

Hindi-Urdu Phrase Structure Annotation Guidelines

Version 0.2 (Reorganized for Joint PS-DS Guidelines)

Rajesh Bhatt, with Annahita Farudi and Owen Rambow

`bhatt@linguist.umass.edu`

November 10, 2013

Contents

| | | |
|-----------|---|-----------|
| I | General Introduction | 11 |
| 1 | General Introduction | 13 |
| II | Orthography, Morphology, and Parts of Speech | 15 |
| 2 | Tokenization | 17 |
| 3 | Parts of Speech | 19 |
| 3.1 | Introduction | 19 |
| 3.2 | Objective | 20 |
| 3.3 | Some Assumptions | 20 |
| 3.4 | Issues in Tag Set Design | 20 |
| 3.4.1 | Fineness vs Coarseness | 21 |
| 3.4.2 | Syntactic Function vs Lexical Category | 22 |
| 3.4.3 | New Tags vs Tags from a Standard Tagger | 22 |
| 3.5 | POS tags Chosen for the Current Scheme | 23 |
| 3.5.1 | Nouns, Nominals, Postpositions, and Pronouns | 23 |
| 3.5.2 | Verbs | 27 |
| 3.5.3 | JJ: Adjective | 30 |
| 3.5.4 | RB: Adverb | 30 |
| 3.5.5 | RP: Particle | 31 |
| 3.5.6 | CC: Conjuncts (co-ordinating and subordinating) | 31 |
| 3.5.7 | QF: Quantifiers | 32 |
| 3.5.8 | QC: Cardinals | 32 |
| 3.5.9 | QO: Ordinals | 32 |
| 3.5.10 | INTF: Intensifier | 32 |

| | | |
|--|---|-----------|
| 3.5.11 | INJ: Interjection | 33 |
| 3.5.12 | NEG: Negative | 33 |
| 3.5.13 | UT: Quotative | 33 |
| 3.5.14 | SYM: Special Symbol | 33 |
| 3.5.15 | XC: Compounds | 33 |
| 3.5.16 | RDP: Reduplication | 34 |
| 3.5.17 | ECH: Echo words | 35 |
| 3.5.18 | UNK: Unknown | 35 |
| 3.6 | Some Special Cases | 36 |
| 3.6.1 | <i>vAlA</i> -type constructions | 36 |
| 3.6.2 | Honorifics in Indian languages | 37 |
| 3.6.3 | Foreign words | 37 |
| 3.7 | A Special Note | 37 |
| 3.8 | Chunk Tags Chosen for the Current Scheme | 38 |
| 3.8.1 | Definition of a chunk | 39 |
| 3.8.2 | Some Special Cases | 45 |
| 3.9 | Annotation Procedure | 47 |
| 3.10 | Conclusion | 47 |
| 3.11 | References | 47 |
| 3.12 | Acknowledgements | 48 |
| 3.13 | Table of POS Tags | 48 |
| III Introduction: Syntax | | 53 |
| 4 Introduction to Syntactic Annotation | | 55 |
| 4.1 | Background | 55 |
| 5 Introduction: Phrase Structure Annotation | | 57 |
| 5.1 | General Background | 57 |
| 5.2 | Representational Issues | 58 |
| 5.3 | The Representation of Arguments | 58 |
| 5.3.1 | Background: The Representation of Arguments | 60 |
| 5.4 | Summary of Phrase Markers Used in PS | 63 |

| | |
|---|-----------|
| <i>CONTENTS</i> | 5 |
| 6 Comparison Between the Levels of Representation | 65 |
| IV Verbs and Their Arguments | 67 |
| 7 Transitive Verbs | 69 |
| 7.1 The First Argument | 69 |
| 7.1.1 Simple Transitives | 69 |
| 7.2 The Second Position | 70 |
| 7.2.1 Transitives | 71 |
| 7.3 Background | 72 |
| 7.3.1 Background: the First Argument | 72 |
| 7.3.2 Background: the Lower Argument | 75 |
| 8 Intransitives | 77 |
| 8.1 Unergatives | 77 |
| 8.2 Unaccusatives | 78 |
| 8.3 Background | 80 |
| 9 Dative Constructions | 81 |
| 9.1 General Observations on Datives | 81 |
| 9.2 Ditransitives | 81 |
| 9.3 Dative Subject Constructions | 83 |
| 10 Clausal Complementation | 85 |
| 10.1 Non-Finite Complements | 85 |
| 10.2 Finite Complements and Extraposition | 87 |
| 10.3 Missing Subject with Predicative Adjective and <i>ki</i> Complement Clause | 91 |
| 10.4 Result Clauses | 92 |
| 10.5 Background: Clausal Complementation | 94 |
| 10.5.1 Finite Complements and Extraposition | 94 |
| 11 Existentials, Possessives, and Locatives | 95 |
| 11.1 Summary | 95 |
| 11.2 Existentials | 96 |
| 11.3 Locative Possessives | 98 |

| | |
|--|------------|
| 11.4 Non-Locative Possession | 99 |
| 11.4.1 Inalienable Possession | 99 |
| 11.4.2 Experiencer Possession | 103 |
| 11.5 Background: The Definiteness Effect and Exceptions | 104 |
| 12 Non-Verbal Predicates | 107 |
| 12.1 Predicative Use of Nouns, Adjectives, and Postpositions | 107 |
| 12.2 Adjectival Predicates with Eventive Meaning | 109 |
| 12.3 Non-Verbal Nominal Predicates with <i>ban/banaa</i> ‘become/make’ | 111 |
| 13 Adjuncts: Instrument, Source, Location | 115 |
| 13.1 Adverbs | 116 |
| 14 Clausal Adjuncts | 117 |
| 14.1 General | 117 |
| 14.2 Participial Adjunct Clauses | 118 |
| 14.2.1 Gerunds | 122 |
| 15 Negation | 125 |
| 15.1 Sentential Negation | 125 |
| 15.2 Constituent Negation | 127 |
| 16 Auxiliaries | 129 |
| V Types of Clauses and Word Order in Clauses | 131 |
| 17 Passive Voice | 133 |
| 18 Causativization | 135 |
| 18.1 Initial Non-Binary Branching Analysis | 136 |
| 18.2 Revised Binary Branching Analysis | 137 |
| 18.3 Transitivity Alternations that we do Not Handle | 147 |
| 19 Clausal Modifiers of Nouns | 149 |
| 19.1 English Style Headed Relative Clauses | 150 |
| 19.2 Correlatives | 152 |

| | | |
|-----------|--|------------|
| 19.3 | Participial Relative Clauses | 155 |
| 19.4 | <i>vaalaa</i> Modifiers of Nouns | 157 |
| 20 | Light Verb Constructions | 161 |
| 20.1 | Linguistic Description | 161 |
| 20.2 | A Typology of Light Verb Constructions | 162 |
| 20.2.1 | No External Argument, Internal Argument has Nominal Case | 162 |
| 20.2.2 | No External Argument, Internal Argument has Clausal Case | 163 |
| 20.2.3 | External Argument, Internal Argument has Nominal Case | 164 |
| 20.2.4 | External Argument, Internal Argument has Clausal Case | 165 |
| 20.3 | Dative Arguments associated with Event Nominals | 166 |
| 20.4 | Light Verb Constructions with Foreign Eventive Expressions | 169 |
| 20.5 | Placement of Negation | 170 |
| 20.6 | Pronominal Predicates with Extraposed Clause | 172 |
| 20.7 | Non-Finite Complement Clauses | 173 |
| 20.8 | Small Clauses | 174 |
| 20.9 | Finite Complement Clauses | 175 |
| 20.10 | Result Clauses | 176 |
| 20.11 | Conditionals | 177 |
| 21 | Local Scrambling | 179 |
| 22 | Cross-clausal Scrambling | 183 |
| 23 | 'Rightward' Scrambling | 189 |
| VI | Noun Phrases and Other Phrases | 191 |
| 24 | Postpositional Phrases | 193 |
| 24.1 | Linguistic Description | 193 |
| 24.1.1 | True Postpositions (P) | 193 |
| 24.1.2 | Spatio-Temporal Nouns (NST) | 195 |
| 24.2 | PS Analysis | 196 |
| 24.2.1 | True Postpositions (P) | 196 |
| 24.2.2 | Spatio-Temporal Nouns (NST) | 198 |

| | |
|---|------------|
| 25 Nominal Phrases | 199 |
| 25.1 Structure of Noun Phrases | 199 |
| 25.2 Empty NPs | 203 |
| 25.3 Noun Compounds | 203 |
| 25.3.1 Additional considerations and alternative analyses | 206 |
| 25.4 Nominal Appositives | 207 |
| 26 Adjectival Phrases and Comparatives | 209 |
| | |
| VII Ellipsis and Coordination | 237 |
| | |
| 27 Coordination | 239 |
| | |
| 28 Shared arguments | 243 |
| 28.1 Description and data | 243 |
| 28.2 Analysis | 244 |
| | |
| 29 Ellipsis | 249 |
| 29.1 NP-Ellipsis | 249 |
| 29.2 Gapping | 250 |
| 29.3 Gapping | 250 |
| 29.4 Right Node Raising | 260 |
| 29.5 Sluicing | 264 |
| | |
| VIII Miscellaneous Syntax | 265 |
| | |
| 30 Questions | 267 |
| | |
| 31 Imperatives | 269 |
| | |
| 32 Vocatives | 271 |
| | |
| 33 Parentheticals | 273 |
| | |
| 34 Fragments | 275 |
| | |
| 35 Idiomatic Constructions | 277 |

| | |
|---|------------|
| <i>CONTENTS</i> | 9 |
| 35.1 Introduction | 277 |
| 35.2 <i>X-ke calte hue</i> Constructions | 277 |
| 35.3 <i>X-ke rang-me bhang paR gayaa</i> Construction | 279 |
| 36 Names | 281 |
| 37 Foreign Words | 285 |
| IX Non-Syntactic Phenomeona | 287 |
| 38 Punctuation | 289 |
| 39 Other Formatting in Written Language | 291 |
| 39.1 Speaker Marking in Dialog | 291 |
| 40 Dates and Expressions with Numbers | 293 |
| X Appendix | 295 |
| A Confusion Areas | 297 |
| B Information about the Treebank | 299 |
| C Conversion | 301 |
| D Summary of Empty Categories | 303 |
| D.1 The Role of Empty Categories | 304 |
| D.1.1 Types of Empty Categories | 304 |
| D.1.2 Linguistic Motivation for Empty Categories | 304 |
| D.1.3 Empirical Motivation for Empty Categories | 305 |

Part I

General Introduction

Chapter 1

General Introduction

This is a general introduction to the whole combined guidelines. To be jointly written at the end.

Part II

Orthography, Morphology, and Parts of Speech

Chapter 2

Tokenization

This text is identical for all levels. Person in charge: tbd.

Chapter 3

Parts of Speech

This analysis is identical for all levels of syntactic annotation, except that the discussion of chunking (Section 3.8) only applies to DS.

3.1 Introduction

The significance of large annotated corpora in the present day NLP is widely known. Annotated corpora serve as an important tool for investigators of natural language processing, speech recognition and other related areas. It proves to be a basic building block for constructing statistical models for automatic processing of natural languages.

Many such corpora are available for languages across the world and have proved to be a useful step towards natural language processing. Coming to the scenario for Indian languages, not much work has been carried out on the front of automatic processing of Hindi or any other Indian language. The main bottleneck being unavailability of an annotated corpora, large enough to experiment statistical algorithms.

Annotation of corpora (AnnCorra) can be done at various levels viz, part of speech, phrase/clause level, dependency level, etc. Part of speech tagging forms the basic step towards building an annotated corpus. Chunking can form the next level of tagging.

The task of annotating corpora of several Indian languages has been taken up in the Indian Language Machine Translation (ILMT) project. ILMT is a project in which a number of institutes have come together to form a consortium and work towards developing MT systems for various Indian language pairs.

A primary requirement of such an effort is defining standards for various sub tasks. Thus, standardization of annotation schemes for various annotation tasks becomes a crucial step in this direction.

The issues related to defining standards for POS/Chunk tagging schemes were discussed by scholars from various Indian institutes by way of holding meetings etc. and some standards have been arrived at.

3.2 Objective

The purpose of the meetings was to arrive at standard tagging scheme for POS tagging and chunking for annotating Indian languages (AnnCorra) and come up with the tags which are exhaustive for the task of annotation for a larger group of languages, specially, Indian languages. The present document gives a detailed description of the tags which have been defined for the tagging schemes and elaborates the motivations behind the selection of these tags. The document also discusses various issues that were addressed while preparing the tag sets and how they have been resolved.

3.3 Some Assumptions

During the workshop it was decided to base the discussion and decisions about various tags on the following basic assumptions which everybody agreed on :

1. The tags should be common for all Indian languages.
2. It should be comprehensive/ complete.
3. It should be simple. Maintaining simplicity is important for the following two reasons :
 - (a) Ease of Learning
 - (b) Consistency in annotation

Another important point which was discussed and agreed upon was that POS tagging is NOT a replacement for morph analyzer. A 'word' in a text carries the following linguistic knowledge a) grammatical category and b) grammatical features such as gender, number, person etc. The POS tag should be based on the 'category' of the word and the features can be acquired from the morph analyzer.

3.4 Issues in Tag Set Design

This section deals with some of the issues related to any POS tagger and the policy that we have adopted to deal with each of these issues for our purpose.

The first step towards developing POS annotated corpus is to come up with an appropriate tags. The major issues that need to be resolved at this stage are:

1. Fineness vs Coarseness in linguistic analysis
2. Syntactic Function vs lexical category
3. New tags vs tags close to existing English tags

We discuss them in turn.

3.4.1 Fineness vs Coarseness

An issue which always comes up while deciding tags for the annotation task is whether the tags should capture 'fine grained' linguistic knowledge or keep it 'coarse'. In other words, a decision has to be taken whether or not the tags will account for finer distinctions of the parts of speech features. For example, it has to be decided if plurality, gender and other such information will be marked distinctly or only the lexical category of a given word should be marked.

It was decided to come up with a set of tags which avoids 'finer' distinctions. The motivation behind this is to have less number of tags since less number of tags lead to efficient machine learning. Further, accuracy of manual tagging is higher when the number of tags is less.

However, an issue of general concern is that in an effort to reduce the number of tags we should not miss out on crucial information related to grammatical and other relevant linguistic knowledge which is encoded in a word, particularly in agglutinating languages, eg, Tamil, Telugu and many other Indian languages. If tags are too coarse, some crucial information for further processing might be missed out. As mentioned above, primarily the required knowledge for a given lexical item is its grammatical category, the features specifying its grammatical information and any other information suffixed into it. For example,

Telugu word ' rAmudA (Is it Ram ?)' contains the following information ;category (noun)+grammatical features(masculine, singular) + question;. The word by itself is a bundle of linguistic information. Morph analyser provides all the knowledge that is contained in a word. It was decided that any linguistic knowledge that can be acquired from any other source (such as morph analyser) need not be incorporated in the POS. As mentioned above, POS tagger is not a replacement for morph analyser. In fact, features from morph analyser can be used for enhancing the performance of a POS tagger. The additional knowledge of a POS given by a POS tagger can be used to disambiguate the multiple answers provided by a morph analyser.

On the other hand, we agree that too coarse an analysis is not of much use. Essentially, we need to strike a balance between fineness and coarseness. The analysis should not be so fine as to hamper machine learning and also should not be so coarse as to miss out important information. It is also felt that fine distinctions are not relevant for many of the applications(like sentence level parsing, dependency marking, etc.) for which the tagger may be used in future.

However, it is well understood that plurality and other such information is crucial if the POS tagged corpora is used for any application which needs the agreement information. In case such information is needed at a later stage, the same tag set can be extended to encompass information such as plurality etc as well. This can be done by providing certain heuristics or linguistic rules.

Thus, to begin with, it has been decided to adopt a coarse part of speech analysis. At the same time, wherever it is found essential, finer analysis is incorporated. Also, there is a basic understanding that wherever/whenever essential, the tags containing finer linguistic knowledge can be incorporated. An example of where finer analysis becomes crucial has been given below. Take the Hindi sentence (h1) below :

h1. AsamAna_N meM_PSP uDane_VM vAlA_PSP ghoDA_NN 'sky' 'in' 'flying' 'horse nIce_NST utara_VM AyA_VAUX. 'down' descend came

In (h1) above uDane is a noun derived from a verb. The word AsamAna is an argument of uDane and not of 'nIce utara AyA another verb in the sentence. It is crucial to retain the information that uDane, though functioning as a noun now, is derived from a verb and can take its own arguments. In order to preserve such crucial information a finer analysis is essential. Therefore, a distinct tag needs to be introduced for such expressions. In the current tagging scheme uDane will be annotated as a 'main verb (VM)' at the POS level. However, the information that it is functioning like a noun will be captured at the chunking level by introducing a distinct chunk tag VGNN (discussed in details under Section III on Chunking).

3.4.2 Syntactic Function vs Lexical Category

A word belonging to a particular lexical category may function differently in a given context. For example, the lexical category of harijana in Hindi is a noun . However, functionally, harijana is used as an adjective in (h2) below,

h2. eka dina pAzca baje khabara AyI ki koI harijana 'one' 'day' 'five' 'o'clock' 'news' 'came' 'that' 'some' 'harijana' bAlaka unase milanA cAhatA hE 'young boy' 'him' 'to meet' 'wants' 'is' One day, a message came at five o'clock that some 'harijana' boy wanted to meet him.

Such cases require a decision on whether to tag a word according to its lexical category or by its syntactic category. Since the word in a context has syntactic relevance, it appears natural to tag it based on its syntactic information. However, such a decision may lead to further complications.

In AnnCorra, the syntactic function of a word is not considered for POS tagging. Since the word is always tagged according to its lexical category there is consistency in tagging. This reduces confusion involved in manual tagging. Also the machine is able to establish a word-tag relation which leads to efficient machine learning.

In short, it was decided that syntactic and semantic/pragmatic functions were not to be the basis of deciding a POS tag.

3.4.3 New Tags vs Tags from a Standard Tagger

Another point that was considered while deciding the tags was whether to come up with a totally new tag set or take any other standard tagger as a reference and make modifications in it according to the objective of the new tagger. It was felt that the later option is often better because the tag names which are assigned by an existing tagger may be familiar to the users and thus can be easier to adopt for a new language rather than a totally new one. It saves time in getting familiar to the new tags and then work on it.

The Penn tags are most commonly used tags for English. Many tag sets designed subsequently have been a variant of this tag set (eg. Lancaster tag set). So, while deciding the

tags for this tagger, the Penn tags have been used as a benchmark. Since the Penn tag set is an established tag set for English, we have used the same tags as the Penn tags for common lexical types. However, new tags have been introduced wherever Penn tags have been found inadequate for Indian language descriptions. For example, for verbs none of the Penn tags have been used. Instead AnnCorra has only two tags for annotating verbs, VM (main verb) and VAUX (auxiliary verb).

3.5 POS tags Chosen for the Current Scheme

This section gives the rationale behind each tag that has been chosen in this tag set.

3.5.1 Nouns, Nominals, Postpositions, and Pronouns

NN: Noun

The tag NN for nouns has been adopted from Penn tags (%%Ref) as such. The Penn tag set makes a distinction between noun singular (NN) and noun plural (NNS). As mentioned earlier, distinct tags based on grammatical information are avoided in IL tagging scheme. Any information that can be obtained from any other source is not incorporated in the POS tag. Plurality, for example, can be obtained from a morph analyzer. Moreover, as mentioned earlier, if a particular information is considered crucial at the POS tagging level itself, it can be incorporated at a later date with the help of heuristics and linguistic rules. This approach brings the number of tags down, and helps achieve simplicity, consistency, better machine learning with a small corpora etc. Therefore, the current scheme has only one tag (NN) for common nouns without getting into any distinction based on the grammatical information contained in a given noun word

Example cases :

The words which are nouns have been tagged as NN in h2 below (repeated for the sake of convenience)

h2. eka dina_N pAzcA baje khabara_N AyI ki koI harijana_N bAlaka_N unase milanA cAhatA hE

Some Special Cases Add syntactic tests for determining when a token should be considered a noun Inclusion criteria Exclusion criteria to be added with examples Certain categories which have 'pronominal ending' in Dravidian languages would be included as noun Pronominalised verbs have to be excluded Pronominalised pronouns will also have to be excluded

NST: Noun denoting spatial and temporal expressions

A tag NST has been included to cover an important phenomenon of Indian languages. Certain expressions such as 'Upaara' (above/up), 'nIce' (below) 'pahale' (before), 'Age' (front) etc are content words denoting time and space. These expressions, however, are used in

various ways. For example, these words often occur as temporal or spatial arguments of a verb in a given sentence taking the appropriate vibhakti (case marker):

h3. vaha Upara so rahA thA . 'he' 'upstairs' 'sleep' 'PROG' 'was' He was sleepign upstairs.

h4. vaha pahale se kamare meM bEThA thA . 'he' 'beforehand' 'from' 'room' 'in' 'sitting' 'was' He was sitting in the room from beforehand

h5. tuma bAhara bETho 'you' 'outside' 'sit' You sit outside.

Apart from functioning like an argument of a verb, these elements also modify another noun taking postposition 'kA'.

h6. usakA baDZA bhAI Upara ke hisse meM rahatA hE 'his' 'elder' 'brother' 'upstairs' 'of' 'portion' 'in' 'live' 'PRES' His elder brother lives in the upper portion of the house.

Apart from occurring as a nominal expression, they also occur as a part of a postposition along with 'ke'. For example,

h7. ghaDZe ke Upara thAll rakhI hE. 'pot' 'of' 'above' 'plate' 'kept' 'is' The plate is kept on the pot.

h8. tuma ghara ke bAhara bETho 'you' 'home' 'of' 'outside' 'sit' You sit outside the house.

'Upara' and 'bAhara' are parts of complex postpositions 'ke Upara' and 'ke bAhara' in (h6) and (h7) respectively which can be translated into English prepositions 'on' and 'outside'.

For tagging such words, one possible option is to tag them according to their syntactic function in the given context. For example in 5.2.2 (h7) above, the word 'Upara' is occurring as part of a postposition or a relation marker. It can, therefore, be marked as a postposition. Similarly, in 5.2.1. (h3) and (h6) above, it is a noun, therefore, mark it as a noun and so on. Alternatively, since these words are more like nouns, as is evident from 5.2.1 above they can be tagged as nouns in all there occurrences. The same would apply to 'bAhAra' (outside) in examples examples (h4), (h5) and (h8). However, if we follow any of the above approaches we miss out on the fact that this class of words is slightly different from other nouns. These are nouns which indicate 'location' or 'time'. At the same time, they also function as postpositions in certain contexts. Moreover, such words, if tagged according to their syntactic function, will hamper machine learning. Considering their special status, it was considered whether to introduce a new tag, NST, for such expressions. The following five possibilities were discussed :

1. Tag both (h5) & (h8) as NN
2. Tag both (h5) & (h8) as NST
3. Tag (h5) as NN & (h8) as NST
4. Tag (h5) as NST & (h8) as PSP
5. Tag (h5) as NN & (h8) as PSP

After considering all the above, the decision was taken in favour of (2). The decision was primarily based on the following observations:

1. 'bAhara' in both (h5) and (h8) denotes the same expression (place expression 'outside')
2. In both (h5) and (h8), 'bAhara' can take a vibhakti like a noun (bAhara ko bETho, ghara ke bAhara ko bETho)
3. If a single tag is kept for both the usages, the decision making for annotators would also be easier.

