage is 100%. The precision is evaluated based on the ranks of the correct tags. Table 6.14 shows results of 400 words with a coarse as well as fine grained tagset.

Thus, if we consider only the first rank, the precision with 8 tags is 76.1% and with 55 tags, it is 62.0%. The precision increases substantially to 95.8% and 83.5% respectively if we take first three ranks into account.

Rank	with 55 tags		with 8 tags	
	no of words	% of words	no of words	% of words
1	250	62.0	307	76.1
2	59	14.6	56	13.8
3	28	6.9	24	5.9

Table 6.8 : Precision of Type Identifier.

Effect of the size of manually tagged corpus

It is known that the Natural Language corpus suffers from sparseness. There are also various studies to predict the 'ideal' size of the corpus for a specific task. We state here our experience with the manually tagged compound for type identification.

With 150K corpus, the results were as follows :-

- (i) Around 10% of the compounds were repeated.
- (ii) The 12,796 tokens of compounds contain 2,630 types of left word and 7,106 types of right word.
- (iii) The frequency distribution of highly frequent tags is shown in Table 6.10. To study the effect of fine-grained-ness we also merged the sub-types. Table 6.12 gives the frequency distribution of major tags, after merging the sub-types.

Tag	% of words
T6	28.35
Bs6	12.45
K1	9.63
Tn	8.56
Di	7.23
U	5.73

Table 6.10: Distribution of Fine-grain-Tags

Tag	% of words
T	52.43
B	18.96
K	12.04
D	8.84
U	5.73

Table 6.12: Distribution of Coarse-Tags

- (iv) Finally the performance of type identifier on this two data was

Thus, if we consider only the 1st rank, the precision with 8 tags is 72.7% and with 55 tags, it is 63.0%. The precision increases substantially to 95.4% and 81.1% respectively if we take 1st three

Rank	with 55 tags		with 8 tags	
	no of words	% of words	no of words	% of words
1	252	63.0	291	72.7
2	44	10.9	53	13.2
3	29	7.2	38	9.5

Table 6.14 : Precision of Type Identifier.

ranks into account.

What can be concluded is, to improve the performance of type identifier, one may need to explore other parameters, other than the left or right word of a compound. For example, it might be useful to check whether the component is in neuter gender or not, whether it is a non-finite (कृदन्त) form ending in क्त or not, and so on. Such features may further improve the performance.