Therefore, a new tag NST is introduced for such expressions. The tag NST will be used for a finite set of such words in any language. For example, Hindi has *Age* ('front'), *pIche* ('behind'), *Upara* ('above/upstairs'), *nIce* ('below/down'), *bAda* ('after'), *pahale* ('before'), *andara* ('inside'), *bAhara* ('outside') etc.

What about expressions such as 'sambandha' (Bengali), 'badale' (Hindi), etc . They ARE NOT NST but they need another look - Do we extend the scope of NST to all such nouns with dual property ? Calling them as NST may be confusing. So???

What about 'rUpa' in 'rAma ke rUpa meM' ??? 'sahAyatA' in 'cAkU kI sahAyatA se' ??
Provide tests/examples to disambiguate from a PSP and NST for exclusion and inclusion

NNP: Proper Nouns

The need for a separate tag for proper nouns and its usability was discussed. Following points were raised against the inclusion of a separate tag for proper nouns :

a) Indian languages, unlike English, do not have any specific marker for proper nouns in orthographic conventions. English proper nouns begin with a capital letter which distinguishes them from common nouns. b) All the words which occur as proper nouns in Indian languages can also occur as common nouns denoting a lexical meaning. For example, English : John, Harry, Mary occur only as proper nouns whereas Hindi : aTala bihArI, saritA, aravinda etc are used as 'names' and they also belong to grammatical categories of words with various senses . For example given below is a list of Hindi words with their grammatical class and sense.

aTala adj immovable bihArI adj from Bihar saritA noun river aravinda noun lotus

Any of the above words can occur in texts as common lexical items or as proper names. (h9) - (h11) below show their occurrences as proper nouns,

h9. atala bihAri bAjapaI bhArata ke pradhAna mantrI the. 'Atal' 'Bihari' 'Vajpayee' 'India' 'of' 'prime' 'minister' 'was' Atal Behari Vajpayee was the Prime Minister of India.

h10. merI mitra saritA tAIvAna jA rahI hE. 'my' 'friend' 'Sarita' 'Taiwan' 'go' 'PROG' 'is' My friend Sarita is going to Taiwan

h11. aravinda ne mohana ko kitAba dI. 'Aravind' 'erg' 'Mohan' 'to' 'book' 'gave' Aravind gave the book to Mohan.

Therefore, in the Indian languages' context, annotating proper nouns with a separate tag will not be very fruitful from machine learning point of view. In fact, the identification of proper nouns can be better achieved by named entity filters.

Another point that was considered in this context was the effort involved in manual tagging of proper nouns in a given text. It is felt that not much extra effort is required in manual tagging of proper nouns. However, the data annotated with proper nouns can be useful for certain applications. Therefore, there is no harm in marking the information if it does not require much effort.

Finally, it was decided to have a separate tag for proper nouns for manual annotation and ignore it for machine learning algorithms. Following this decision, the tag NNP is included in the tag set. This tag is the same as the Penn tag for proper nouns. However, in this case also AnnCorra has only one tag for both singular and plural proper nouns unlike Penn tags where a distinction is made between proper noun singular and proper noun plural by having two tags NNP and NNPS respectively.

Needs more thinking for the second part of the above word sequences Should they be marked as a 'Proper noun compounds' as XC NNP or differently. Problem with the above is that a common noun would be tagged as an NNP.

A note on how to deal with the 'titles' should be included o

PRP: Pronoun

Penn tags make a distinction between personal pronouns and possessive pronouns. This distinction is avoided here. All pronouns are marked as PRP. In Indian languages all pronouns inflect for all cases (accusative, dative, possessive etc.). In case we have a separate tag for possessive pronouns, new tags will have to be designed for all the other cases as well. This will increase the number of tags which is unnecessary. So only one tag is used for all the pronouns. The necessity for keeping a separate tag for pronouns was also discussed, as linguistically, a pronoun is a variable and functionally it is a noun. However, it was decided that the tag for pronouns will be helpful for anaphora resolution tasks and should be retained.

A note about the ambiguity of 'wh' pronouns to be included when to mark them as 'PRP' and when to mark them as 'WQ'

Some discussion on 'conjuncts' such as a 'isaliye' etc should be included whether to mark them as 'pronouns' or CC and a test for 'decision making'

Also what should be the category for 'aba' . If pronoun then we should have some evidence to prove it and differentiate it from 'Aja, kala, parasoM' etc

DEM: Demonstratives

The tag 'DEM' has been included to mark demonstratives. The necessity of including a tag for demonstratives was felt to cover the distinction between a pronoun and a demonstrative. For example,

h12. vaha ladakA merA bhAI hE (demostrative) 'that' 'boy' 'my' 'brother' 'is' h13. vaha merA bhAI hE (pronoun) 'he' 'my' 'brother' 'is'

Many Indian languages have different words for demonstrative adjectives and pronouns. A

better evidence for including a separate tag for demonstratives is from the following Telugu examples,

t1. A abbAyi nA tammudu 'that' 'boy' 'my' 'brother' t2. atanu nA tammudu 'he' 'my' 'brother' (Telugu does not have a copula 'be' in the present tense)

It has to be marked as 'RP'

PSP: Postposition

All Indian languages have the phenomenon of postpositions. Postpositions express certain grammatical functions such as case etc. The postposition will be marked as PSP in the current tagging scheme. For example,

h22. mohana kheta meM khAda dAla rahA thA 'Mohan' 'field' 'in' 'fertilizer' 'put' 'PROG' 'was'

meM in the above example is a postposition and will be tagged as PSP. A postposition will be annotated as PSP ONLY if it is written separately. In case it is conjoined with the preceding word it will not be marked separately. For example, in Hindi pronouns the postpositions are conjoined with the pronoun,

h23. mEne usako bAzAra meM dekhA 'I' 'him' 'market' 'in' 'saw'

(h23) above has three instances of 'postposition' (in bold) usage. The postpositions 'ne' and 'ko' are conjoined with the pronouns mEM and usa respectively. The third postposition 'meM' is written separately. In the first two instances, the postposition will not be annotated. Such words will be annotated with the category of the head word. Therefore, the three instances mentioned above will be annotated as shown in (h23a) below :

h23a. mEne_{PRP} usako_{PRP} bAzAra_{NN} meM_{PSP} dekhA

3.5.2 Verbs

VM: Verb Main

Verbal constructions in languages may be composed of more than one word sequences. Typically, a verb group sequence contains a main verb and one more auxiliaries (V AUX AUX). In the current tagging scheme the support verbs (such as dAlanA in kara dAlAtA hE, uThanA in cOMka uThA thA etc) are also tagged as VAUX. The group can be finite or non-finite. The main verb need not be marked for finiteness. Normally, one of the auxiliaries carries the finiteness feature.

The necessity of marking the finiteness or non-finiteness in a verb was discussed extensively and everybody agreed that it was crucial to mark the distinction. However, languages such as Hindi, which have auxiliaries for marking tense, aspect and modalities pose a problem. The finiteness of a verbal expression is known only when we reach the last auxiliary of a verb group. Main verb of a finite verb group (leaving out the single word verbal expressions of the finite type eg vaha dilli gayA) does not contain finiteness information. For example,

h14. laDZakA seba khAtA raHA wA 'boy' 'apple' 'eating' 'PROG' 'was' The boy had kept eating.

h15. seba khAtA huA laDZakA jA rahA thA 'apple' 'eating' 'PROG' 'boy' 'go' 'PROG' 'was' The boy eating the apple was going.

The expression khAtA raHA in (h29) above is finite and khAtA huA in (h3) is non finite. However, the main verb 'khAtA' is non-finite in both the cases.

So, the issue is - whether to (1a) mark finiteness in khAtA rahA thA (had kept eating) at the lexical level on the main verb (khA) or (1b) on the auxiliary containing finiteness (wA) or (2) not mark it at the lexical level at all. All the three possibilities were discussed; 1) Mark the finiteness at the lexical level.

If we mark it at the lexical level, following possibilities are available :

1a) Mark the finiteness on the main verb, even though we know that the lexical item itself is not finite.

In this case, the annotator interprets the finiteness from the context. (The POS tags VF, VNF and VNN were earlier decided based on this approach). The main verb, therefore, is marked as finite consciously with a view that the group contains a 'verb root' and its auxiliaries (as TAM etc) is finite even though the main verb does not carry the finiteness at the lexical level. Although, this approach facilitates annotation of both the main verb and the finiteness (of the group) by a single tag, it allows tagging a lexical item (main verb) with the finiteness feature which it does not actually carry. So, this is not a neat solution.

1b) The second possibility is, mark the finiteness on the last auxiliary of the sequence. Here again the decision has to be taken from the context. This possibility was not considered since this also involves marking the verb finiteness at the lexical level.

2) Don't mark the finiteness at the lexical level. Instead mark it as indicated in (2a) or (2b) below.

2a) Introduce a new layer which groups the verb group and mark the verb group as finite or non-finite. This approach proposes the following :

(i) Annotate the main verb as VM (introduce a new tag). Thus,

h14a. laDZakA seba khAtA_{VM} raHA thA 'boy' 'apple' 'eating' 'PROG' 'was'

h15a. seba khAtA_{VM} huA laDZakA jA rahA thA 'apple' 'eating' 'PROG' 'boy' 'go' 'PROG' 'was'

(ii) Annotate the auxiliaries as VAUX,

h14a. laDZakA seba khAtA_{VM} raHA_{VAUX} thA_{VAUX} 'boy' 'apple' 'eating' 'PROG' 'was'

h15a. seba khAtA_{VM} huA_{VAUX} laDZakA jA rahA thA 'apple' 'eating' 'PROG' 'boy' 'go' 'PROG' 'was'

(iii) Group the verb group (before chunking) and annotate it as finite or non-finite as the case may be,

h14a. laDZakA seba [khAtA_{VM} raHA_{VAUX} wA_{VAUX}]_{VF} 'boy' 'apple' 'eating' 'PROG' 'was' h15a. seba [khAtA_{VM} huA_{VAUX}]_{VNF} laDZakA jA rahA thA 'apple' 'eating' 'PROG'

'boy' 'go' 'PROG' 'was'

This approach is more faithful to the available linguistic information. However, it requires introducing another layer. So, this was not considered useful.

2b) Mark the finiteness at the chunk level,

In this approach, the lexical items are marked as in (2). No new layer is introduced. Instead, the decision is postponed to the chunk level. Since the finiteness is in the group, it is marked at the chunk level. This offers the best solution as it facilitates marking the linguistic information as it is without having to introduce a new layer.

h14a. laDZakA seba ((khAtA_VM raHA_VAUX wA_VAUX))_VGF 'boy' 'apple' 'eating' 'PROG' 'was'

h15a. seba ((khAtA_VM huA_VAUX))_VGNF laDZakA jA rahA thA 'apple' 'eating' 'PROG' 'boy' 'go' 'PROG' 'was'

In this case also the decision is made by looking at the entire group. (2b) was most preferred as it facilitates marking the linguistic information correctly, at the same time no new layer needs to be introduced. Therefore, the current tagging scheme has adopted this approach. Thus, the main verbs in a given verb group will be marked as VM, irrespective of whether the total verb group is finite or non finite. Given underneath are some examples of other verb group types :

1) Non finite verb groups - Non-finite verb groups can have two functions :

a) Adverbial participial, for example : khAte-khAte in the following Hindi sentence,

h16. mEMne khAte khAte ghode ko dekhA 'I erg' 'while eating' 'horse' 'acc' 'saw' I saw a horse while eating.

The main verb in (h16) would be annotated as follows :

h16a. mEMne khAte khAte_VM ghode ko dekhA

b) Adjectival participial, for example : 'khAte Hue' in the following Hindi sentence ,

h17. mEMne ghAsa khAte_VM hue ghoDe ko dekhA * 'I erg' 'grass' 'eating' 'PROG' 'horse' 'acc' 'saw' I saw the horse eating grass.

(* (h17) is ambiguous in Hindi. The other sense that it can have is, I saw the horse while (I was) eating grass. In such cases, the annotator would disambiguate the sentence depending on the context and mark accordingly.)

2) Gerunds

Functionally, gerunds are nominals. However, even though they function like nouns, they are capable of taking their own arguments, eg. pInA in the following Hindi sentence can occur on its own or take an argument (given in parenthesis):

h18. (sharAba) pInA_VM sehata ke liye hAnikAraka hE. 'liquor' 'drinking' 'health' 'for' 'harmful' 'is' Drinking (liquor) is bad for health

h19. mujhe khAnA_VM acchA lagatA hai 'to me' 'eating' 'good' 'appeals' I like eating

h20. sunane meM saba kuccha acchA lagatA hE 'listening' 'in' 'all' 'things' 'good' 'appeal'

'is'

As mentioned above, noun 'sharAba' in (h18) is an object of the verb 'pInA' and has no relation to the main verb (hE). In order to be able to show the exact verb-argument structure in the sentence, it is essential that the crucial information of a noun derived from a verb is preserved. Therefore, even gerunds have to be marked as verbs. It is proposed that in keeping with the approach adopted for non-finite verbs, mark gerunds also as VM at the lexical level. For capturing the information that they are gerunds, such verbs will be marked as VGNN (see the section on Chunk tags for details) at the chunk level to capture their gerundial nature. The verbs having 'vAIA' vibhakti will also be marked as VM. For example, 'khonevAIA' (one who loses).

VAUX: Verb Auxiliary

All auxiliary verbs will be marked as VAUX. This tag has been adopted as such from the Penn tags. (For examples, see h14 h16 above).

3.5.3 JJ: Adjective

This tag is also taken from Penn tags. Penn tag set also makes a distinction between comparative and superlative adjectives. This has not been considered here. Therefore, in the current scheme for Indian languages, the tag JJ includes the 'tara' (comparative) and the 'tama' (superlative) forms of adjectives as well. For example, Hindi adhikatara (more times), sarvottama (best), etc. will also be marked as JJ.

inclusion patterns

3.5.4 RB: Adverb

For the adverbs also, the tag RB has been borrowed from Penn tags. Similar to the adjectives, Penn tags make a distinction between comparative and superlative adverbs as well. This distinction is not made in this tagger. This is in accordance with our philosophy of coarseness in linguistic analysis. Another important decision for the use of RB for adverbs in the current scheme is that :-

(a) The tag RB will be used ONLY for 'manner adverbs' . Example, h21. vaha jaldI jaldI khA rahA thA 'he' 'hurriedly' 'eat' 'PROG' 'was'

(b) The tag RB will NOT be used for the time and manner expressions unlike English where time and place expressions are also marked as RB. In our scheme, the time and manner expressions such as 'yahAz vahAz, aba waba ' etc will be marked as PRP.

Definition to be more clear What should be done with 'jZora se' Should it be a 'adverb' or a 'noun PSP' sequence ? Clear examples, more examples to help annotators List the adverbs Items which are not 'manner adverbs' but are clear adverbs otherwise, include them.

3.5.5 RP: Particle

Expressions such as bhI, to, jI, sA, hI, nA, etc in Hindi would be marked as RP. The nA in the above list is different from the negative nA. Hindi and some other Indian languages have an ambiguous 'nA' which is used both for negation (NEG) and for reaffirmation (RP). Similarly, the particle wo is different from CC wo. For example in Bangla and Hindi:

Bangla : (b1) tumi nA_{RP} khub dushtu 'you' 'particle' 'very' 'naughty' You are very naughty (comment)

Hindi : (h24) tuma nA_{RP}, bahuta dushta ho 'you' 'particle very naughty You are very naughty (comment)

Bangla : (b2) cheleta dushtu nA_{NEG} 'the boy' 'naughty' 'not' The boy is not naughty Hindi : (h25) mEM nA_{NEG} jA sakUMgA 'I' 'not' 'go' 'will able' I will not be able to go

Bangla : (b3) binu yYoxi khAya to_{CC} Ami khAba 'Binu' 'if' 'eats' 'then' 'I' 'will eat' If Binu eats then I will eat (too) Hindi : (h26) yadi binu khAyegA wo_{CC} mEM khAUMgI 'if' 'Binu' 'eats' 'then' 'I' 'will eat' Only if Binu eats, I will eat (too) Bangla : (b4) Ami to_{RP} jAni nA 'I' 'particle' 'know' 'not' I dont know Hindi : (h27) mujhako to_{RP} nahIM patA 'I' 'particle' 'not' 'know' I dont know

Is 'baje' in 'tIna baje' an RP or a VM

Should we put 'kevala' and 'sirpha' as RP only OR RP and RB OR JJ and RB

3.5.6 CC: Conjuncts (co-ordinating and subordinating)

The tag CC will be used for both, co-ordinating and subordinating conjuncts. The Penn tag set has used IN tag for prepositions and subordinating conjuncts. Their rationale behind this is that subordinating conjuncts and prepositions can be distinguished because subordinating conjuncts are followed by a clause and prepositions by a noun phrase.

But in the current tagger all connectives, other than prepositions, will be marked as CC.

h28. mohana bAzAra jA rahA hE Ora_{CC} ravi skUla jA rahA hE 'Mohan' 'market' 'go' 'PROG' 'is' 'and' 'Ravi' 'school' 'go' 'PROG' 'is' Mohan is going to the market and Ravi is going to the school

h29. mohana ne mujhe batAyA ki_{CC} Aja bAzAra banda hE 'Mohan' 'erg' 'to me' 'told' 'that' 'today' 'market' 'close' 'is' Mohan told me that the market is closed today.

Is 'cAhe' a CC or VM ? Such cases to be discussed

5.11 WQ Question Words The Penn tag set makes a distinction between various uses of 'wh-' words and marks them accordingly (WDT, WRB, WP, WQ etc). The 'wh-' words in English can act as questions, as relative pronouns and as determiners. However, for Indian languages we need not keep this distinction. Therefore, we tag the question words as WQ.

h30. kOna AyA hE ? 'who' 'come' 'has' Who has come ?

h31. tuma kala kyA kara rahe ho ? 'you' 'tomorrow' what' 'doing' 'are' What are you doing tomorrow ?

h32. tuma kala kahAz jA rahe ho ? 'you' 'tomorrow' 'where' 'going' 'are' Where are you going tomorrow ?

h33. kyA tuma kala Aoge ? '?' 'you' 'tomorrow' 'will come' Will you come tomorrow ?

3.5.7 QF: Quantifiers

All quantifiers like Hindi kama (less), jyAdA (more), bahuwa (lots), etc. will be marked as QF.

h34. vahAz bahuta_{QF} loga Aye the 'there' many' 'people' 'came' 'was' Many people came there.

In case these words are used in constructions like 'baHutoM ne jAne se inkAra kiyA' ('many' 'by' 'to go' 'refused'; Many refused to go) where it is functioning like a noun, it will be marked as NN (noun). Quantifiers of number will be marked as below.

3.5.8 QC: Cardinals

Any word denoting a cardinal number will be tagged as QC. Penn tag set has a tag CD for cardinal numbers and they have not talked of ordinals. For example:

h35. vahAz tIna_{QC} loga bEThe the 'there' 'three' 'people' 'sitting' 'were' Three people were sitting there

3.5.9 QO: Ordinals

Expressions denoting ordinals will be marked as QO.

h36. mEMne kitAba tIsare_{QO} laDake ko dI thI 'I' 'book' 'third' 'boy' 'to' 'give' 'was' I gave the book to the third boy

3.5.10 INTF: Intensifier

This tag is not present in Penn tag set. Words like 'bahuta', 'kama', etc. when intensifying adjectives or adverbs will be annotated as INTF. Example,

h37. hEdarAbAda meM aMgUra bahuta_{INTF} acche milate hEM 'HyderabAd' 'in' 'grapes' 'very' 'good' 'available' 'are' Very good grapes are available in Hyderabad.

vaha badZde joSa meM bolA Whether 'badZe' should be marked as a JJ or QF . Chk - replace it with 'bahuta' and see

3.5.11 INJ: Interjection

The interjections will be marked as INJ. Apart from the interjections, the affirmatives such as Hindi 'HAz'('yes') will also be tagged as INJ. Since, this is the only example of such a word, it has been clubbed under Interjections.

h38. arre_IINJ, tuma A gaye ! 'oh' 'you' 'come' 'have' Oh! you have come

h39. hAz_IINJ, mEM A gayA 'yes', 'I' 'come' 'have' Yes, I have come.

3.5.12 NEG: Negative

Negatives like Hindi 'nahIM' (not), 'nA' (no, not), etc. will be marked as NEG. For example,

h40. vaha Aja nahIM_NEG A pAyegA 'he' 'today' 'not' 'come' 'will be able'

Also, see examples (b2) and (h25) given above. Indian languages have reiteration of NEG in certain constructions. For example,

b5. tumi chobitA dekhbe ? 'you' 'picture-def' 'will see' ? Will you see the picture ? b6. nA_NEG, xekhabo nA_NEG 'no' 'will see (I)' 'not' No, I will not see (it)

The first occurrence of 'nA' in such constructions will also be marked as NEG.

3.5.13 UT: Quotative

****ocr This should be removed****

A quotative introduces a quote. Typically, it is a verb. Many Indian languages use quotatives. Given below is an example from Bengali,

b7. she Ashbe bole bolechilo 'he' 'will come' 'quotative' 'told' He told that he will come.

3.5.14 SYM: Special Symbol

All those words which cannot be classified in any of the other tags will be tagged as SYM. This tag is similar to the Penn 'SYM'. Also special symbols like \$, %, etc are treated as SYM. Since the frequency of occurrence of such symbols is very less in Indian languages, no separate tag is used for such symbols.

3.5.15 XC: Compounds

In XC, X is a variable of the type of the compound of which the current word is a member of.

The issue of including a tag for marking compounds was discussed extensively. Results of algorithms using IIIT-H tag set which included NNC (part of compound nouns) and NNPC (part of proper nouns) showed that these two tags contributed substantially to the low

accuracy of the tagger. Since most elements which occur as NNC or NNPC can also occur as NN and NNP, it affected the learning by the machine. So, the question was, why to include tags which contributed more to the errors ? The other aspect, however, was that while human annotators are annotating the data, they know from the context when a certain element is NNC or NN, NNPC or NNP and if marked, this information can be useful for certain applications. The argument is same as the one in favor of including a tag for proper nouns.

Another point which was discussed was that any word class can have compound forms in Indian languages (including adjectives and adverbs). Therefore, if we decide to have a tag for showing compounds of each type, the number of tags will go very high. The final decision on this was to include a *C tag which will be realised as catC tag of the type of compound that the element is a part of. For example, if a certain word is part of a compound noun, it will be marked as NNC, if it is part of a compound adjective, it will be marked as JJC and so on and so forth. Some examples are given below :

Hindi compound noun keMdra sarakAra (Central government) will be tagged as keMdra_NNC sarakAra_NN.

In this example, 'keMdra' and 'sarakAra' are both nouns which are forming a compound noun. All words except the last one, of a compound words will be marked as NNC. Thus any NNC will be always followed by another NNC or an NN. This strategy helps identify these words as one unit although they are not conjoined by a hyphen. Similarly, a compound proper noun will be marked as NNPC excluding the last one. eg. aTala_NNPC bihArI_NNPC vAjapeyI_NNP The first two words, in the above example, will be tagged as NNPC and the last one will be tagged as NNP. Similar to the NNC tag for common nouns, NNPC tag helps in marking parts of a proper noun.

h41. rAma, mohana aur shyAma ghara gaye. 'Ram', 'Mohan' 'and' 'Shyam' 'home' 'went'
Ram, Mohan and Shama went home.

h42. bagIce meM ranga_JJC biraMge_JJ phUla khile the 'garden' 'in' 'colourful' 'flowers'
'flowered' 'were' The garden had colorful flowers

Titles such as Dr., Col., Lt. etc. which may occur before a proper noun will be tagged as NNC. All such titles will always be followed by a Proper Noun. In order to indicate that these are parts of proper nouns but are nonetheless nouns themselves, they will be tagged as NNC, eg. Col._NNC Ranjit_NNPC Deshmukh_NNP

3.5.16 RDP: Reduplication

****ocr No reduplication in Hindi?****

In this phenomenon of Indian languages, the same word is written twice for various purposes such as indicating emphasis, deriving a category from another category etc. eg. choTe choTe ('small' 'small'; very small), lAla lAla ('red' 'red'; red), jaldI jaldI ('quickyl' 'quickly' ; very quickly) There are two ways in which such word sequences may be written. They can be written (a) separated by a space or (b) separated by a hyphen.

The question to be resolved is that in case, they are written as two words (separated by space) how should they be tagged? Earlier decision was to use the same tag for both the words. However, in this approach, the morphological character of reduplication is missed out. That is, the reduplicated item will then be treated exactly like two independent words of the same category. For example,

h43. vaha mahaMgI_JJ mahaMgI_JJ cIjZeM kharIda lAyA 'he' 'expensive' 'expensive' 'things' 'buy' 'bring' He bought all expensive things. h44. una catura_JJ buddhimAna_JJ baccoM ne samasyA sulajhA II 'those' 'smart' 'intelligent' 'children' 'erg' 'problem' 'solved' Those smart and intelligent children solved the problem.

Both (h43) and (h44) have a sequence of adjectives - mahaMgI_JJ mahaMgI_JJ and catura_JJ buddhimAna_JJ respectively. In the first case, the sequence of two adjectives is a case of reduplication (same adjective is repeated twice to indicate the intensity of 'expensive') whereas in the second case the two adjectives refer to two different properties attributed to the following noun. Since reduplication is a highly productive process in Indian languages, it is proposed to include a new tag RDP for annotating reduplicatives. The first word in a reduplicative construction will be tagged by its respective lexical category and the second word will be tagged as RDP to indicate that it is a case of reduplication distinguishing it from a normal sequence such as in (h44) above. Some more examples are given underneath to make it more explicit,

h45. vaha dhIre_RB dhIre_RDP cala rahA thA. 'he' 'slowly' 'slowly' 'walk' 'PROG' 'was' He was walking (very) slowly. h46. usake bAla choTe_JJ choTe_RDP the. 'his' 'hair' 'short' 'short' 'were' He had (very) short hair h47. yaha bAta galI_NN galI_RDP meM phEla gayI. 'this' 'talk' 'lane' 'lane' 'in' 'spread' 'went' The word was spread in every lane.

3.5.17 ECH: Echo words

Indian languages have a highly productive usage of echo words such as Hindi 'cAya-vAya' ('tea' 'echo'), where 'cAya' is a regular lexical item of Hindi vocabulary and 'vAya' is an echo word indicating the sense etc. These words, on their own, are 'nonsense' words and do not find a place in any dictionary. Thus, the gloss for 'cAya-vAya' would be 'tea etc'. It is proposed to add the tag ECH for such words.

3.5.18 UNK: Unknown

A special tag to indicate unknown words is also included in the tag set. The annotators can use this tag to mark the words whose category they are not aware of. This tag has to be used very cautiously and sparsely, i.e., only if it is absolutely necessary.

3.6 Some Special Cases

This section gives the details of certain aspects of Indian languages which need to be dealt with separately in the tagger. These are issues that cannot be handled by just changing or adding tags.

3.6.1 *vAlA*-type constructions

'vAlA' is a kind of suffix used in Hindi and some other Indian languages. It conjoins with nouns (Case I, below) or verbs (Case II) to form adjectives or even nouns. It is also used as an aspectual TAM in a verbal construction (Case III).

h48. lAla kamIjZa vAlA AdamI merA bhAI hE . 'red' 'shirt' 'in' 'man' 'my' 'brother' 'is'
The man in red shirt is my brother.

h49. mehanawa karane vAle vyakti ko inAma milegA . 'hard work' 'doing' 'adj' 'person' 'to'
'prize' 'will get' The person who works hard will get a prize.

These cases are elaborated below.

Case I: The suffix 'vAlA' can occur with a noun. For example, lAThI vAlA ('stick' 'with' -The one with a stick).

h50. lAThI vAle AdamI ko bulAo 'with stick' 'man' 'acc' 'call' Call the man with the stick.

This suffix 'vAlA' in Hindi (a) may be written separately or (b) may be attached to the preceding noun.

(a) In case it is written separately as in 'lAThI vAlA' above, the word 'lAThI' will be tagged as NN and the word 'vAlA' will be tagged as PSP.

The whole expression 'lAThI vAlA' is an adjective, in which 'lAThI' is a noun and 'vAlA' is a suffix which derives an adjective from a noun. Since 'lAThI' and 'vAlA' written separately in the above example, they have to be tagged individually. 'vAlA' in such cases will be treated like a postposition and will be tagged as PSP.

(b) The second possibility is of 'lAThI' and 'vAlA' written together as 'lAThIvAlA'. In such cases it will be treated as one word and will be marked as JJ since 'lAThIvAlA' is an adjective.

Case II: 'vAlA' can also occur after a verb. Example, karane vAlA ('doing' 'one' The one who does something)

h51. mehanata karanevAle ko phala milatA hE 'hard' 'working one' acc 'fruit' 'get' 'PRES'
The one who works hard gets the fruits.

As mentioned earlier, the suffix 'vAlA' also joins a verb in its nominal form and makes it an adjective. In this case also, the two words may be written separately (karane vAle) or together (karanevAlA). In the former case, the two words will be marked as VM and PSP respectively (karane_{VM} vAle_{PSP}). In the latter case, being a single word (karanevAlA) it will be tagged as VM (karanevAle_{VM}). It is crucial to retain the 'verb' information in these cases, so that at a later stage if we want to annotate its argument structure we should be

able to do so (discussed earlier in the document).

Case III: 'vAlA' can also occur as part of TAM. For example,

h52. mEM wo jAne vAlA hI thA. 'I' 'particle' 'to go' 'about' 'part' 'was' I was about to go

Although the word 'jAne' has a 'vAlA' suffix in (h52) above, the entire expression is not an adjective but is a verb having the aspectual information of 'shortly'. In this case, the sequence 'jAne vAlA' will be marked as jAne_VM vAlA_VAUX. The alternative writing convention of writing the sequence as one word (jAnevAlA) is possible in this case also. Like the earlier cases, the word will be marked for the category of the content morpheme which is verb in this case. Thus jAnevAle will be tagged as jAnevAle_VM.

Here again we stand by our policy that the tag will be decided on the basis of the part of speech and not on the basis of the category of the word in the given sentence (syntactic function). This avoids confusion at the level of manual tagging and aids machine learning as well. So the tag (VM) remain same although the function of the words is different in two different places, it is adjective in Cases I and II and verbal in Case III.

3.6.2 Honorifics in Indian languages

Hindi (and some other Indian languages) has particles such as 'jI' or 'sAHaba' etc. after proper nouns or personal pronouns. These particles are added to denote respect to the referred person. Such honorific words will be treated like particles and will be tagged RP like other particles.

h53. mantr_{IN}N jI_{RP} sabhA meM dera se pahuMce . 'minister' 'hon' 'meeting' 'in' 'late' 'part' 'reached' The minister reached late for the meeting.

3.6.3 Foreign words

Presence of loan words is a fairly common phenomenon in languages. Most Indian languages have a number of loan words from English. One may also come across words from other Indian languages or Sanskrit in a given text. Such foreign words will be tagged as per the syntactic function of the word in the given context. In special cases, such as when the annotator is not sure of the category of a word, it will be tagged as UNK.

3.7 A Special Note

There may be situations when an annotator does not feel very confident about the tag for a particular word. The annotator may then assign it different tags in different places. Inconsistency in the manual tagging can affect the learning considerably. Since this is a task which involves a number of human annotators, the methods have to be evolved to check and cross validate the human annotation. Another practical problem in annotation is that in the initial stages of annotation, the annotators need time to get familiar with the tagging scheme and the concept behind each tag. Thus they take some time before coming

to a stable stage of decision making for various instances, particularly various ambiguous cases. Especially, in the initial stages, the annotators may often come across cases where their confidence level may not be very high. They may feel the need of some clarifications for these cases. Since the task of annotation has to go on and immediate clarification may not be possible, the annotators may be forced to take decisions and mark a case as they consider appropriate at that point of time. Over a period of time, with better understanding of the tags and tagging scheme, they may reach a stable stage. However, by then they may already have tagged a given case differently in different places thus introducing inconsistency in the annotated corpus. At a later stage, it will be difficult to go back to all the cases that have been annotated by then and correct them. So the chances are that the annotators may proceed with the revised decision and leave the earlier annotation as such. This will introduce inconsistencies in the annotated corpus.

To control such a situation, it is decided to provide a way by which the annotators can initially mark the uncertainty of their decision so that they can easily extract these cases easily and take them up for discussions and clarifications.

This 'uncertainty' will be annotated as follows :

1. The annotators first mark such a case with a tag that they consider appropriate at the time of annotation.
2. Along with the chosen tag, they also put a question mark (?) against that tag. The question mark will indicate that this case is not yet resolved and will be finalized after clarification or discussion.
3. All the cases with a question mark can be later taken out and placed for discussion. An annotator will be responsible for bringing such cases for discussion and once the cases are resolved, the annotator will go back and correct the tag. In case the tag assigned by the annotator initially itself is correct, the annotator will remove the question mark against it.

This is a purely temporary measure and the data finally submitted by an annotator should not have any words having a question mark.

3.8 Chunk Tags Chosen for the Current Scheme

This section deals with the chunk tags. Not many of the issues discussed above hold for defining the chunk tags. Various points which have been deliberated upon in relation to chunking scheme are:

1. Definition of a chunk
2. Chunk Types
3. Some Special Cases
4. Annotation method/procedure

3.8.1 Definition of a chunk

Following issues related to the definition of a chunk were discussed : What constitutes a 'chunk' ?

A typical chunk consists of a single content word surrounded by a constellation of function words (Abney,1991). Chunks are normally taken to be a 'correlated group of words'.

The next issue, however, is - How to define the boundaries of these 'correlated word groups' for our purpose?

For example, which case in the following pairs should be grouped as a chunk ?

((xillI meM)) OR ((xillI)) meM 'Delhi' 'in' 'Delhi' 'in' ((rAjA kA betA)) OR ((rAjA kA)) ((betA)) 'king' 'of' 'son' 'king' 'of' 'son' ((rAjA ke bete kI paxnI)) OR ((rAjA ke)) ((bete kI)) ((paxnI)) 'king' 'of' 'son' 'of' 'wife' 'king' 'of' 'son' 'of' 'wife'

Following definition of a 'chunk' was evolved through discussion :

A minimal (non recursive) phrase(partial structure) consisting of correlated, inseparable words/entities, such that the intra-chunk dependencies are not distorted. Each chunk type discussed and the decided upon is described below . 8.2. Chunk Types

Based on the above definition of chunk, issues related to various chunk types were discussed. A chunk would contain a 'head' and its modifiers.

NP: Noun Chunk

Noun Chunks will be given the tag NP and include non-recursive noun phrases and postpositional phrases. The head of a noun chunk would be a noun. Specifiers will form the left side boundary for a noun chunk and the vibhakti or head noun will mark the right hand boundary for it. Descriptive adjective/s modifying the noun will be part of the noun chunk. The particle which anchors to the head noun in a noun chunk will also be grouped within the chunk. If it occurs after the noun or vibhakti, it will make the right boundary of the chunk. Some example noun chunks are :

((bacce_NN))_NP, ((kucha_QF bacce_NN))_NP, 'children' 'some' 'children' ((kucha_QF acche_JJ bacce_NN))_NP, ((Dibbe_NN meM_PSP))_NP, 'some' 'good' 'children' 'box' 'in' ((eka_QC kAlA_JJ ghoDZA_NN))_NP, 'one' 'black' 'horse' ((yaha_DEM nayL_JJ kitAba_NN))_NP, 'this' 'new' 'book' ((isa_DEM nayL_JJ kitAba_NN meM_PREP))_NP, 'this' 'new' 'book' 'in' ((isa_DEM nayL_JJ kitAba_NN meM_PSP bhLRP))_NP 'this' 'new' 'book' 'in' 'also'

The issue of genitive marker and its grouping with the nouns that it relates to was discussed in detail. For example, the noun phrase 'rAma kA beTA' contains two nouns 'rAma' and 'beTA'. The two nouns are related to each other by the vibhakti 'kA'. The issue is whether to chunk the two nouns separately or together? Linguistically, 'beTA' is the head of the phrase rAma kA beTA. 'rAma' is related to 'beTA' by a genitive relation which is expressed through the vibhakti 'kA'. Going by our definition of a 'chunk' we should break 'rAma kA beTA' into two chunks (((rAma kA))_NP, ((beTA))_NP) by breaking 'rAma kA' at 'kA' vibhakti . Moreover, if we chunk 'rAma kA beTA' as one chunk, linguistically, we will end up

with a recursive noun phrase as a single chunk (((rAma kA)) beTA)) which also is against our definition of a chunk.

Therefore, it was decided that the genitive markers will be chunked along with the preceding noun. Thus, the noun group 'rAma kA beTA' would be chunked into two chunks.

h54. ((rAma kA))NP ((beTA))NP acchA hE Ram's son is good h55. ((kitAba))NP ((rAma kI))NP hE The book belongs to Ram

For the noun groups such as usakA beTA it was decided that they should be chunked together.

Verb Chunks

The verb chunks would be of several types. A verb group will include the main verb and its auxiliaries, if any. Following are some examples of verb chunks from Hindi,

((khAyA)), ((khA rahA hE)), ((khA sakawe hEM)) 'ate' 'eat' 'PROG' 'is' 'eat' 'can' 'PRES'

The types of verb chunks and their tags are described below.

VGF: Finite Verb Chunk

As mentioned in 5.4 above, a verb group sequence (V VAUX VAUX . .) contains a main verb and its auxiliaries. The group itself can be finite or non-finite. In case of it being finite, the main verb in such a sequence may not be finite. The finiteness is known by the auxiliaries. Therefore, it is decided to mark the finiteness of the verb at the chunk level. Thus, any verb group which is finite will be tagged as VGF. For example,

h56. mEMne ghara para khAnA ((khAyA_VM))_VGF 'I erg' 'home' 'at' 'meal' 'ate'

h57. vaha cAvala ((khA_VM rahA_VAUX hE_VAUX))_VGF 'he' 'rice' 'eat' 'PROG' 'is'

VGNF: Non-finite Verb Chunk

A non-finite verb chunk will be tagged as VGNF. For example,

h15a) seba ((khAtA_VM huA_VAUX))_VGNF laDZakA jA rahA thA 'apple' 'eating' 'PROG' 'boy' 'go' 'PROG' 'was'

h16a) mEMne ((khAte khAte_VM))_VGNF ghode ko dekhA 'I erg' 'while eating' 'horse' acc 'saw' h17a) mEMne ghAsa ((khAte_VM hue_VAUX))_VGNF ghoDe ko dekhA 'I erg' 'grass' 'eating' 'PROG' 'horse' acc 'saw'

The IIIT-H tagset had initially included three tags for the non-finite verbal forms. Unlike Penn tagset, all non finite verbs, which are used as adjectives, were marked as VJJ at the POS level. Similarly, to mark adverbial non-finite verbs, the POS tagset had VRB tag. A tag VNN was included to mark the nominalized verbs.

However, during the discussions IL standards, it was pointed out that inclusion of too many finer tags hampers machine learning. Moreover, the marking is based on syntactic informa-

tion, which we should avoid at the POS level, unless it is contributing to further processing in a substantial way. On the other hand, it is important to mark finite non-finite distinction in a verbal expression as it is a crucial information and is also easy to learn. As discussed under 5.4 above, it was decided to mark this distinction at the chunk level, rather than at the POS level. Therefore, the tag VGNF has been included to mark non-finite adverbial and adjectival verb chunk.

VGINF: Infinitival Verb Chunk

This tag is to mark the infinitival verb form. In Hindi, both, gerunds and infinitive forms of the verb end with a -nA suffix. Since both behave functionally in a similar manner, the distinction is not very clear. However, languages such as Bangla etc have two different forms for the two types. Examples from Bangla are given below.

b8. Borabela ((snAna karA))_VGNN SorIra pokze BAlo 'Morning' 'bath' 'do-verbal noun' 'health-gen' 'for' 'good' Taking bath in the early morning is good for health

b9. bindu Borabela ((snAna karawe))_VGINF BAlobAse 'Bindu' 'morning' 'bath' 'take-inf' 'love-3pr' Bindu likes to take bath in the early morning

In Bangla, the gerund form takes the suffix A / -Ano, while the infinitive marker is we. The syntactic distribution of these two forms of verbs is different. For example, the gerund form is allowed in the context of the word darakAra necessary while the infinitive form is not, as exemplified below:

b10 Borabela ((snAna karA))_VGNN darakAra 'Morning' 'bath' 'do-verbal noun' 'necessary' It is necessary to take bath in the early morning

b11. *Borabela ((snAna karawe))_VGINF darakAra

Based on the above evidence from Bangla, the tag VGINF has been included to mark a verb chunk.

VGNN: Gerunds

A verb chunk having a gerund will be annotated as VGNN. For example,

h18a. sharAba ((pInA_VM))_VGNN sehata ke liye hAnikAraka hE. 'liquor' 'drinking' 'health' 'for' 'harmful' 'is' Drinking (liquor) is bad for health

h19a. mujhe rAta meM ((khAnA_VM))_VGNN acchA lagatA hai 'to me' 'night' 'in' 'eating' 'good' 'appeals' I like eating at night

h20a. ((sunane_VM meM_PSP))_VGNN saba kuccha acchA lagatA hE 'listening' 'in' 'all' 'things' 'good' 'appeal' 'is'

JJP: Adjectival Chunk

An adjectival chunk will be tagged as JJP. This chunk will consist of all adjectival chunks including the predicative adjectives. However, adjectives appearing before a noun will be

grouped together with the noun chunk. A JJP will consist of phrases like

h58. vaha laDaZkI hE((suMdara_JJ sLRP))_JJP 'she' 'girl' 'is' 'beautiful' 'kind of'

h59. hAthI AyA ((moTA_*C tagadA_JJ))_JJP 'elephant' 'came' 'fat' 'powerful'

h60. vaha laDakI ((bahuta_INTF sundara_JJ))_JJP hE 'she' 'girl' 'very' 'beautiful' 'is'

Cases such as (h61) below will not have a separate JJP chunk. In such cases, the adjectives will be grouped together with the noun they modify. Thus forming a NP chunk.

h61. ((kAle_JJ ghane_JJ laMbe_JJ bAla_NN))_NP 'black' 'thick' 'long' 'hair'

The following examples from Hindi present a special problem:

h62. xillI meM rahanevAlA merA BAI kala A rahA hE . 'Delhi' 'in' 'staying' 'my' 'brother' 'tomorrow' 'come' 'PROG' 'is' My brother who stays in Delhi is coming tomorrow.

h63. usane Tebala para rakhA huA seba khAyA. '(s)he erg' 'table' 'on' 'kept' 'apple' 'ate' He ate the apple kept on the table.

In (h62) above 'rahanevAlA' is an adjectival participle. But we do NOT mark it as JJP. Instead, it will be marked as a VGNF. The decision to tag it as a VGNF is based on the fact that such adjectival participles are derived from a verb can have their arguments. This information is useful for processing at the syntactic level. Thus, 'rahanevAlA' in (h62) will be annotated as follows:

h62a. xillI meM ((rahanevAlA_VM)_VGNF merA BAI kala A rahA hE .

Similarly, in (h63) above, the chunk 'rakhA huA' is an adjective but will also be marked as a VGNF since this also derived from a verb and chunks like 'Tebala pra' etc are its arguments. So the chunk name will be VGNF and the POS tag will be VM which might be followed by an auxiliary verb tagged as VAUX. (h63a) shows how 'rakhA huA' will be annotated :

h63a. usane Tebala para ((rakhA_VM huA_VAUX))_VGNFseba khAyA.

RBP: Adverb Chunk

This chunk name is again in accordance with the tags used for POS tagging. This chunk will include all pure adverbial phrases.

h64. vaha ((dhIre-dhIre_RB))_RBP cala rahA thA. 'he' 'slowly' 'walk' 'PROG' 'was' He was walking slowly

Now consider the following examples:

h65. vaha dagamagAte hue cala rahA thA . 'he' 'walk' 'PROG' 'was' He was walking

h66. vaha khAnA khAkara ghara gayA . 'he' 'meal' 'after eating' 'home' 'went' He went home after eating his meal

In the above examples, 'dagamagAte hue' and 'khAkara' are non finite forms of verbs used as adverbs. Similar to adjectival participles these will also be chunked as VGNF and not as RBP. The reason for this is that we need to preserve the information that these are underlying verbs. This will be a crucial information at the level of dependency marking

where the arguments of these verbs will also be marked.

((isa.PRP nayI.JJ kitAba.NN meM.PSP bhI.RP))_NP 'this' 'new' 'book' 'in' 'also'

NEGP: Negatives

(i) In case a negative particle occurs around a verb, it is to be grouped within verb group. For example,

h67. mEM kala dillI ((nahIM_NEG jA_VM rahI_VAUX))-VGF "I" "tomorrow" "Delhi" "not" "go" "Cont"

h68. ((binA_NEG bole_VM))-VGNF kAma ((nahIM_NEG calatA_VM))-VGF "without" "saying" "work" "not" "happen"

However ,

h69. binA kucha bole kAma nahIM calatA "without" "something" "saying" "work" "not" "happen"

In the above sentence, the noun "kucha" is coming between the negative "binA" and verb "bole". Here, it is not possible to group the negative and the verb as one chunk. At the same time, "binA" cannot be grouped within an NP chunk, as functionally, it is negating the verb and not the noun. To handle such cases an additional NEGP chunk is introduced. If a negative occurs away from the verb chunk, the negative will be chunked by itself and chunk will be tagged as NEGP. Thus,

h69a. ((binA))-NEGP ((kucha))-NP ((bole))-VG ((kAma))-NP ((nahIM calatA))-VG

CCP: Conjuncts

Conjuncts are functional units information about which is required to build the larger structures. Take the following examples of conjunct usages :

h70. (rAma kitAba paDha rahA thA) Ora (mohana Tennisa khela rahA thA). Ram was reading a book and Mohan was playing tennis

h71. (rAma ne batAyA) ki (usakI kitAba acchI hE). Ram said that his book is good

h72. (rAma) Ora (mohana) Tennisa khela rahe the. Ram and Mohan were playing tennis.

h73. (merA bhAI rAma) Ora (usakA dostA mohana) Tennisa khela rahe the. My brother Ram and his friend Mohan were playing tennis.

h74 . rAma (saphZeda kapade) Ora (nIle jute) pahane thA. Ram was wearing white clothes and blue shoes.

h75. rAma eka (halkI) Ora (nIII) bOla lAyA. ram brought a light and blue ball.

The sentences above have various types of conjoined structures. To represent these conjoined structures, it is decided to form separate chunks for conjuncts and the elements a conjunct conjoins. Thus (h70) and (h71) above will be chunked as (h70a) and (h71a) given below,

h70a. ((rAma))-NP ((kitAba))-NP ((paDha rahA thA))-VG ((Ora))CCP ((mohana))-NP

((TennisA))_NP ((khela rahA thA))_VG.

h71a. ((rAma ne))_NP ((batAyA))_NP ((ki))_CCP ((usakI))_NP ((kitAba))_NP ((acchI))_JJP ((hE))_VG.

Expression 'rAma Ora mohana' in example (h72) is a complex NP. Though complex, the expression can be annotated as a single NP chunk as functionally it is the subject of the verb 'play'. However, example (h73) presents a case where it would be better to form three independent chunks for the complex subject NP. Though the conjunct 'Ora' is conjoining 'rAma' and 'mohana', both 'rAma' and 'mohana' have their respective modifiers. To make it explicit, it is better to treat them as two independent NP chunks conjoined by a CCP.

h73a. ((merA bhAI rAma)) ((aura))_CCP ((usakA dosta mohana))_NP ((TennisA))_NP ((khela rahe the))_VG.

Following this, the subject NP of (h72) would also be annotated similarly. Therefore,

h72a. ((rAma))_NP ((aura))_CCP ((mohana))_NP ((TennisA))_NP ((khela rahe the))_VG.

The annotation for cases such as (h74) and (h75) would be as follows :

h74a. ((rAma))_NP ((safeda kapade))_NP ((aura))_CCP ((nIle jute))_NP ((pahane thA))_VG.

h75a. ((rAma))_NP ((eka))_JJP ((halkI))_JJP ((aura))_CCP ((nIII))_JJP ((bOla))_NP ((lAyA))_VG

Thus the decision for conjuncts is - the conjoined entities will be broken into separate chunks.

eg. ((rAma))_NP ((Ora))_CCP ((SyAma))_NP

FRAGP: Chunk Fragments

Some times certain fragments of chunks are separated from the chunks to which they belong. For example :

h76. rAma (jo merA baDZA bhAI hE) ne kahA ... 'Ram' 'who' 'my' 'elder' 'brother' 'is' 'erg' 'said'

In the above example, vibhakti 'ne', which is a case marker of the noun 'rAma', is separated from it by an intervening clause. Syntactically, 'ne' is a part of the noun chunk 'rAma ne'. However, at times it can be written separately. The following was decided for such fragments :

(i) There will be a separate chunk for the vibhakti in constructions where it gets separated from the noun it would normally be grouped with. This chunk can have more than one entity within it.

h77. ((rAma))_NP, mere dilli vAle bhAI, ((ne))_FRAGP kahA 'Ram' 'my' 'Delhi' 'from' 'brother' 'erg' 'said'

(ii) If the entities embedded between the noun and it's vibhakti are a series of nouns the entire group will be chunked as a single noun chunk.

h78. ((isa 'upanyAsa samrATa' Sabda kA))_NP 'this' 'Novel' 'King' 'word' 'of'

BLK: Miscellaneous entities

Entities such as interjections and discourse markers that cannot fall into any of the above mentioned chunks will be kept within a separate chunk. eg. ((oh_INJ))_BLK, ((arre_INJ))_BLK

3.8.2 Some Special Cases

Apart from the above, some special cases related to certain lexical types are discussed below.

Conjunct Verbs

****ocr Are these in fact light verb constructions?****

The issue whether to treat the noun/adjective which is part of a conjunct verb differently by marking it with a special tag (NVB/JVB) or to treat it as a noun like any other noun at the POS level was deliberated on. The question was based on the following observations :

a) NVB/JVB , as part of conjunct verbs, are most often not recognized by the learning algorithms.

b) Having NVB at the POS level is based on syntactic considerations. Therefore, do we really need to go for it ? Instead, at the POS level we mark the noun as a noun and leave the decision of marking a conjunct verb as single unit for a later level.

c) Moreover, since the noun, which is part of a conjunct verb (Kriyamula), can occur away from its 'verbaliser', it becomes difficult to differentiate it from a 'noun' which may be an argument of the verb. This also creates problem for chunking of the verb group. The two components of the chunk have to be separately marked and have to be joined at the syntactic level.

d) If NVB is marked at the POS level, a natural consequence would be to group it with its verbaliser as a VG chunk. In fact, that is the purpose of identifying it as different from a noun. However, sometimes one comes across expressions such as 'mEMne unase eka prashna kiyA' (I posed a question to him). In this sentence, 'eka' is a modifier of 'prashna'. 'prashna karana' is recognized as a conjunct verb in Hindi by most Hindi speakers. Following example shows the problem of grouping 'praSna karanA' as a single VG:

POS : mEMne_PRP unase_PRP eka_QC prashna_NVB kiyA_VM Chunk : ((meMne_PRP))_NP ((unase_PRP))_NP ((eka_QC))_JJP ((prashna_NVB kiyA_VM))_VGF

Once praSna karanA are grouped together as a chunk, it will be difficult to show the relation between 'eka' and 'prashna' subsequently.

Thus, an alternative was proposed wherein, the noun of the conjunct verb is tagged as NN at the POS level which is accordance with the decision to tag the lexical item based on its lexical category. Thereafter, the noun is grouped with its preceding adjectival modifiers as an NP chunk. The only problem in this approach is that the information of a noun verb sequence being a conjunct verb is not captured till the chunk level and the noun of the conjunct verb is separated from its verbaliser. However, the approach has following advantages :

1) At the POS level, the word is tagged for its grammatical category and not for its syntactic function. This eases the decision making at the POS level. And marking the information, that the conjunct verbs which are composed of two words form one lexeme semantically, is postponed to a later level.

2) It allows us to show the modifier-modified relation between an adjective such as 'eka' in the above example with its modified noun 'praSna'.

3) Since the information of a noun verb sequence being a 'kriyamula' is crucial at the syntactic level, it will be captured at that level by marking the relation between the 'noun' and its verbaliser by an appropriate tag. Therefore, the decision is :

The noun/adjective and verb (internal components of a conjunct verb) will be chunked separately. eg. prashna karanA - ((prashna))NP ((kiyA))VG ucita kiyA - ((ucita))JJP ((kiyA))VG

Particles

Regarding the particles, it was decided that the particles will be chunked with the same chunk as the anchor word they occur with. Thus, eg. ((rAma ne bhI))_NP, ((mEM wo))_NP, 'Ram' 'erg' 'also' 'I' 'emph'

Quantifiers

The issue of chunking quantifiers was discussed in great details. Numbers can occur (a) as noun modifiers before a noun (haZaroM ladakoM ne 'thousands' 'boys' 'erg') or (b) can occur without a noun (hazAroM ne 'thousands' 'erg') with a nominal inflection. The issue of whether to treat the quantifiers of the type (b) as nouns was discussed. The issue is whether (b) is a case of an ellipsis of the noun after a number or whether it is the number itself which is the noun. If the latter has to be followed then the POS tag for quantifiers in such cases should be NN. Following decisions were taken :

(i) A 'QC' or 'QO' occurring with a noun will be part of the noun chunk.

h79. ((hazAroM_QC logoM_NN ne_PSP))_NP yaha driSya dekhA 'thousands' 'people' 'erg' 'this' 'scene' 'watched' Thousands of people watched this scene.

h80. ((dUsare_QO ladake_NN ne_PSP))_NP isa samasyA ko sulajhA diyA 'second' 'boy' 'erg' 'this' 'problem' acc 'solve' 'did' The second boy solved this problem.

(ii) All categories occurring without a noun, with nominal inflections (overt or otherwise) will be tagged as noun.

h81. ((hazAroM_NN ne_PSP))_NP yaha driSya dekhA 'thousands' 'erg' 'this' 'scene' 'watched' Thousands watched this scene. h82. ((mote_NN ne_PSP))_NP ((chote_NN ko_PSP))_NP ((mArA))_VGF 'fat' 'erg' 'small' 'to' 'killed'

Punctuation

All punctuations, with an exception of sentence boundary markers and clausal conjuncts, will be included in the preceding chunk. For example

h83. ((usane_PRP))_NP ((kahA_VM _SYM)_VGF 'He erg' 'said' (" _SYM yaha_PRP))_NP ((Thika_JJ))_JJP ((hE_VM " _SYM))_VGF 'this' 'proper' 'is' He said, this is not right .

h84. rAma AyA ((,_SYM))_CCP mohana gayA 'Ram' 'came' , 'Mohan' 'went' Ram came and Mohan left.

Punctuations such as (a) hyphens and (b) quote marks will be taken care of by the tokenizer.

(a) Hyphens: Identified to be of two types:- - Without space on either sides, as in the case of compound nouns eg. mAtA-pitA(mother-father)

With spaces, as in the case of

h85. rAma ne kahA yaha thIk hE 'Ram' 'erg' 'said' 'this' 'proper' 'is'

(b) Quote marks (single and double both) : Identified to be of two types:- (i) opening (ii) closing

3.9 Annotation Procedure

To maintain consistency in the data format and the annotation, it was decided to use 'San-chay', a facility developed at IIIT, Hyderabad for the annotation task.

3.10 Conclusion

The annotation standards for POS tagging and chunking for Indian languages include 26 tags for POS (Table-1 in Appendix) and 11 chunk tags (Table-2 in Appendix). The tags are decided on coarse linguistic information with an idea to expand it to finer knowledge if required.

****ocr These are notes:*

AzsU-bharii 'bharii' as a PSP or JJ ?

*Onomatopoeic words have to be considered. A discussion should be included.****

3.11 References

Steven Abney. Parsing by Chunks. In: Robert Berwick, Steven Abney and Carol Tenny (eds.), Principle-Based Parsing. Kluwer Academic Publishers, Dordrecht. 1991.

3.12 Acknowledgements

Following participated in the meetings/discussions :

IIIT Hyderabad Rajeev Sangal, Dipti M Sharma, Soma Paul, Lakshmi Bai, Research students of LTRC, IIIT, Hyderabad. University of Hyd. Amba Kulkarni, G. Uma Maheshwar Rao, Rahmat Yousufzai IITB Pushpak Bhattacharya, Om Damani, Rajat Kumar Mohanty, Sushant S Develkar, Pranesh, Maneesh IITKGP Sudeshna Sarkar, Anupan Basu, Pratyush Banerjee, Sandipan Dandapat CDAC, Pune Saurabh Singhal, Abhishek Gupta, Ritu Bara, Mahendra Pandey CDAC, Noida Vijay Kumar, K K Arora IIIT Allahabad Ratna Sanyal Tamil University S. Rajendran AUKBC, Chennai Sobha Nair Jadhavpur University Shivaji Bandopadhyay

Ms Pranjali Karade prepared the initial document describing IIIT-H tagging scheme which has been an immense help in preparing the current document. Thanks to Soma Paul and Vasudhara Sarkar for providing Bangla examples given in the text.

3.13 Table of POS Tags

****ocr We need to carefully check all uses of POS tags and reconcile them. We need a single table.****

****ocr The following two tables are from the POS document.****

POS Tag Set for Indian Languages (Nov 2006, IIIT Hyderabad)

| Sl No. | Category | Tag name | | Example |
|--------|----------------|-----------|----|----------------------------|
| | | DS and PB | PS | |
| 1.1 | Noun | NN | NN | |
| 1.2 | NLoc | NST | | |
| 2. | Proper Noun | NNP | | |
| 3.1 | Pronoun | PRP | | |
| 3.2 | Demonstrative | DEM | | |
| 4 | Verb-finite | VM | | |
| 5 | Verb Aux | VAUX | | |
| 6 | Adjective | JJ | | |
| 7 | Adverb | RB | | *Only manner adverb |
| 8 | Post position | PSP | | |
| 9 | Particles | RP | | bhI, to, hI, jI, hA.N, na, |
| 10 | Conjuncts | CC | | bole (Bangla) |
| 11 | Question Words | WQ | | |
| 12.1 | Quantifiers | QF | | bahut, tho.DA, kam |
| 12.2 | Cardinal | QC | | |
| 12.3 | Ordinal | QO | | |
| 12.4 | Classifier | CL | | |
| 13 | Intensifier | INTF | | |
| 14 | Interjection | INJ | | |
| 15 | Negation | NEG | | |
| 16 | Quotative | UT | | mAne |
| 17 | Sym | SYM | | |
| 18 | Compounds | *C | | |
| 19 | Reduplicative | RDP | | |
| 20 | Echo | ECH | | |
| 21 | Unknown | UNK | | |

It was decided that for foreign/unknown words that the POS tagger may give a tag UNK

Chunk Tag Set for Indian Languages

| Sl. No | Chunk Type | Tag Name | Example |
|--------|------------------------|----------|---|
| 1 | Noun Chunk | NP | ((merA nayA ghara))_NP 'my new house' |
| 2.1 | Finite Verb Chunk | VGF | mEMne ghara para khAnA ((khAyA_VM))_VGF |
| 2.2 | Non-finite Verb Chunk | VGNF | mEMne ((khAte khAte_VM))_VGNF ghode ko dekhA |
| 2.3 | Infinitival Verb Chunk | VGINF | Bangla : bindu Borabela ((snAna karawe))_VGINF BA |
| 2.4 | Verb Chunk (Gerund) | VGNN | mujhe rAta meM ((khAnA_VM))_VGNN acchA lagatA |
| 3 | Adjectival Chunk | JJP | vaha laDaZkI hE((suMdara_JJ sI_RP))_JJP |
| 4 | Adverb Chunk | RBP | vaha ((dhIre-dhIre_RB))_RBP cala rahA thA |
| 5 | Chunk for Negatives | NEGP | ((binA))_NEGP ((kucha))_NP ((bole))_VG ((kAma))_L |
| 6 | Conjuncts | CCP | ((rAma))_NP ((Ora))_CCP ((SyAma))_NP |
| 7 | Chunk Fragments | FRAGP | rAma (jo merA baDZA bhAI hE) ne kahA... |
| 8 | Miscellaneous | BLK | |

****ocr The following table was created before the POS guidelines were incorporated. It comes from conversion data, I think. It only contains the POS tags that are different at DS and PS. All others are the same. We need also a table of all other POS tags.****

****ocr Need to reorder table****

| Description | DS and PB POS | XPOS | PS |
|--|---------------|-----------------------|----------------|
| Adjectives | JJ | [A <fs >] | A |
| Coordinating conjunction | CC | [CC-Coord <fs >] | CC |
| Subordinating conjunction | CC | [CC-Subord <fs >] | C |
| Determiner | DEM | [DET <fs >] | Det |
| General Quantifier | QF | [QF <fs >] | Q |
| Relative Quantifier (3 words) | QF | [DEG <fs >] | Deg |
| Empty causative head | EC | [EC <fs ectype=CAUS>] | V |
| Empty complementizer head | EC | [EC <fs ectype=COMP>] | C |
| Focus Marker | RP | [FOC <fs >] | Foc |
| Intensifier | INTF | [INTF <fs >] | Q |
| Common Noun | NN | [N <fs >] | N |
| Compound Common Noun | NNC | [NC <fs >] | NC |
| Pronoun with case marker | PRP | [N-P <fs >] | N+P |
| Pronoun with case marker Special case: sabase | QF | [N-P <fs >] | N+P |
| Proper Noun | NNP | [NProp <fs >] | NProp]] |
| Compound Proper Noun | NNPC | [NPropC <fs >] | NPropC |
| Negation marker | NEG | [NEG <fs >] | Neg |
| Spatio-Temporal Nouns | NST | [NST <fs >] | NST |
| Postposition | PSP | [PSP <fs >] | P |
| Cardinals | QC | [QC <fs >] | Num |
| Ordinal (??) | QO | [QC <fs >] | Num |
| Adverb | RB | [RB <fs >] | Adv |
| Partial Reduplication (prev. POS: π) | RDP | [π <fs >] | π |
| Partial Reduplication (prev. POS: π) | ECH | [π <fs >] | π |
| Punctuation and Symbols | SYM | [SYM <fs >] | PUNCT |
| Auxiliary | VAux | [VAux <fs >] | VAux |
| Passive Auxiliary | VAux | [V-Passive <fs >] | VAux |
| Aspectual Light Verb | VAux | [V-Vector <fs >] | VAux V+VAux |
| Verb | VM | [VM <fs >] | V |
| Non-finite verb | VM | [VM-NF <fs >] | V |
| Gerundive verb | VM | [VM-NN <fs >] | V |
| Nominal Question Word | WQ | [N <fs >] | N |
| Adverbial Question Word | WQ | [RB <fs >] | Adv |
| Determiner Question Word | WQ | [DET <fs >] | Det |
| Quantifier Question Word | WQ | [QF <fs >] | Q |
| Unknown | UNK | [UNK] | UNK |
| Interjection | INJ | [INJ] | INJ |
| Other Compounds | XC | | XC |

Part III

Introduction: Syntax

Chapter 4

Introduction to Syntactic Annotation

4.1 Background

These annotation guidelines of Hindi/Urdu are part of a multi-representational and multi-layered treebanking project [?, ?]. As part of this larger project, the following annotations are being made on top of corpora:

- A dependency annotation (DS) based on Paninian grammar. This annotation is done manually.
- An annotation of lexical predicate-argument structure (PropBank), in conjunction with the creation of a lexicon of verb meanings which has complete coverage for the annotated corpus. The lexicon creation and the annotation are done manually. The corpus annotation happens on top of the DS annotations.
- A phrase-structure annotation (PS) based on a theory following the principles-and-parameters methodology, and following the framework developed by Chomsky and his followers since the 1980s. This annotation will be derived automatically from DS and PropBank. Despite being derived automatically, it is carefully motivated and described in this manual.

Chapter 5

Introduction: Phrase Structure Annotation

5.1 General Background

Following the work in the Chomskyan tradition, our phrase structure does not just aim at representing surface structure, but it also aims at representing the underlying lexical predicate-argument structure. This is achieved in a monostratal representation by using traces to indicate the underlying position from which constituents moved; the underlying position represents the lexical predicate-argument structure through a uniform mapping from argument structure into specified positions in a canonical syntactic representation; from these specified positions arguments can then undergo movement to their surface position. Examples of syntax which involve such movement include passive, the unaccusative, and scrambling (these are not the same types of movement). We note that our approach allows for some arguments to be base generated in their final position; put differently, movement is not required in our representation.

We would like to contrast this Chomskyan approach with Lexical-Functional Grammar (LFG, [?]), in which the lexical predicate-argument structure and the surface syntax are represented in two different structures, the former (“f-structure”) as a dependency tree¹, and the latter (“c-structure”) as a phrase structure tree. We note that the the Chomskyan tradition and LFG aim for accounting for the same phenomena, but the different representational choices mean that the theories often look rather different. In the context of our annotation project, we note that the DS annotation is in some sense similar to LFG’s f-structure, and if we remove all traces (which is what much computational work in parsing does anyway), our PS structure is a plausible c-structure in LFG. Thus, our project can also be seen as producing an LFG-style treebank.

¹F-structure allows “re-entrancy” which creates a directed acyclic graph (dag); however, this is restricted to local contexts and the content of this structure can also be represented using features, thus preserving the arborescent nature of the f-structure.

5.2 Representational Issues

How flat/hierarchical should our trees be? One could argue that several aspects of hierarchical structure are very directly encoded in the linear order of arguments and that for this reason we should go for flatter trees. If we follow this line of thinking, for simple clauses, we would have trees that look pretty much like PDG trees.

But the DS treebank already exists and the phrase structure treebank should ideally provide a different way of looking at the data. With this in mind, we will go for binary branching as far as possible. This fits in well with several strains of linguistic theorizing and also provides a clear treebanking heuristic.

5.3 The Representation of Arguments

Instead of talking about “subject” and “object(s)” or making a traditional distinction between arguments and adjuncts, we identify and represent the arguments that are structurally identified. The syntax of Hindi provides clear diagnostics for two arguments that are structurally identified because they can agree with the verb, and appear in a clause without a postposition (or “case marking”). In addition, some verbs have a dative-marked argument, for which we introduce a privileged position. Finally, we add a fourth position, which however is rarely made explicit in the phrase structure: a lower position from which the internal arguments originate.

****ocr Need to discuss and rewrite as needed:*

- *Our view: 3 types of dependents:*

+ *first and second position: structurally defined, always there, may be filled by verb; if verb does not fill (true unergatives), they are there and empty*

+ *dative: structurally defined, but only there when filled by verb; same with NP genitives, and degree adverbials in APs (check this latter claim)*

+ *all others: not structurally defined and only there when filled by verb*

+ *Special case: stative be, which does not have the 2nd position*

- *We need to define internal and external argument.*

v- Is there a difference between a “privileged position” and a “structurally identified” position? We know there are differences between the positions in terms of what happens there, but are the positions intrinsically different?

- *I want to get rid of all instances of “structurally licensed” and change them to “structurally privileged”. Ok?*

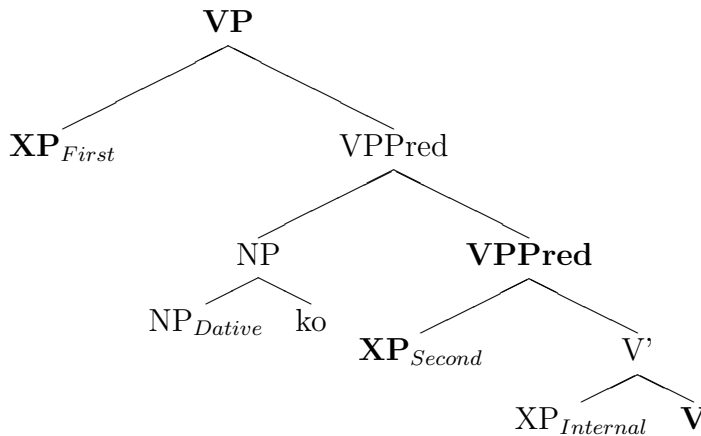
- *Related: are there 2 or 3 structurally identified positions?*

- *What is underlying model of how syntax relates to lex pred-arg structure? (A minor question...)*

Going up the projection from the verb, we have the following four positions.

- The Internal Argument Position. This is the position in which internal arguments originate, and certain clausal arguments can remain here (small clauses). This position is only represented in phrase structure if it is filled. In all standard intransitive and transitive clauses, this position is not filled with overt material and is not represented in the annotation. However, in some sense, it is always present.
- The Second Position. This is the lower of the two positions which can show agreement with the verb. In a typical transitive sentence, the internal argument (“object”) ends up in this position.
- The Dative Position. This is the position in which the dative postposition *-ko* is assigned, both in ditransitive and in dative subject constructions. The dative subjects almost always scramble leftward. This position is only present if there is a dative argument. We assume that if a verb has no dative argument, then this position is not present at all.
- The First Position. This is the higher of the two positions which can show agreement with the verb. In a typical transitive sentence, the external argument (“subject”) is in this position. It is the position of ergative-marked arguments.

The phrase structure is as follows. The elements of the structure which are always present are in **bold face**.



All other XPs will be treated like syntactic adjuncts, i.e., XPs which are not structurally licensed. In this document, we will use “in a structurally privileged position” and “syntactic argument” interchangeably, and “not structurally licensed” and “syntactic adjunct” as well. *Par abus de langage*, we will sometimes omit the word “syntactic” when using “argument” or “adjunct”, but in this document, these terms (*argument* and *adjunct*) do not have

their standard lexical-semantic meaning, but a syntactic meaning specific to our analysis Hindi/Urdu.

We note that one of the reasons we can allow ourselves to ignore the traditional meaning of “subject”, “object”, and “argument” is that this project includes, from the beginning, PropBanking. This means that any practical concerns about the absence of traditional argument labeling and argument/adjunct distinction will be handled at a lexical-semantic level, for which a lexicographic methodology has been elaborated.

5.3.1 Background: The Representation of Arguments

In any Hindi clause, there are at most two structurally licensed arguments. The higher one corresponds in many cases to the notion ‘subject’ and the lower one to the notion ‘object’. But this is not a definitional property. The structural licensing will be represented using node labels (V, V’, VP-PRED, VP), and thus no dashtags will be needed to identify these two arguments (just as in the English PTB the syntactic object is not marked with a dashtag). In the diagrams in this document we use subscripts for clarity, but they are not part of the proposed annotation.

Subject and Objects

The first decision that we will need to make is whether subjects occupy a distinguished position in syntax. The problem is that there is no very clear subject diagnostic in Hindi. There are a number of diagnostics and they diverge from each other in the tricky cases.

(1) Diagnostics for subjecthood:

a. initial position in default word order

This diagnostic is actually rather useful despite its rather squishy nature. We could just represent divergences from the ‘default’ word order independent of the notion of subjecthood.

b. nominative case [unmarked case licensed by T]

Problem: Ergative subjects are subjects by any one’s estimation but they are not nominative. Moreover in ‘dative subject’ constructions, there are nominative arguments that are not particularly subject like.

c. agreement trigger: in many languages, the verb agrees with the subject:

Problem: existence of object agreement.

d. binding of reflexives etc.

Problem: in Dative Subject constructions, more than one argument can bind reflexives.

e. control

Unlike in Icelandic, in Hindi dative subjects cannot be controlled. Only nominative arguments can become big PRO.

We take the above conflicting diagnostics to show that the subject/object distinction that is familiar to us from English is not very revealing when it comes to Hindi.

Structurally Licensed Arguments and Others

Instead of talking about “subject” and “object(s)” or making a traditional distinction between arguments and adjuncts, we will identify and represent the arguments that are structurally licensed. The syntax of Hindi provides clear diagnostics for up to two arguments that are structurally licensed and we will use these diagnostics. All other XPs will be treated like syntactic adjuncts, i.e., XPs which are not structurally licensed. In this document, we will use “structurally licensed” and “syntactic argument” interchangeably, and “not structurally licensed” and “syntactic adjunct” as well. *Par abus de langage*, we will sometimes omit the word “syntactic” when using “argument” or “adjunct”, but in this document, these terms (*argument* and *adjunct*) do not have their standard lexical-semantic meaning, but a syntactic meaning specific to Hindi/Urdu. The distinction may lead to some initially counter-intuitive results,

We note that one of the reasons we can allow ourselves to ignore the traditional meaning of “subject” and “object” is that this project includes, from the beginning, PropBanking. This means that any practical concerns about the absence of traditional argument labeling and argument/adjunct distinction will be handled at a lexical-semantic level, for which a lexicographic methodology has been elaborated.

In any Hindi clause, there are at most two structurally licensed arguments. The higher one corresponds in many cases to the notion ‘subject’ and the lower one to the notion ‘object’. But this is not a definitional property. The structural licensing will be represented using node labels (V, V’, VP), and thus no dashtags will be needed to identify these two arguments (just as in the English PTB the syntactic object is not marked with a dashtag). In the diagrams in this document we use subscripts for clarity, but they are not part of the proposed annotation.

Agreement

Under the right circumstances, the low argument can trigger agreement. Since only structurally represented arguments can ever agree, this is a powerful and reliable test and we can use it to distinguish real low arguments from dependents that resemble them externally.

(2) with transitives:

a. (K2Lower-Argument-PS-5) ergative subject, object agreement

| | | |
|-----------|---------------|------------|
| Atif-ne | kitaab | paṛh-ii |
| AwiPZA ne | kiwAba | paDZI |
| Atif-Erg | book.f | read-Pfv.f |

‘Atif read the book.’

- b. (K2Lower-Argument-PS-6) dative subject, object agreement

Atif-ko **kitaab** parh-nii hai
 AwiPZa ko kiwAba paDZanI hE
 Atif-Dat book.f read-Inf.f be.Prs.Sg

‘Atif has to read (a/the) book.’

This test comes in handy with unaccusative verbs of motion which often appear with an unmarked location argument. The agreement test reveals that this argument is not a structurally licensed low argument. It does not trigger agreement.

(3) Unaccusative Motion Verbs

- a. (K2Lower-Argument-PS-7)

Atif *Dilli* jaa rahaa hai
 AwiPZa xillI jA rahA hE
 Atif.M Delhi.f go Prog.MSg be.Prs.Sg

‘Atif is going to Delhi.’

- b. (K2Lower-Argument-PS-8) *Dilli* does not trigger agreement:

Atif-ko *Dilli* jaa-naa thaa
 AwiPZa ko xillI jAnA WA
 Atif-Dat Delhi.f go-Inf/-Inf.F be.Pst/be.Pst.F

‘Atif has to go to Delhi.’

5.4 Summary of Phrase Markers Used in PS

We have a single type of dashtags, where a “dashtag” is a tag added to the main symbol (or to another dashtag) after a dash or a plus sign. This dashtag is “-Pred”, which attaches to a variety of XPs to indicate that it takes a subject and projects further to an SC.

We also use the plus sign ‘+’ (as in N+P and V+VAux), which indicates that two words which are usually spelled separately have been spelled as one word (as preferred by the Government of India).

The following is a list of projections. This list covers all nonterminal symbols used in the PS part of the Hindi and Urdu Treebank.

| Type | POS | Intermediate | | Maximal |
|----------------------|------------------------------------|--------------|-----------|----------------------|
| Nominal | N, NC, N+P NProp, NPropC | — | — | NP |
| | N, NC, N+P NProp, NPropC | — | NP-Pred | SC |
| | NST | — | — | NSTP |
| | | — | NSTP-Pred | SC |
| P | — | — | — | |
| Verbal | V, V+VAux | V' | VPPred | VP |
| | V (with <i>-naa</i> morphology) | V' | VPPred | NP See note below |
| | VAux | — | — | — |
| Adjective | A | — | — | AP |
| | | — | AP-Pred | SC |
| Adverb | <i>check with DS</i> | | | AdvP |
| Complementizer | C | — | — | CP |
| Conjunction | CC | — | — | CCP |
| Determiner | D | — | — | — |
| Degree Expression | Deg | — | — | DegP |
| | | — | DegP-Pred | SC |
| Demonstrative | Dem | — | — | — |
| Focus Particle | Foc | — | — | — |
| Negation Maker | Neg | — | — | — |
| Numeral | Num | — | — | NumP |
| | | — | NumP-Pred | SC |
| Participle | Part | — | — | — |
| Compound category | XC | — | — | — |
| Quantifier | Q | — | — | — |
| Punctuation Mark | PUNCT | — | — | PUNCTP |

Note: In the case of a verb projecting to NP, the higher argument is genitive or PRO.

Note: For XC, X can be any preterminal. It always adjoins to (but does not project to) XP.

Note: PUNCTP is used when a punctuation mark is used as a head; it represents an extra-synactic analysis.

Explanations:

- XC: The X in the label XC is a variable that is replaced in the actual annotation by the relevant part of speech tag. As such, the label for non-final words in non-proper compound nouns is NC , the label for non-final parts of compound proper name parts is $NPropC$.

****ocr Is the above still correct?****

Chapter 6

Comparison Between the Levels of Representation

This is a comparison. To be based on existing papers. In charge: tbd.

Part IV

Verbs and Their Arguments

Chapter 7

Transitive Verbs

7.1 The First Argument

The first argument will always be represented in the [Spec,V] position i.e. [VP NP_{First} VP-PRED].

7.1.1 Simple Transitives

The higher structurally licensed argument can be nominative or dative, and for transitive verbs, ergative.

- (4) Transitive:
- a. (Transitive-Verbs-PS-1) Nominative:
Atif kitaab paḥ-egaa
AwiPZa kiwAba paDZegA
Atif.M book.f read-Fut.3MSg
'Atif will read (a/the) book.'
 - b. (Transitive-Verbs-PS-2) Dative:
Atif ko kitaab paḥ-nii hai
AwiPZa ko kiwAba paDZanI hE
Atif Dat book.f read-Inf.f be.Prs.Sg
'Atif has to read (a/the) book.'
 - c. (Transitive-Verbs-PS-3) Ergative:
Atif ne kitaab paḥ-ii
AwiPZa ne kiwAba paDZI
Atif Erg book.f read-Pfv.F
'Atif read (a/the) book.'

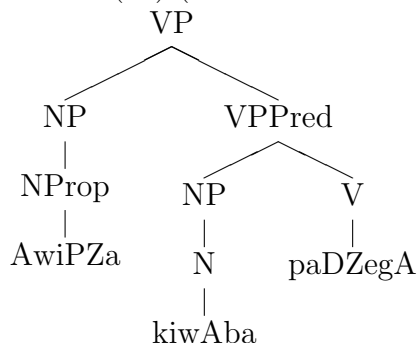
The case marker in the second position is not relevant. In all of the above examples, the second position has unmarked case but the object could very well have had

accusative *-ko*.

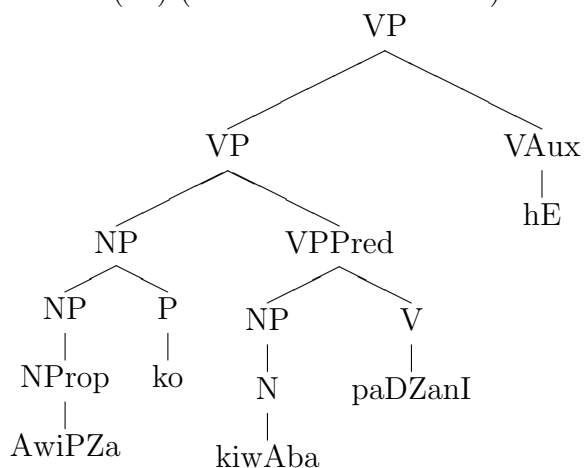
Trees:

(5) Transitive:

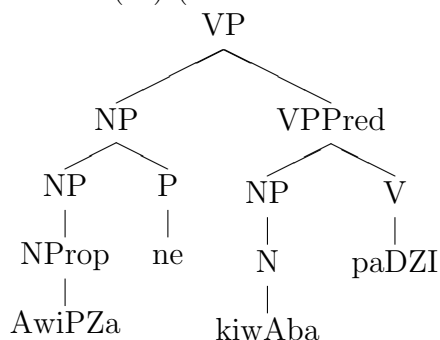
a. Tree for (4a) (Transitive-Verbs-PS-1) Nominative:



b. Tree for (4b) (Transitive-Verbs-PS-2) Nominative:



c. Tree for (4c) (Transitive-Verbs-PS-3) Nominative:



7.2 The Second Position

The issue of a second position only arises with two classes of verbs: transitives and ditransitives. The other classes have only one structurally licensed argument.

The possible case markers for the second position are only ϕ or *-ko*.

In this section, we discuss transitives to illustrate the two case marking options on the second position.

7.2.1 Transitives

These examples concentrate on the realization of the second argument. More examples can be found in Section 7.1.1.

(6) Transitive:

a. (K2Lower-Argument-PS-1) without *-ko*:

Atif **kitaab** uṭhaa rahaa hai
 AwiPZa kiwAba uTA rahA hE
 Atif.M book.f lift Prog.MSg be.Prs.Sg

‘Atif is picking up (a/the) book.’

b. (K2Lower-Argument-PS-2) with *-ko*:

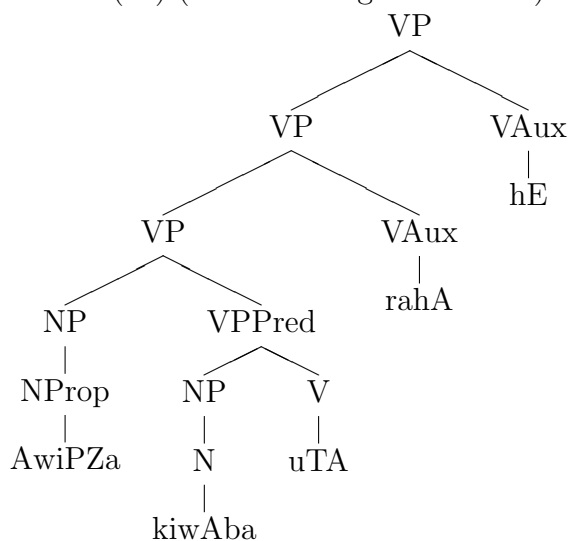
Atif **kitaab-ko** uṭhaa rahaa hai
 AwiPZa kiwAba ko uTA rahA hE
 Atif.M book.f-KO lift Prog.MSg be.Prs.Sg

‘Atif is picking up the book.’

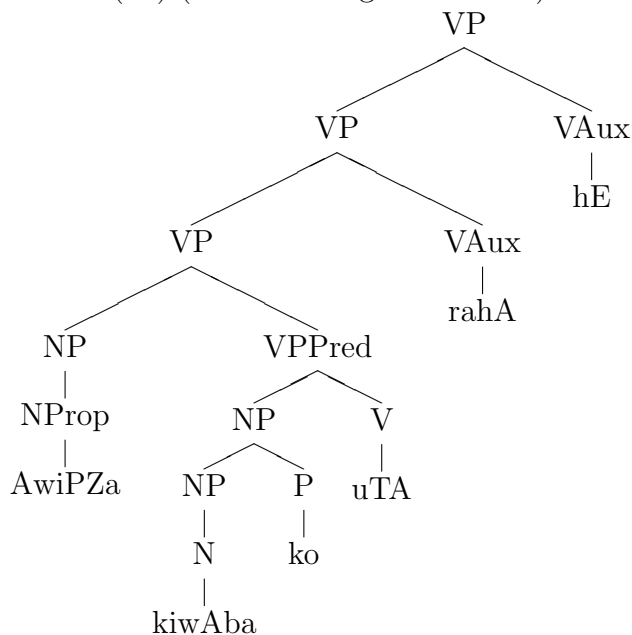
Trees:

(7) Transitive:

a. Tree for (6a) (K2Lower-Argument-PS-1) Nominative:



- b. Tree for (6b) (K2Lower-Argument-PS-2) Nominative:



7.3 Background

7.3.1 Background: the First Argument

General Diagnostics for the Higher Argument

We use the term “dependent” as a generic term for syntactic arguments and adjuncts. Case Marking:

- (8) If the case-marking is anything other than *-ne* ‘ERG’, *-ko* ‘DAT’, or $-\phi$, the argument cannot be the higher argument.

Ergative Marking:

- (9) If a dependent can take the ergative marker *-ne*, it is the higher argument.
- only useful in transitives and ditransitives
 - if we are testing transitives/ditransitives in a non-perfective aspect, we need to change the aspect to perfective and remove any intransitive light verbs.

Dative of Obligation:

- (10) If a dependent can appear bearing the dative of obligation, it is the highest argument.
- (K1Higher-Argument-PS-1)

Arif Dilli jaa rahaa hai
 AriPZa xilli jA rahA hE
 Arif.m Delhi go Prog.MSg be.Prs.Sg
 ‘Arif is going to Delhi.’

- b. (K1Higher-Argument-PS-2)

Arif ko Dilli jaa-naa hai
 AriPZa ko xilli jAnA hE
 Arif Dat Delhi go-Inf be.Prs.Sg
 ‘Arif has to go to Delhi.’

- (11) However, certain non-referential dependents cannot (happily) take the dative of obligation even when they are the highest argument. For example, the NP’s in NP+be constructions.

- a. (K1Higher-Argument-PS-3)

[_{NP}**mandir kaa udghaatan**] kal ho-gaa
 manxira kA uxGAtana kala hogA
 temple Gen inauguration yesterday be-Fut.MSg
 ‘The temple will be inaugurated tomorrow.’

- b. (K1Higher-Argument-PS-4)

??[_{NP}**mandir ke udghaatan**] ko kal ho-naa hai
 manxira ke uxGAtana ko kala honA hE
 temple Gen.Obl inauguration Dat tomorrow be-Inf be.Prs.Sg
 ‘The temple is to be inaugurated tomorrow.’

(This confound is limited to subjects of unaccusatives. Subjects of unergatives and transitives can always take the dative of obligation.)

Non-Finite Contexts:

- (12) The highest argument (typically) goes to zero in non-finite contexts. When zero, it can be controlled by an argument of an embedding verb. One way of realizing the highest argument overtly in non-controlled non-finite contexts is to mark it with the genitive *-kaa*.

- (13) infinitival subjects

- a. (K1Higher-Argument-PS-5) transitives: (ditransitives behave the same)

[ϕ /**Atif-kaa**/***Atif** kitaab parh-naa] acchii baat hai
 AwIPZa kA kiwAba padZanA acCI bAwa hE
 ϕ /Atif-Gen/Atif book read-Inf goodf .thing.f be.Prs.Sg
 ‘(Atif’s) reading books is a good thing.’

- b. (K1Higher-Argument-PS-6) unergatives:

[ϕ /Atif-kaa/*Atif roz nahaa-naa] acchii baat hai
 AwiPZa kA rojZa nahAnA acCI bAwa hE
 ϕ /Atif-Gen/Atif daily bathe-Inf good.f thing.f be.Prs.Sg

‘(Atif’s) bathing daily is a good thing.’

- c. (K1Higher-Argument-PS-7) unaccusatives, motion verb:

[ϕ /Atif-kaa/*Atif roz yahā: aa-naa] acchii baat hai
 AwiPZa kA rojZa yahAz AnA acCI bAwa hE
 ϕ /Atif-Gen/Atif daily here come-Inf good.f thing.f be.Prs.Sg

‘(Atif’s) coming here everyday is a good thing.’

- d. (K1Higher-Argument-PS-8) dative subject:

[ϕ /tumhaaraa/*tum boss-ko office-me dikh-naa] acchii
 wumhArA bOYsa ko OYPisa meM xiKanA acCI
 ϕ /you-Gen/you boss-Dat office-in see-Inf good.f thing.f be.Prs.Sg
 baat hai
 bAwa hE

‘(Your) being seen in the office by the boss is a good thing.’

- (14) Unaccusative verbs from the Transitivity Alternation class show some complications with this test. The genitive is fine but zero is odd, presumably because the relevant position is typically inanimate and so is incompatible with the animate PRO_{arb} that the zero receives. The additional complication is that the subject of the unaccusative when inanimate can surface in the unmarked case.

- a. (K1Higher-Argument-PS-9) genitive ok:

[pattō-kaa gir-naa] acchii baat hai
 pawwoM kA giranA acCI bAwa hE
 leaves.Obl-Gen fall-Inf good.f thing be.Prs.Sg

‘The falling of leaves is a good thing.’

- b. (K1Higher-Argument-PS-10) zero is a bit marked:

??[ϕ gir-naa] acchii baat hai
 giranA acCI bAwa hE
 ϕ fall-Inf good.f thing be.Prs.Sg

‘Falling is a good thing.’

- c. (K1Higher-Argument-PS-11) unmarked case is also good!:

[patte gir-naa] acchii baat hai
 pawwe giranA acCI bAwa hE
 leaves fall-Inf good.f thing be.Prs.Sg

‘Leaves falling is a good thing.’

The conclusion then is that when we are doing the ‘argument goes to zero’ test to test for the higher argument, we should use a proper name/pronominal subject if possible. Also if an infinitive allows for unmarked subjects, we know that the predicate is unaccusative.

Control Structures:

- (15) The high argument is the one that is always controlled. The examples involve complement control but the facts also hold for adjunct control.
- a. (K1Higher-Argument-PS-12) Transitives: (same for ditransitives)
- | | | | | | |
|--------|----------|--------|-----------|--------------|-----------|
| Atif | [ϕ | kitaab | paṛh-naa] | caah-taa | hai |
| AwiPZa | NULL | kiwAba | paDZanA | cAhawA | hE |
| Atif.M | | book | read-Inf | want-Hab.MSg | be.Prs.Sg |
- ‘Atif wants to read (a/the) book.’
- b. (K1Higher-Argument-PS-13) Unergatives: (same for motion verb unaccusatives)
- | | | | | |
|--------|----------|------------|--------------|-----------|
| Atif | [ϕ | nahaa-naa] | caah-taa | hai |
| AwiPZa | NULL | nahAnA | CAhawA | hE |
| Atif.M | | bathe-Inf | want-Hab.MSg | be.Prs.Sg |
- ‘Atif wants to bathe.’
- c. (K1Higher-Argument-PS-14) Dative Subject Unaccusatives:
- | | | | | | |
|--------|------------------|--------|-----------|-----------|-------------------|
| Atif | [Mona-ko | ϕ | bazaar-me | dikh-naa] | caah-taa |
| AwiPZa | monA | ko | NULL | bAjZARA | meM xiKanA cAhAwA |
| Atif.M | Mona-Dat | | market-in | see-Inf | want-Hab.MSg |
- hai
hE
be.Prs.Sg
- ‘Atif wants to be seen by Mona in the market.’
- d. (K1Higher-Argument-PS-15) Transitivity Alternation Unaccusatives: (this can be a bit tricky because the high argument of these verbs is often inanimate and hence the control examples come out pragmatically odd.)
- | | | | | |
|---------------|----------|----------|--------------|-----------|
| Atif/per | [ϕ | gir-naa] | caah-taa | hai |
| Atif.M/tree.M | | fall-Inf | want-Hab.MSg | be.Prs.Sg |
- ‘Atif/the tree wants to fall.’

7.3.2 Background: the Lower Argument

- (16) Possible Cases: only ϕ or *-ko*
- a. - but sometimes a dependent which is not the lower argument can be ϕ (directional arguments of motion verb unaccusatives)
- b. - and a dependent which is not the lower argument NPs can be *-ko* (datives and certain directional/purpose adjuncts)

The *ko*-alternation

- (17) Inanimate NP objects of transitives can typically appear with or without *-ko*. The absence/presence of *-ko* has an effect on the specificity/definiteness of the object.

When this alternation is possible, we have a pretty good test for the lower argument.

- a. (K2Lower-Argument-PS-1) without *-ko*:

Atif **kitaab** uṭḥaa rahaa hai
 AwiPZa kiwAba uTA rahA hE
 Atif.M book.f lift Prog.MSg be.Prs.Sg

‘Atif is picking up (a/the) book.’

- b. (K2Lower-Argument-PS-2) with *-ko*:

Atif **kitaab-ko** uṭḥaa rahaa hai
 AwiPZa kiwAba ko uTA rahA hE
 Atif.M book.f-KO lift Prog.MSg be.Prs.Sg

‘Atif is picking up the book.’

To test for this alternation, one needs to try objects that are not proper names or pronouns because they always require *-ko*. One can use this test also with ditransitives but it is a little more involved and not all speakers like the resulting instances with two *-ko*'s.

- (18) a. (K2Lower-Argument-PS-3) without *-ko*:

Atif Mona-ko kitaab de rahaa hai
 AwiPZa monA ko kiwAba xe rahA hE
 Atif.M Mona-Dat book.f give Prog.MSg be.Prs.Sg

‘Atif is giving a book to Mona.’

- b. (K2Lower-Argument-PS-4) with *ko*, obligatory scrambling of object over dative:

Atif mujhe/kitaab-ko Mona-ko de rahaa hai
 AwiPZa muJe/kiwAba ko se rahA hE
 Atif me.Acc/book-KO Mona-Dat give Prog.MSg be.Prs.Sg

‘Atif is giving me/the book to Mona.’

Chapter 8

Intransitives

8.1 Unergatives

Unergative verbs are, very roughly speaking, simple transitives (verbs with only one argument) in which the first position plays a traditional first position role, such as agent. They can take the same cases as the higher positions of transitive verbs (nominative, ergative, or dative of obligation); they simply lack a lower position.

(19) Unergatives:

a. (Unergatives-PS-1) Nominative:

Atif nahaa-egaa
AwiPZa nahAegA
Atif.M bathe-Fut.3MSg
'Atif will bathe.'

b. (Unergatives-PS-2) Dative:

Atif-ko nahaa-naa hai
AwiPZa ko nahAnA hE
Atif-Dat bathe-Inf be.Prs
'Atif has to bathe.'

c. (Unergatives-PS-3) Ergative:

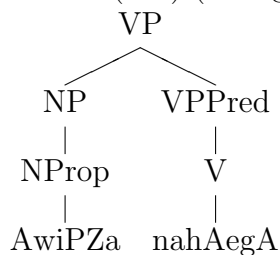
Atif-ne nahaa liyaa
AwiPZa ne nahA liyA
Atif-Erg bathe TAKE.Pfv
'Atif has bathed.'

For unergatives, we include the VPPRED node, but it has only one daughter node, the V node.

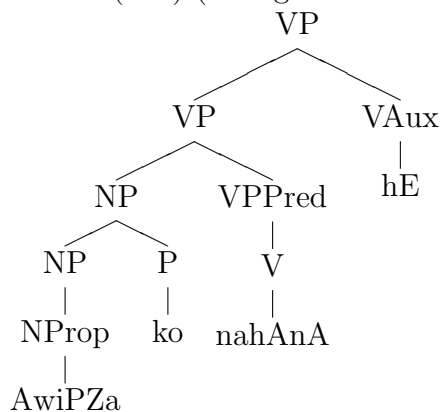
Trees:

(20) Transitive:

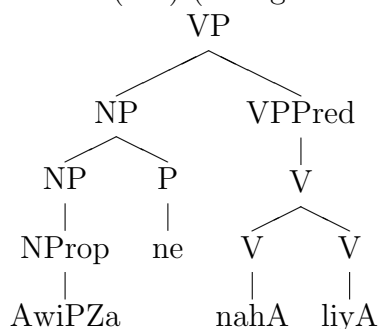
- a. Tree for (19a) (Unergatives-PS-1) Nominative:



- b. Tree for (19b) (Unergatives-PS-2) Nominative:



- c. Tree for (19c) (Unergatives-PS-3) Nominative:



8.2 Unaccusatives

Unaccusative verbs are intransitive verbs (verbs with only one argument) in which, roughly speaking, the first argument has semantic properties usually associated with lower arguments. Ergative case is not an option with unaccusatives, but nominative and obligational dative cases are possible.

- (21) Unaccusatives:

- a. (Unaccusatives-PS-1) Nominative:

darwaazaa khul rahaa hai
 xaravAjZA Kula rahA hE
 door.M open Prog.MSg be.Prs.Sg

‘The door is opening.’

b. (Unaccusatives-PS-2) Dative:

darwaaze-ko khul-naa hai
 xaravAjZe ko KulanA hE
 door-Dat open-Inf be.Prs

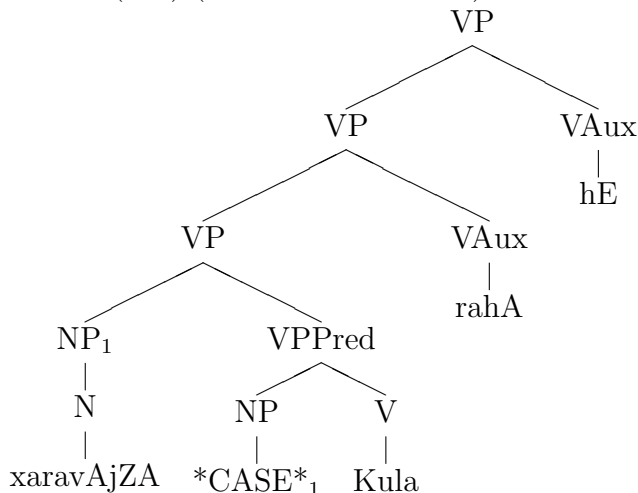
‘The door has to open.’

In the trees for unaccusatives, we insert a special trace in the second position, representing the fact that semantically, the first position comes from somewhere below.

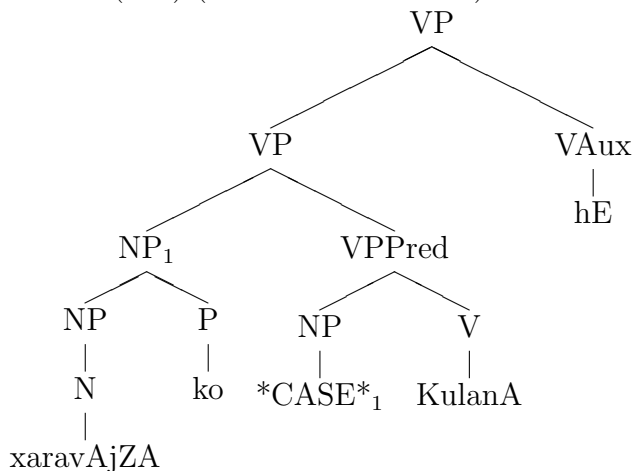
Trees:

(22) Transitive:

a. Tree for (21a) (Unaccusatives-PS-1) Nominative:



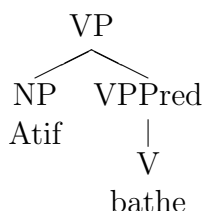
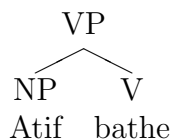
b. Tree for (21b) (Unaccusatives-PS-2) Dative:



8.3 Background

Unergatives

For unergatives, two possible representations are possible.

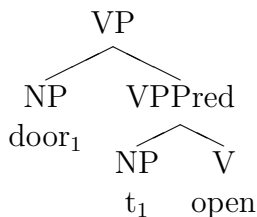
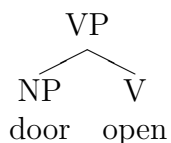


We adopt the second representation because it allows us to have adjuncts between the higher argument and the verb by adjoining them at the V' node, as in transitive verbs. (If the unergative/unaccusative distinction were not available to us in the PDG representation, we would go with the second structure and collapse the distinction between the two.)

Our representation of unergatives thus constitutes the major exception to the general restriction against non-binary branching structures.

Unaccusatives

As was the case with unergatives, two possible representations are possible for unaccusatives:



Again, we adopt the second representation because it allows us to have adjuncts between the higher argument and the verb by adjoining them at the V' node, as in transitive verbs. (If the unergative/unaccusative distinction were not available to us in the PDG representation, we would go with the second structure and collapse the distinction between the two.)

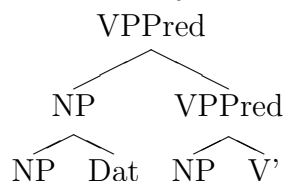
Chapter 9

Dative Constructions

9.1 General Observations on Datives

Lexically selected dative arguments are found with ditransitives and with dative subject constructions. Even though dative arguments are neither the first argument nor the second argument, they display a number of argument properties such as the ability to bind reciprocals/reflexives and the ability to control into certain infinitival clauses. For this reason, we assign them a distinguished structural position.

(23) NP-Dat is adjoined to the minimal VPPred:



We assume that the dative goal in a ditransitive and the experiencer dative in a dative subject construction are introduced at the location, adjoined to the minimal VPPred.

One implication of this treatment is that displacement of these dative arguments will leave traces. In contrast, displacement of adjuncts does not leave traces.

9.2 Ditransitives

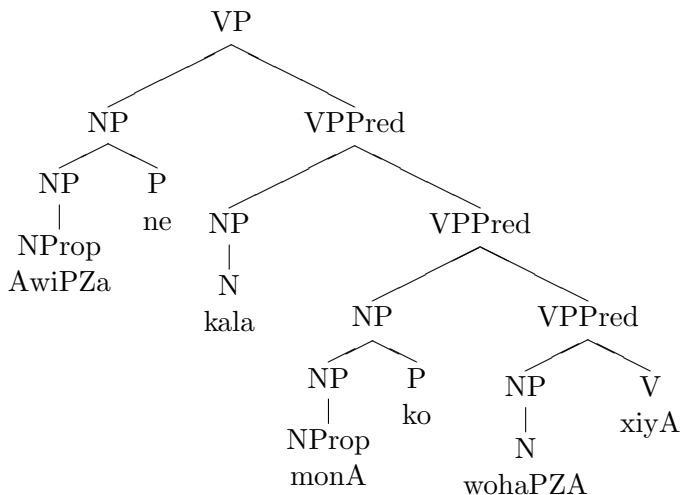
Let us first consider ditransitives. Their treatment is relatively straightforward. The dative argument is adjoined to the minimal VPPred. The second example displays how local scrambling of the dative goal leaves a trace.

(24) a. (Label:Dat-Goal-PS-1)

Atif ne kal Mona ko tohfaa diyaa
 AwiPZa ne kala monA ko wohAPZA xiyA
 Atif Erg yesterday Mona Dat present give.Pfv.MSg

‘Atif gave a present to Mona yesterday.’

Tree for (24a) (Dat-Goal-PS-1):

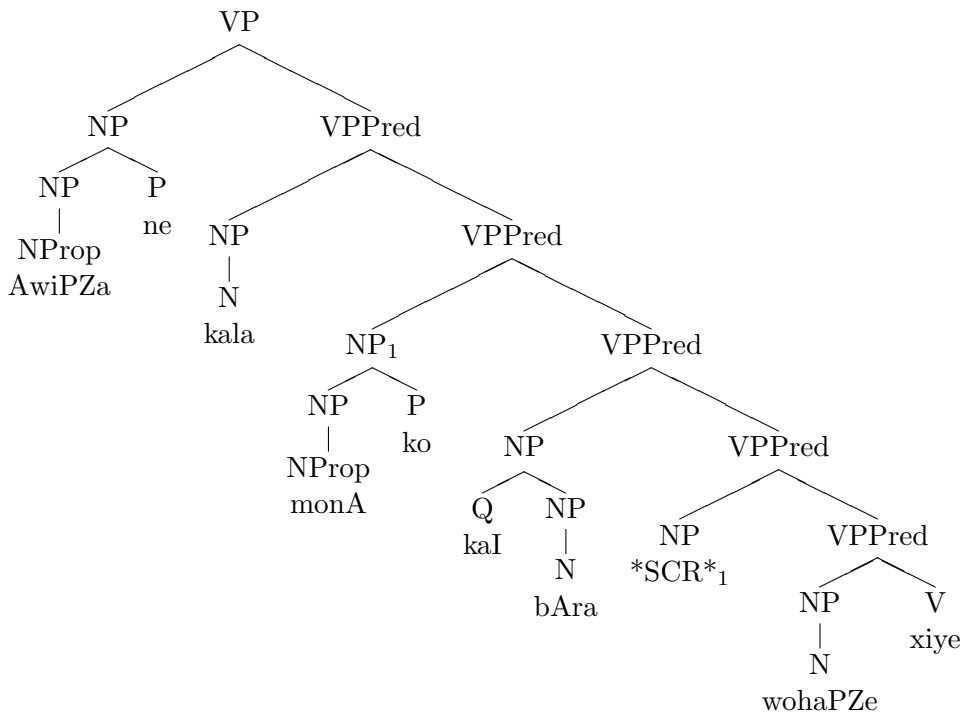


b. (Label:Dat-Goal-PS-2)

Atif ne kal Mona ko kai baar tohfe diye
 AwiPZa ne kala monA ko kaI bAra wohAPZe xiye
 Atif Erg yesterday Mona Dat many times present give.Pfv.MSg

‘Atif gave a present to Mona many times yesterday.’

Tree for (24b) (Dat-Goal-PS-2):



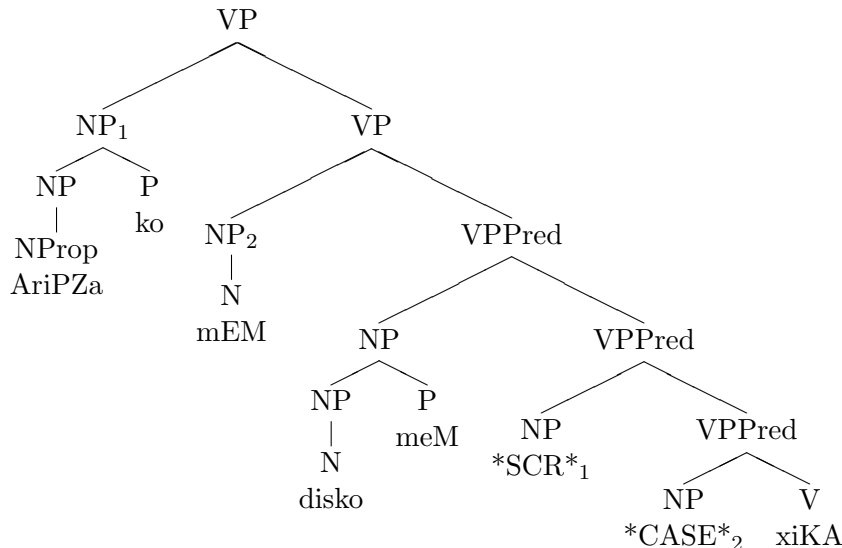
ocr *ocr if we had a name for dative position, we would not need to scramble*****
 ***ocr ***ocr there are no verbs with dative subject and dative goal, obviously
 – maybe say?*****

9.3 Dative Subject Constructions

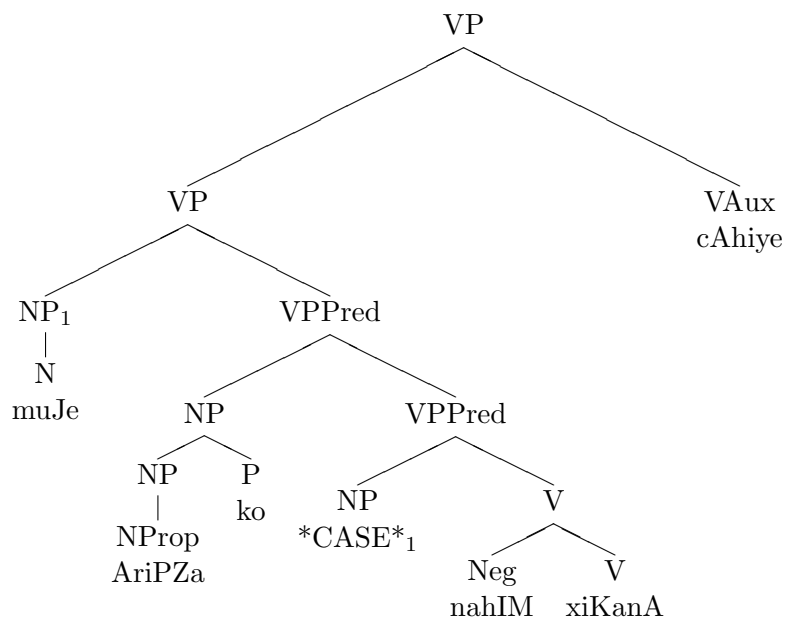
Dative subject constructions raise some interesting issues. They have two arguments, one of which is a lexically licensed dative and the other a structurally licensed nominative. The nominative possesses all the properties of the first argument but the dative subject also has a number of subject-like properties.

As with the dative goal, we assume that the dative argument is introduced adjoined to the VPPred. The nominative argument originates as the internal argument and raises to the high position. The dative subject can optionally scramble to the initial position.

- (25) a. (Label:Dative-Subjects-PS-1)
 Arif ko **mĒ** disco me dikh-aa
 AriPZa ko mEM disko meM xiKA
 Arif Dat I disco in see-Pfv.MSg
 ‘Arif saw me in the disco.’
 Tree for (25a) (Dative-Subjects-PS-1)



- b. (Label:Dative-Subjects-PS-2)
mujhe Arif ko nahī: dikh-naa caahiye
 muJe AriPZa ko nahIM xiKanA cAhiye
 I.Dat Arif Dat Neg see-Inf should
 ‘Arif shouldn’t see me .’
 Tree for (25b) (Dative-Subjects-PS-2)



Interestingly in a dative subject construction, the dative can bind into the nominative and vice versa. This unusual property makes sense given our representation as within this representation, internal to the minimal VP, the dative NP c-commands the nominative NP and vice versa.

ocr ***ocr do you mean the dative c-commands the first position after scrambling? this requires a bit of discussion about scrambling, no?**

Chapter 10

Clausal Complementation

Clausal complementation in Hindi can involve finite *ki* clauses or non-finite *-naa* clauses. *ki*-clauses are always extraposed while *-naa* clauses typically appear pre-verbally.

10.1 Non-Finite Complements

All non-finite clausal complements (as opposed to small clause complements; see Section 12) are infinitival in Hindi (as opposed to being participial etc.). Infinitival clauses are based on the *-naa* form of the verb. In Hindi, infinitives are systematically ambiguous with gerunds. We treat a *-naa* clause with an overt genitive subject as a gerund which are taken to be verbal projections that project to NP. All other *-naa* clauses are treated as ordinary VPs.

The infinitival complements under discussion in this section all have silent subjects. We systematically assume that the silent subject of an infinitival clause is realized by big PRO. We distinguish PRO from little pro because the former is obligatory and restricted to the higher position, while the latter is optional and available for any argument. PRO is controlled by its controller; pro is not (in general).

(26) Non-finite clauses:

- a. (Non-Finite-Complement-Clauses-Control-PS-1) Typically infinitival complements of verbs

Ram ne [ghar jaa-naa] caah-aa
rAma ne Gara jAnA cAhA
Ram Erg home go-Inf want-Pfv

‘Ram wanted to go home.’

- b. (Non-Finite-Complement-Clauses-Control-PS-2) Also infinitival complements of the N in an N-V compound:

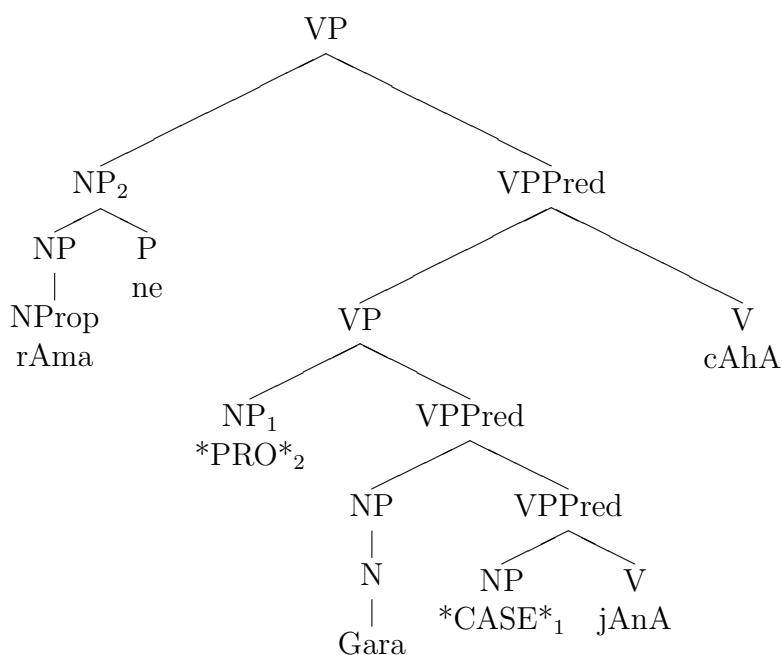
Ram ne [[ghar jAne kii] koshish] kii
rAma ne Gara jAne kI koSiSa kI
Ram Erg home go-Inf Gen attempt.f do.Pfv.f

‘Ram tried to go home.’ (there is also an unaccusative variant of this structure with *koSiSa + be*.)

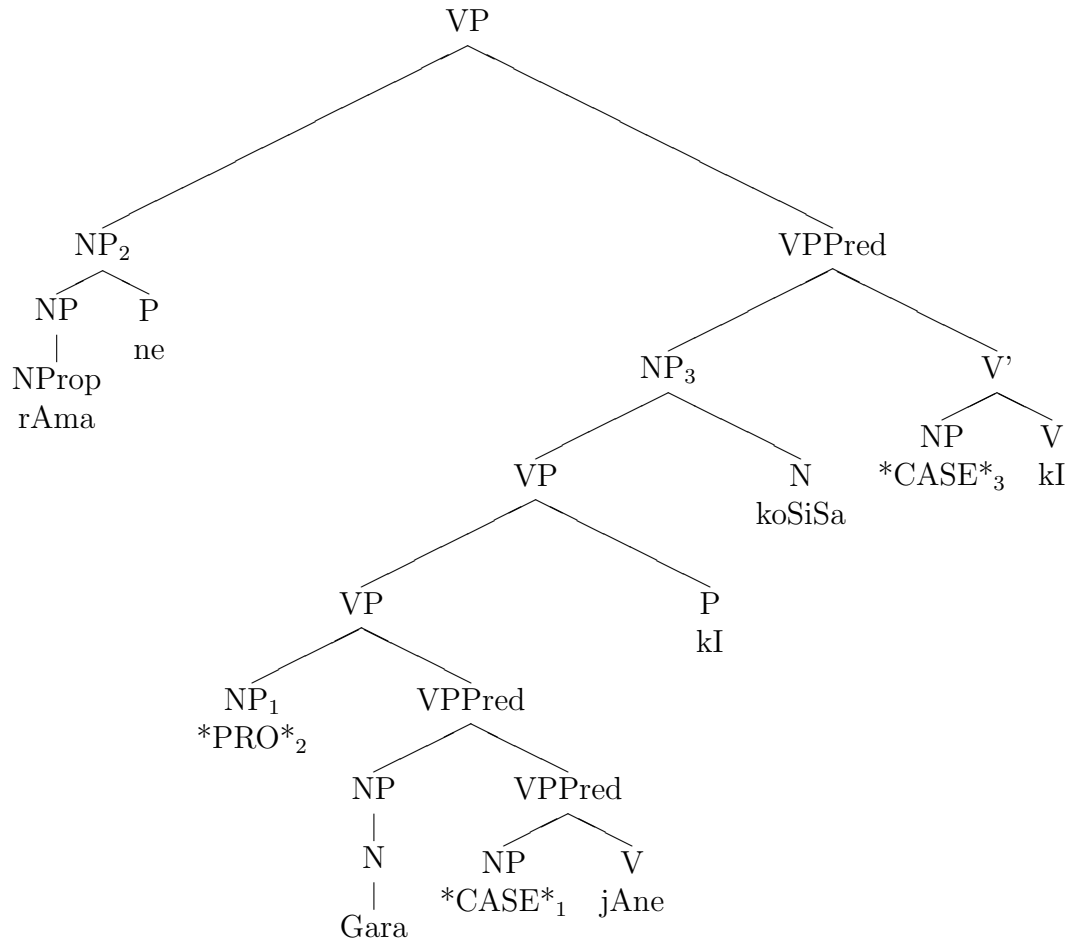
The analysis of (26a) is entirely analogous to similar constructions in other languages. (26b) is more complex. We have a light verb construction (see Section 20), but the clausal argument of the base noun is realized as an infinitival *-naa* clause that combines with a genitive marker.

(27) Non-finite clauses:

- a. Tree for (26a) (Non-Finite-Complement-Clauses-Control-PS-1) Typically infinitival complements of verbs

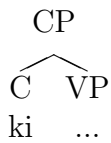


- b. Tree for (26b) (Non-Finite-Complement-Clauses-Control-PS-2)



10.2 Finite Complements and Extraposition

We will treat *ki*-clauses as CPs.



CPs extrapose obligatorily, with an optional co-referential pronoun in the main clause. Cps can also appear together with a full NP in the main clause. When such an NP is present, extraposition is not obligatory.

- (28) a. (Finite-complements-PS-1a) Finite clauses without pronoun , always extraposed
 Ram jaan-taa hai [ki Sita der se aa-egii]
 rAma jAnawA hE ki sIwA xera se AegI
 Ram know-Hab be.Prs.Sg that Sita delay with come-Fut.FSg
 ‘Ram knows that Sita will come late.’

- b. (Finite-complements-PS-1b) Finite clauses with pronoun, always extraposed

Ram yah jaan-taa hai [ki Sita der se aa-egii]
 rAma yaha jAnawA hE ki sIwA xera se AegI
 Ram this know-Hab be.Prs.Sg that Sita delay with come-Fut.FSg

‘Ram knows that Sita will come late.’

- c. (Finite-complements-PS-2) Finite clause complements of N, non-extraposed

Ram [yah baat [ki Sita der se aa-egii]] jaan-taa hai
 rAma yaha bAwa ki sIwA xera se AegI jAnawA hE
 Ram this thing that Sita delay with come-Fut.f know-Hab be.Prs.Sg

‘Ram knows that Sita will come late.’

- d. (Finite-complements-PS-3) Finite clause complements of N, extraposed

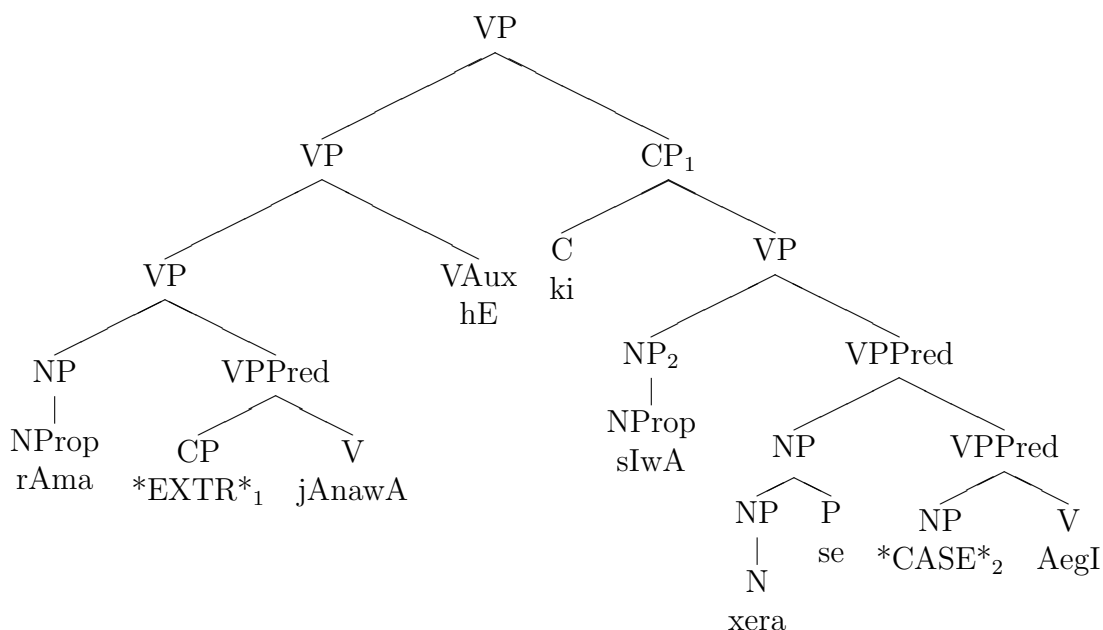
Ram [yah baat] jaan-taa hai [ki Sita der se aa-egii]
 rAma yaha bAwa jAnawA hE ki sIwA xera se AegI
 Ram this thing know-Hab be.Prs.Sg that Sita delay with come-Fut.FSg

‘Ram knows that Sita will come late.’

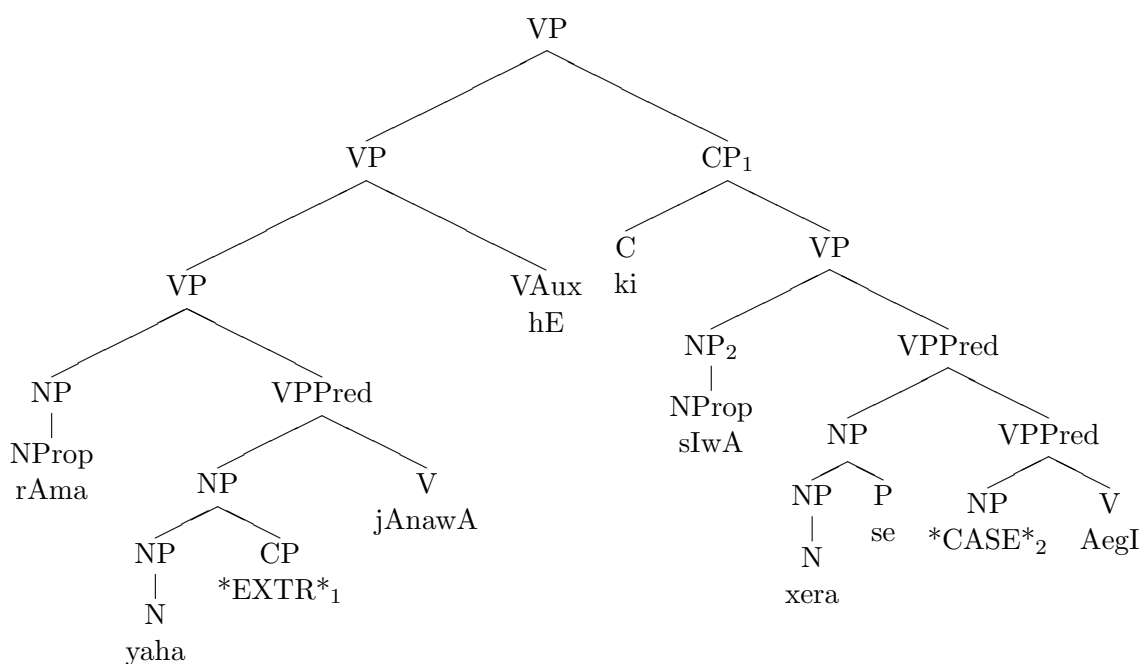
Extrapolation of CPs is marked by an *EXTR* trace. An *EXTR* trace is also used to indicate the relationship between a CP that is discontinuous from its source NP (see 28d). The CP combines with a source NP that contains a demonstrative as an adjunct. Whether we have a plain demonstrative as in (28b) or a demonstrative with a noun as in (28c/28d), the CP is adjoined to the whole NP. This expresses the need for there to be a demonstrative in the source NP. A structure where the CP is a complement of the noun, which is what is usually assumed for the English counterpart of such examples, would not express this relationship.

The trees are as follows.

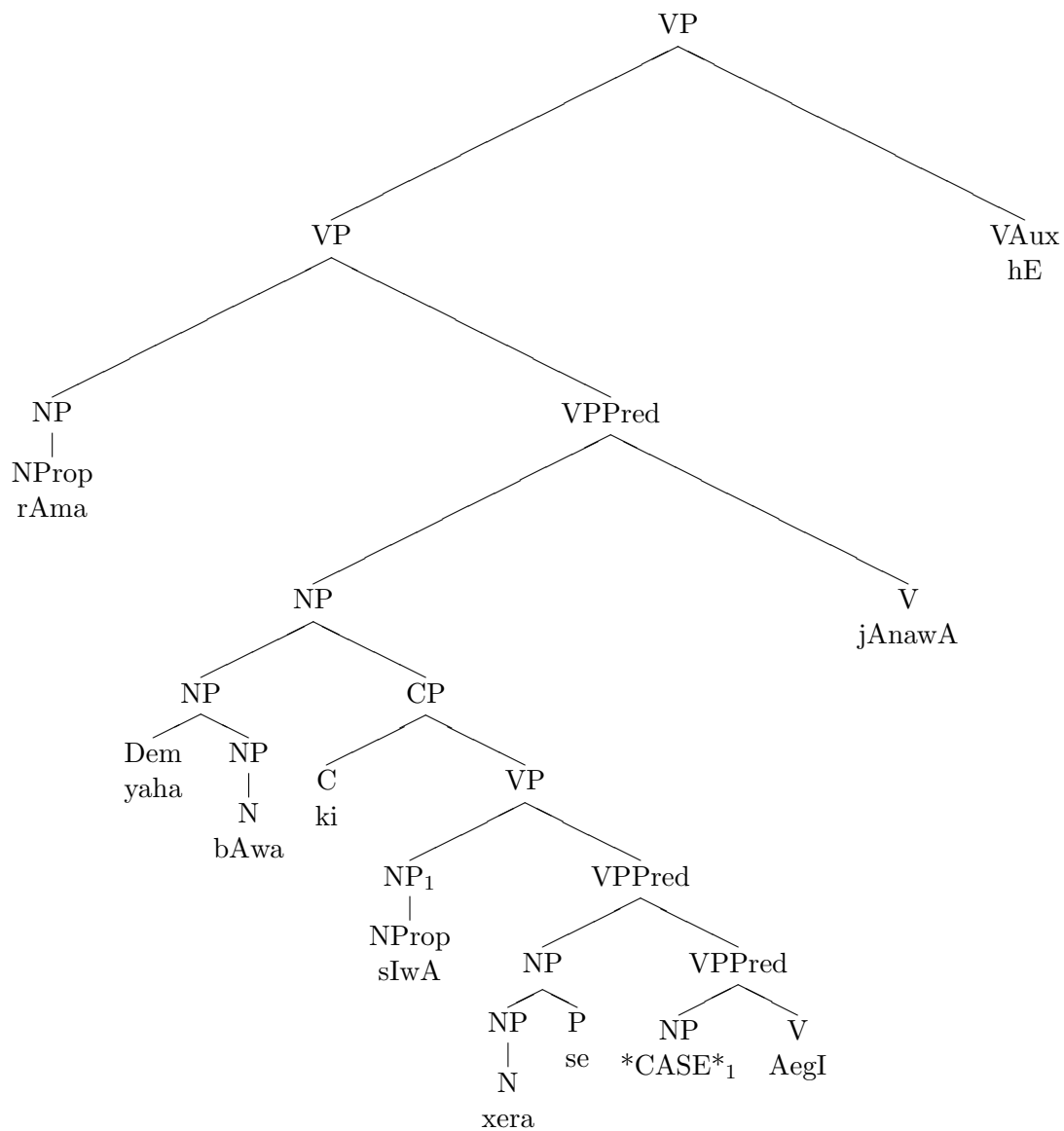
- (29) a. Tree for (28a) (Finite-complements-PS-1a) Finite clauses without pronoun, always extraposed



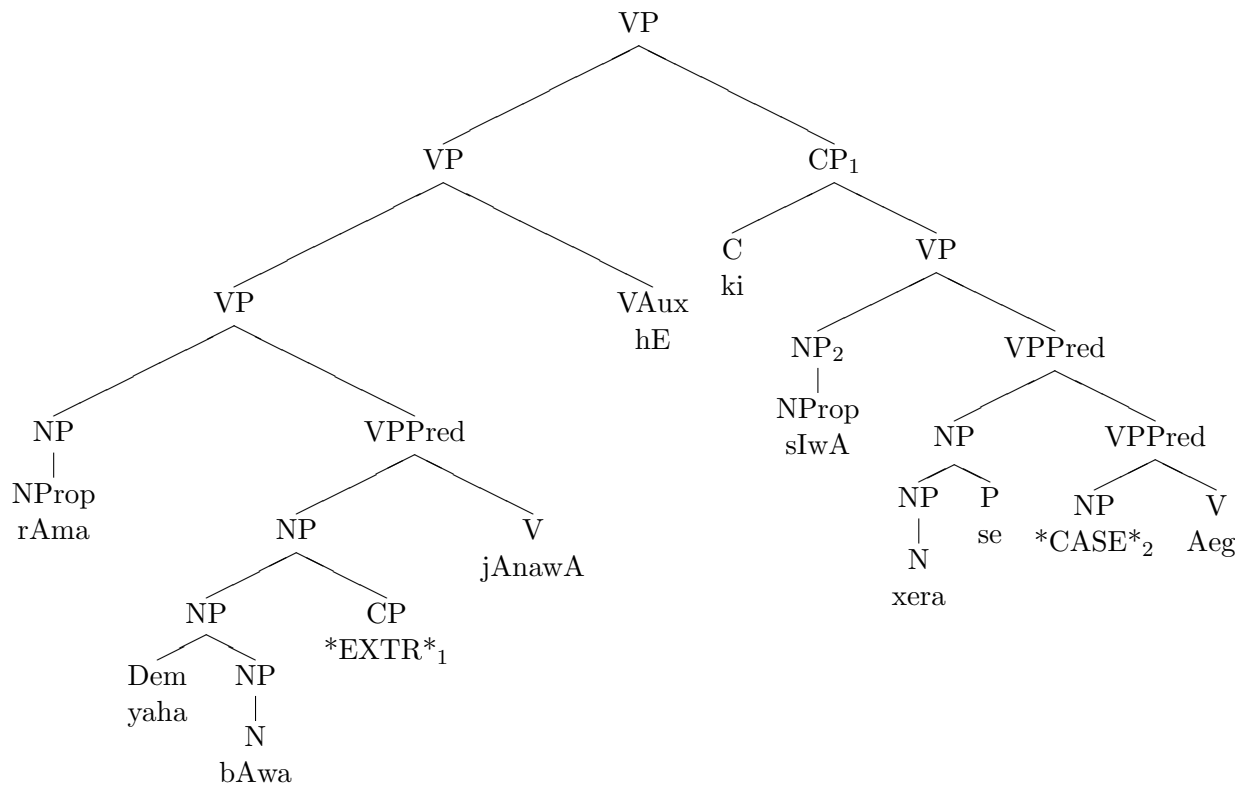
- b. Tree for (28b) (Finite-complements-PS-1b) Finite clauses with pronoun, always extraposed



- c. Tree for (28c) (Finite-complements-PS-2) Finite clause complements of N, non-extraposed



d. Tree for (28d) (Finite-complements-PS-3): Extraposition from NP



The discontinuous relationship between the NP and CP is indicated via extraposition movement which leaves behind an *EXTR* trace.

10.3 Missing Subject with Predicative Adjective and *ki* Complement Clause

Extraposition can also take place from subject position. In such cases, following the DS, we assume that there is a silent pronoun to which the extraposed CP starts off adjoined. (An alternative analysis would be to treat the CP as directly originating in the subject position and obligatorily extraposing. This treatment would be analogous to the treatment of (28a). We adopt the other treatment to maximize consistency with the DS.)

- (30) *pro* *EXTR*_i ullekhaniiya hai [ki unhoN-ne yeh kaam ki-yaa]i
 NULL worth.mentioning be.Pres that they-Erg this work do-Perf
 ‘It is worth mentioning that they did this work.’

10.4 Result Clauses

Result clauses are also realized by finite CPs. They are associated with one or more degree expressions in the main clause.

(31) Transitive:

a. (Result-Clauses-PS-1)

Atif ne itne laḍḍu khaa-e ki us-kaa peṭ
 AwiPZa ne iwane laddU KAe ki usakA peta
 Atif Erg so.many.MPl sweets eat-Pfv.MPl that he-Gen stomach
 dukh-ne lag-aa
 xuKane lagA
 hurt-Inf.Obl start-Pfv.MSg

‘Atif ate so many laddus that his stomach started hurting.’

b. (Result-Clauses-PS-2)

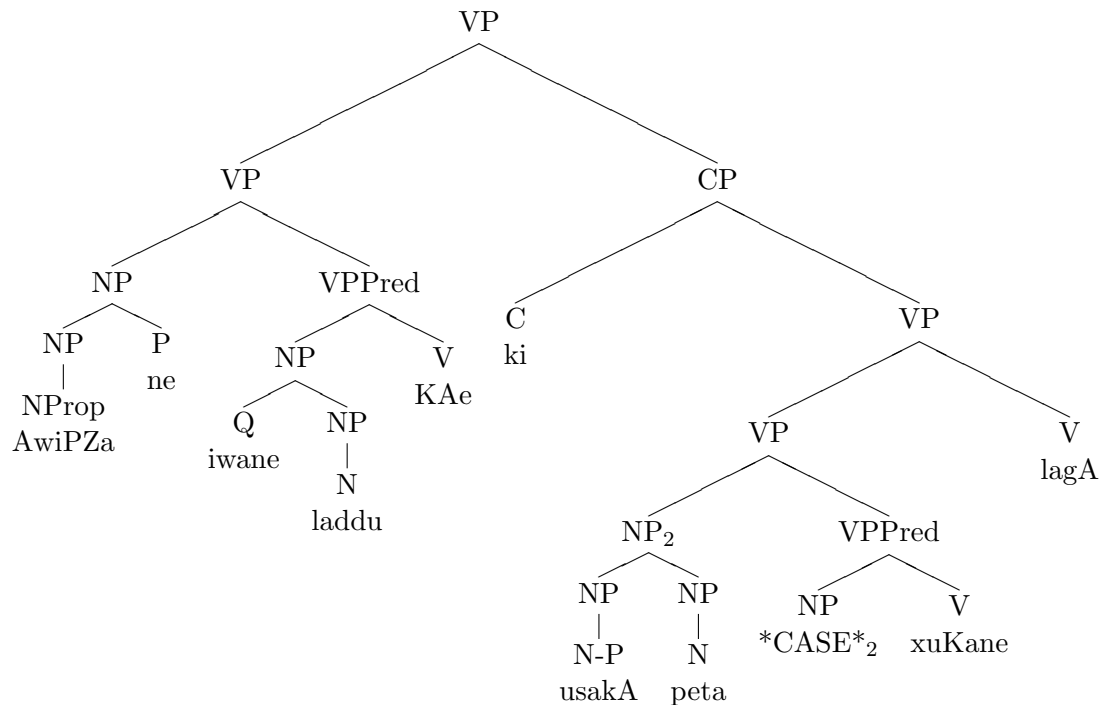
itne logõ ne itne laḍḍu khaa-e ki un-kaa
 iwane logoM ne iwane laddU KAe ki unakA
 so.many.MPl people Erg so.many.MPl sweets eat-Pfv.MPl that they.Gen
 peṭ dukh-ne lag-aa
 peta duKane lagA
 stomach hurt-Inf.Obl start-Pfv

‘So many people people ate so many laddus that their stomach started hurting.’

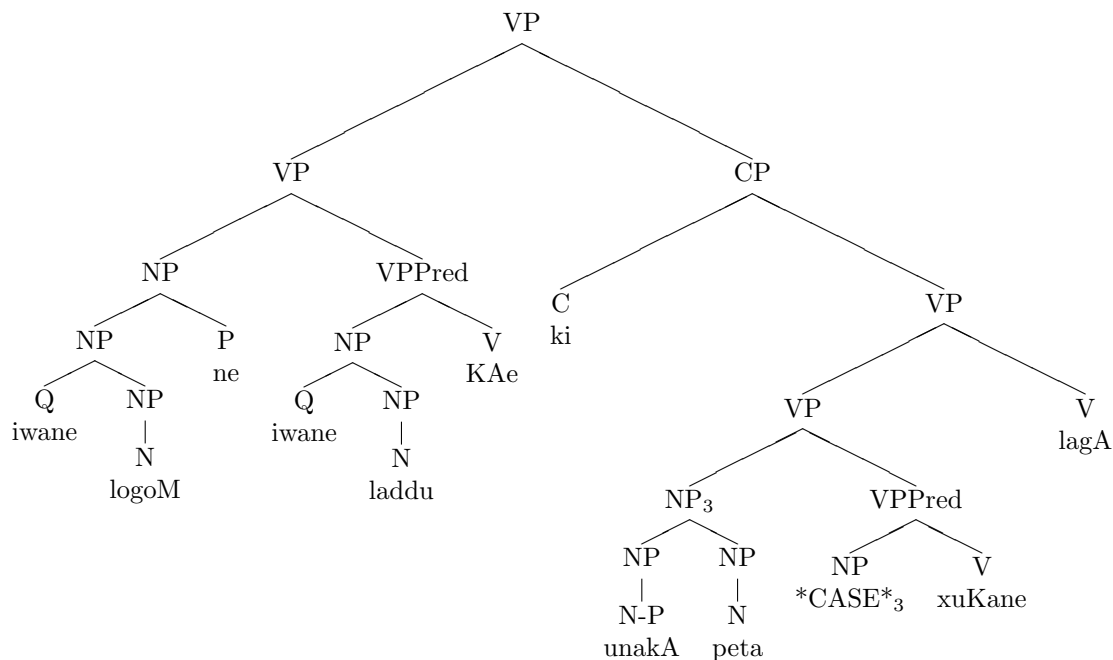
Trees:

(32) Transitive:

a. Tree for (31a) (Result-Clauses-PS-1):



b. Tree for (31b) (Result-Clauses-PS-2):



We do not explicitly indicate the relationship the degree phrases and the result clauses. Note that (31b) shows that this relationship cannot be reduced to movement.

10.5 Background: Clausal Complementation

10.5.1 Finite Complements and Extraposition

The problem is that extraposed CPs seem to behave for scope purposes as if they were pretty low in the structure. But their surface position seems to be pretty high as indicated by data from scope and binding. Object quantifiers can bind pronouns in the complement clause and matrix object pronouns cannot refer to proper names in the complement clause. One possibility is to just adjoin complement clauses at the IP level but coindex them systematically with an extraposition trace (*EXTR*) in/associated with the object position. This is what we will adopt here.

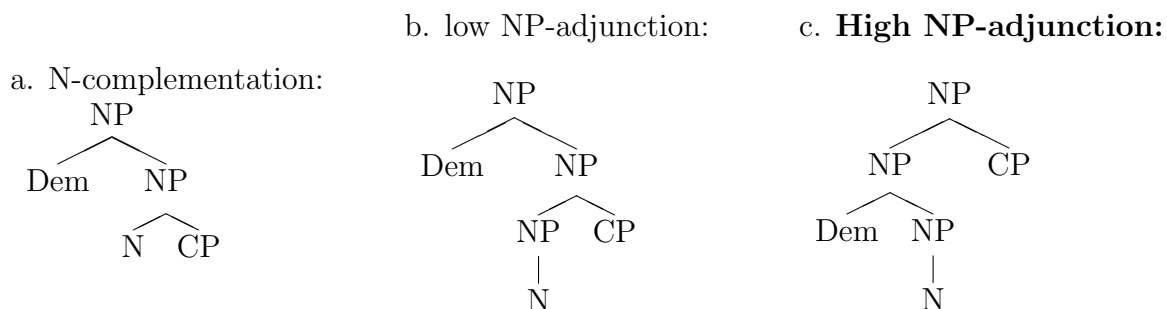
Almost all cases of extraposition will be represented by coindexing the moved XP and an appropriately located trace.

There are, in principle, three cases where coindexation does not correspond to movement:

- (33)
- a. result clauses: coindexation between the relevant D's and the right adjoined CP.
 - b. control: coindexation between controller and PRO
 - c. multi-correlatives: co-indexation between correlative CPs and the demonstrative phrases that they modify

For simplicity, we do not show these cases in the annotation. We take them to be semantic.

One other issue that needed to be sorted out is where the NP internal CP was to be attached. There were a number of options.



We adopted the structure in (c) for clausal complements. This is because there is a dependency between the presence of a demonstrative and the possibility of an extraposed complement. The source NP of the extraposition must contain a demonstrative. This dependency is only expressed in structure (c).

Chapter 11

Existentials, Possessives, and Locatives

Note: for technical reasons, some sentences in this section have numerical labels.

***ocr ***ocr This needs heavy editing (becuase it started life as an independent document).
I propose:

- An introductory subsection called Summary which lists all cases, including a back reference to predicative PPs.
- Then, separate subsections for: Existentials (=main verb *ho* with unaccusative analysis); Locative Possessives; Non-Locative Possessives; Experiencer Possessives; Possesives with Elided Possessed.
- Finally, a Background subsection on the definiteness effect.

All text needs editing as much text does not seem to reflect most recent analysis.*****

11.1 Summary

Predicative PPs have *NP PP be* word order, while possessives and existentials have *PP NP be* word order. Predicative PPs have been discussed in Section 12, page 107; we repeat them in this summary for convenience.

ocr ***ocr Need to refer to previous examples by name and page reference.** All three constructions use the stative version of *ho* ‘to be’, but differ in how arguments are projected. The predicative PP has a small clause complement which contains the predicate and its subject of predication (the NP); the small clause directly establishes a thematic relation between the two. Then, the subject of predication is moved from the small clause into the first position; in contrast, possessives and existentials do not have a small clause complement, and the NP is directly an argument of *ho* with the PP being an adjunct. No direct thematic relation between the NP and the PP is established.

- (34) (9-9.1.0-PS-1) Predicative PP:

Mina kamre me hai
 mInA kamare meM hE
 Mina room in is

‘Mina is in the room.’

For the structural representation, see (50c).

- (35) (Label:9-9.1.0-PS-3) Possessive:

Mona ke paas kai haar hẼ
 monA ke pAsa kaI hAra hEM
 Mona Gen.Obl near many necklace be.Prs.Pl

‘Mona has many necklaces.’

- (36) (9-9.1.0-PS-2) Existential:

us kamre me cuuhe hẼ
 usa kamare meM cUhe hEM
 that.Obl room in rats be.Prs.Pl

‘There are rats in that room.’

For structural representation, see (37).

The distinction between the three constructions should be made primarily semantically (what is the sentence saying in this context?), but the three constructions correlate with other factors:

- The NP in the possessive and the existentials is typically an indefinite.
- Possessive and existentials differ in that the PP in the possessive contains an animate NP (e.g. John-near) while the PP in the existential typically contains an inanimate location (e.g. school-near).
- There are few restrictions on the P that heads existentials and predicative PPs (in, on, inside, under, ...). In contrast, only a very limited set of P’s (-near, the dative, and the genitive) can head the PP in the possessives.

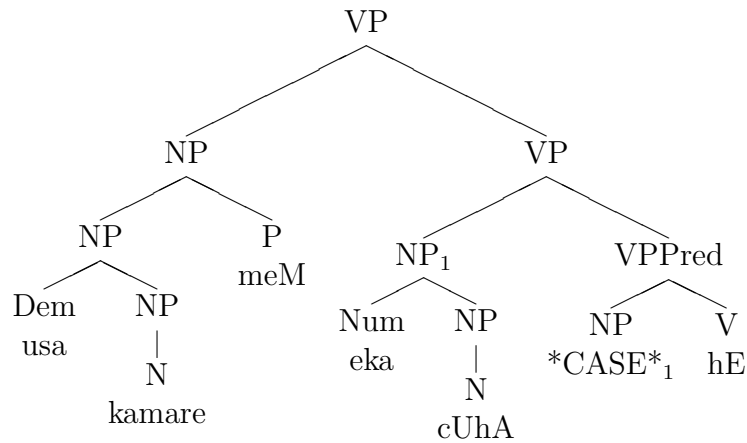
11.2 Existentials

- (37) (Existential-PS-21) Existentials and Possessives: the PP is an adjunct on a particular kind of VP headed by stative
- be*
- : order: PP NP
- be*
- (NP is indefinite)

us kamre me ek cuuhaa hai
 usa kamare meM eka cUhA hE
 that.Obl room in a mouse be.Prs.Sg

‘There is a mouse in that room.’

Tree for (37) (Existential-PS-21)



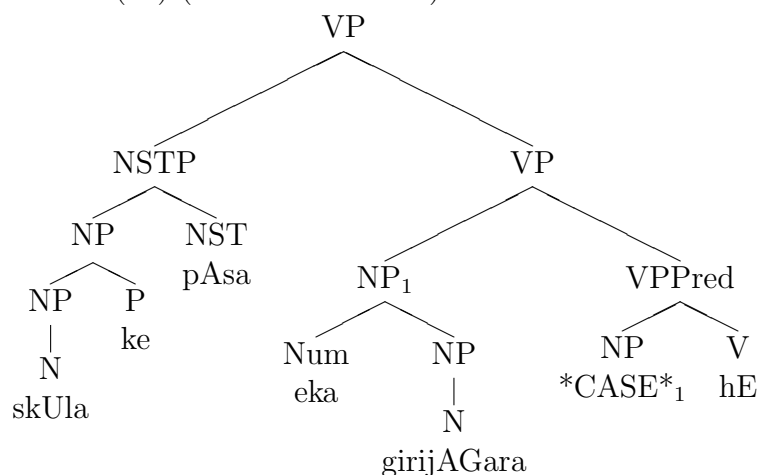
(38) (Existential-PS-22)

skuul ke paas ek girijaghar hai
 skUla ke pAsa eka girijAGara hE
 school Gen.Obl near a church be.Prs.Sg

‘There is a church near the school.’

(A possession reading ‘The school has/owns a church’ is also available. See next section for discussion.)

Tree for (38) (Existential-PS-22)



There is much flexibility with respect to the choice of the postposition with existential constructions. The PP has to denote a location and in the example above, this is done with *-ke paas* ‘near’. But it could also have been done with *-ke upar* ‘above’, *-ke niche* ‘below’, *-ke andar* ‘inside’, *-ke bahar* ‘outside’, *-mē* ‘in’, *-par/pe* ‘on’ etc.

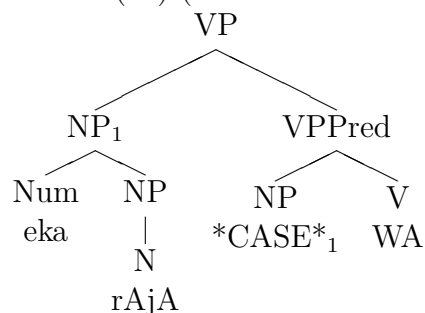
Note though that not all existential constructions have a PP. The analysis of PP-less existentials is not substantially different from that of regular existentials - the two differ only in that the former lacks the PP adjunct.

- (39) (Existential-PS-41) Existential without PP:

ek raajaa thaa
 eka rAjA WA
 a king be.Ps.MSg

‘There was a king.’

Tree for (39) (Existential-PS-41)



11.3 Locative Possessives

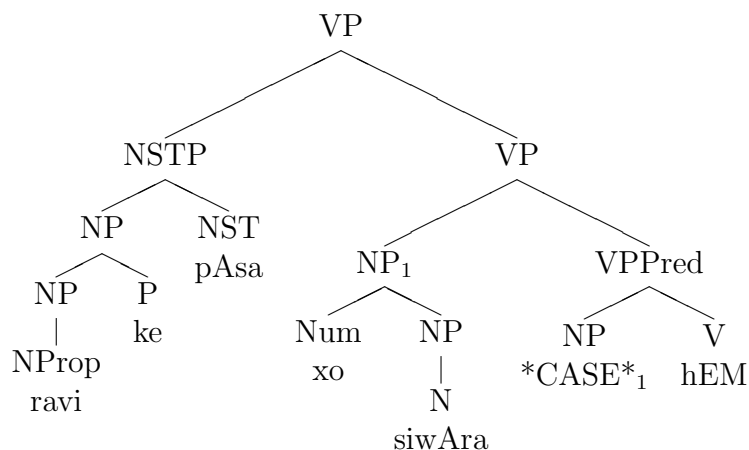
The structure adopted for existentials is also the structure for almost all possessive constructions. The only differences concern the form and contents of the PP. First of all the NP inside the PP in cases of possession is either animate or is the kind of object that can own things (such as a corporation, a city, or a group). Unlike existential constructions, a much smaller range of postpositions is available. The exact form of the postposition on the possessor NP determines the kind of possession at play. If NP is *NP-ke paas*, we get alienable possession. If NP is *NP-ko*, we get experiencer possession. And if NP is *NP-ke*, we get inalienable possession.

- (40) (Possessives-PS-31)

Ravi ke paas do sitaar hÊ
 ravi ke pAsa xo siwAra hEM
 Ravi Gen.Obl near two sitar be.Prs.Pl

‘Ravi has two sitars.’

Tree for (40) (Possessives-PS-31):



The syntax of possessive constructions and the syntax of existentials in Hindi-Urdu is essentially identical and indeed some cases are ambiguous between an existential meaning and a possessive meaning, see (38). As noted above, what distinguishes a possessive from an existential is that the PP in possession constructions is typically animate. Moreover the choice of P for alienable possession is essentially limited to the postposition *-ke paas* ‘near’.

11.4 Non-Locative Possession

As we have already seen, possession structures in Hindi-Urdu have the following word order.

(41) PP NP be

In the cases we have seen so far, the P is the locative postposition *-ke paas* ‘near’ and the structure is interpreted as denoting alienable possession. The language has distinct ways of marking inalienable possession and something we can call ‘experiencer possession’.

11.4.1 Inalienable Possession

The following is an example of an inalienable possession structure. Note that some speakers use the invariant *-ke* in this construction. Other speakers use a genitive marker that agrees with the possessed NP in number and gender *-kaa/-kii/-ke/-kii*.

(42) Inalienable Possession with an agreeing possessor:
order: NP-Gen_{Oblique/Agr} NP_{Agr} be (NP is indefinite)

Ravi ke/-kii do betiyã: hÊ
 ravi ke/kI xo betiyAz hEM
 Ravi Gen/Gen.f two daughters are

‘Ravi has two daughters.’

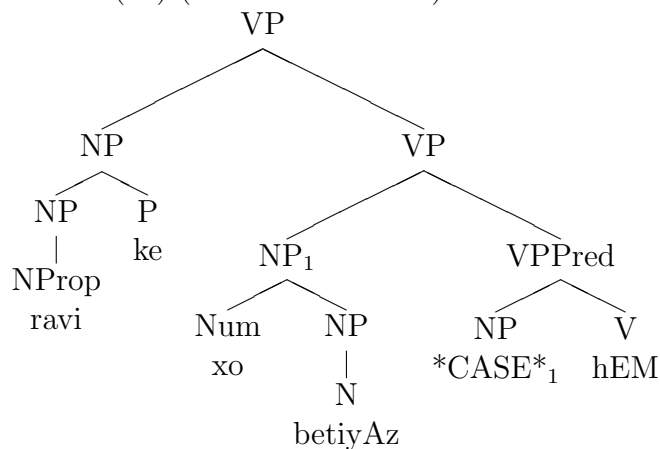
The analytical issue raised by these is that on the surface the agreeing variant of the [PP NP] sequence can be analyzed as a single NP i.e. as ‘Ravi’s two daughters’. So the question is whether we should just treat this as ‘Ravi’s two daughters are’. This seems straightforward but such an analysis is unavailable with the invariant *-ke*. This would mean that we would need to have two quite distinct analyses for the agreeing and non-agreeing variants. Let us first consider the non-agreeing *-ke*.

(43) (Possessives-PS-32)

Ravi ke do beṭiyā: hĒ
 ravi ke xo betiyAz hEM
 Ravi Gen two daughters are

‘Ravi has two daughters.’

Tree for (43) (Possessives-PS-32):



The invariant *-ke* appears in a number of environments not all of which are strictly possessive. This can be seen clearly in (44b, c) where the relationship expressed is not true possession. The *have* in (44b) is not the possessive *have* (‘have a child’) and in (44c), the meaning is not just that his blood start flowing but rather that he started bleeding. For these reasons, it seems reasonable to have a treatment of the oblique *ke* that is independent of the treatment of the agreeing possessor.

(44) a. (Possessives-PS-3) (ambiguous between variant and invariant use)

ghoṛe ke siing nahĩ: ho-te
 GodZe ke siMga nahIM howe
 horse.Obl Gen horn Neg be-Hab.MPl
 ‘A horse doesn’t have horns.’

b. (Possessives-PS-4)

bahut dinõ ke baad un-ke ek beṭaa huua
 bahuwa xinoM ke bAxa unake eka betA huA
 many days Gen.Obl after he-Gen a son be-Pfv
 ‘After a long time, he/she/they had a son.’

- c. (Possessives-PS-5)
- | | | | | |
|-----|-----|-------|--------------|-----------|
| us | ke | khuun | bah-ne | lag-aa |
| usa | ke | KUna | bahane | lagA |
| he | Gen | blood | flow-Inf.Obl | start-Pfv |
- ‘He started bleeding.’

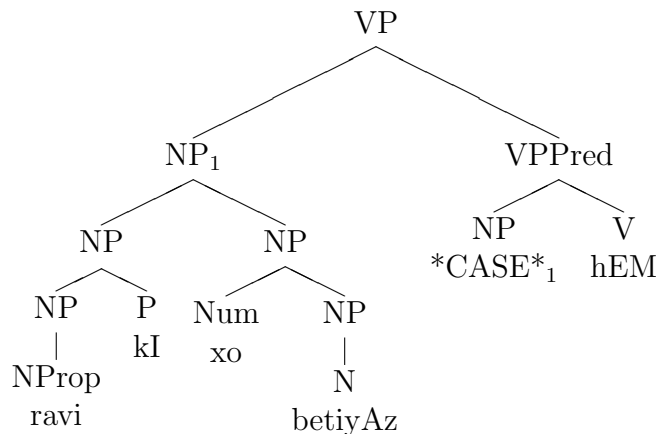
The oblique *-ke* patterns with the ethical dative and dative possessors that undergo what has been called ‘possessor raising’ in other languages. Therefore we will give it an analysis similar to the dative by adjoining the dative phrase in the general vicinity of the NP it is associated with. Thus in (44c), the oblique *-ke* phrase will be adjoined to the minimal VP associated with *bah* ‘flow’. Note that we do not actually move the oblique *-ke* phrase out of the associated NP.

Returning now to the agreeing case, we can now just see this on a parallel with the existential ‘there was a king’.

- (45) (Possessives-PS-33)
- | | | | | |
|------|-----|-----|-----------|-----|
| Ravi | kii | do | betiyã: | hĒ |
| ravi | kI | xo | betiyAz | hEM |
| Ravi | Gen | two | daughters | are |

‘Ravi has two daughters.’

Tree for (45) (Possessives-PS-33):



In some cases where the possessed NP has Masculine Plural features, the oblique genitive and the agreeing genitive have the same form. In such cases, unless the meaning disambiguates in favor of the oblique *-ke*, we will assume an agreeing possessor analysis. To isolate the meanings we might need to consider closely related variants where the features of the head NP help us disambiguate between the two structures.

Seemingly closely related to these cases are cases where the possessed NP appears before the possessor.

- (46) a. (Possessives-PS-61)

yeh computer Ravi kaa hai
 yaha computer ravi kA hE
 this computer.M Ravi Gen.MSg be.Prs.Sg

‘This computer is Ravi’s.’

- b. mĒ tumharii hũ:
 mEM wumhArI hUM
 I you.Gen.f be.Prs.1Sg

‘I_f am yours.’

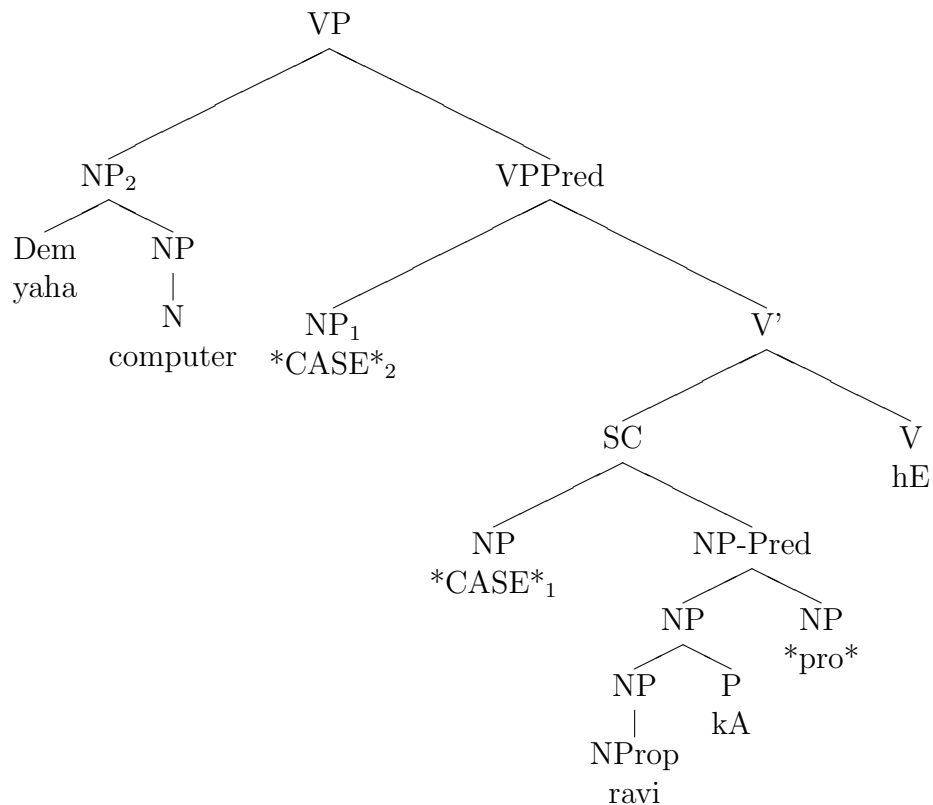
But there is good reason to believe that these cases are quite distinct and do not involve clausal possession at all. This is because they can be used to mark essentially any kind of possession, not just inalienable possession. In fact, since the possessed NP in these cases is typically definite, it is hard to get an inalienable possession reading. A more plausible analysis for these is to assume that these are ‘NP NP be’ copular sentences where a part of the second NP has undergone NP-ellipsis. Thus ‘This computer is Ravi’s’ is analyzed on a par with ‘This computer is Ravi’s computer’.

- (47) a. (Possessives-PS-6)

yeh computer Ravi kaa hai
 yaha computer ravi kA hE
 this computer.M Ravi Gen.MSg be.Prs.Sg

‘This computer is Ravi’s.’

Tree for (47a) (Possessives-PS-6):



11.4.2 Experiencer Possession

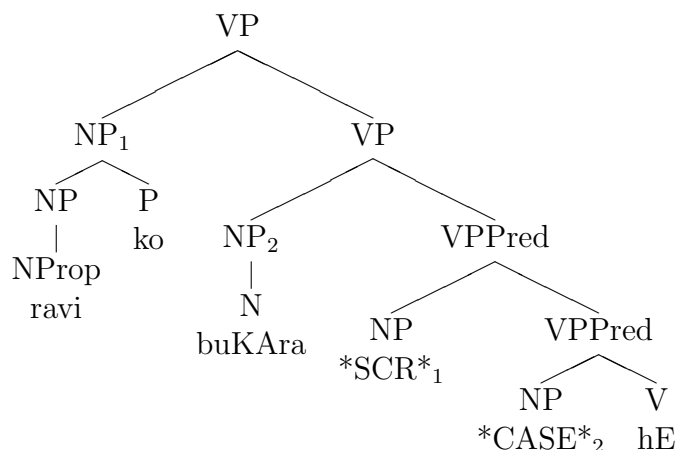
Experiencer Possession refers to a class of structures where the PP bears the dative postposition *-ko*. The PP is animate and is the experiencer of the NP.

(48) (Possessives-PS-7)

Ravi ko bukhaar hai
 ravi ko buKAra hE
 Ravi Dat fever be.Prs.Sg

‘Ravi has fever.’

Tree for (48) (Possessives-PS-7):



The structure we propose has the following properties: *Ravi-ko* is adjoined to VP and *bukhaar* moves from the low position to the high position. This structure resembles closely the structure adopted for dative subject constructions. There too, the dative is treated as an adjunct and the nominative argument moves from the low position to the high position.

11.5 Background: The Definiteness Effect and Exceptions

It was noted earlier that in existentials and locative possessives (order: PP NP be) the NP is typically indefinite. However, this is not always the case.

- (49) a. (9-9.4.0-PS-1) ‘existential’ with definite NP:
 kamre-me Ravi hai
 kamare meM ravi hE
 room-in Ravi be.Prs.Sg
 ‘It’s Ravi who is in the room.’
- b. (9-9.4.0-PS-2) ‘possessive’ with definite NP:
 Ravi ke paas merii car hai
 ravi ke pAsa merI kAra hE
 Ravi-Gen near my.f car be.Prs.Sg
 ‘Ravi has my car.’ (he doesn’t own it, just has it with him)
- c. (9-9.4.0-PS-3) related NP PP be:
 merii car Ravi ke paas hai
 merI kAra ravi ke pAsa hE
 my.f car Ravi-Gen near be.Prs.Sg
 ‘My car is with Ravi.’

¹Corresponding cases do not arise with inalienable possession or experiencer possession. The experiencer NP can be fronted as in the following example but this does not present any special challenges.

It is attractive to analyze these exceptions as being derived from ‘NP PP be’ structure underlyingly followed by scrambling of the PP. Analyzed this way, these structures are not existential/possessive constructions at all and hence are not exceptions to the indefiniteness requirement. However, at this point the DS assigns an identical representation to the existential ‘in the room is a mouse’ and the contrastive locative ‘in the room Ram is’ and so it is not clear that we can assign the desired structures to the definiteness effect violations. We might have to go with assigning an existential/possessive structure to these cases even though there are reasons to believe that this might not be the right move.

(1) (9-9.4.0-PS-4)

bukhaar Ravi-ko hai
 buKAra ravi ko hE
 fever Ravi-Dat be.Prs.Sg

‘It is Ravi who has a fever.’

Chapter 12

Non-Verbal Predicates

12.1 Predicative Use of Nouns, Adjectives, and Postpositions

A number of non-verbal categories in Hindi-Urdu can also function as predicates.

The following are examples of NPs, APs, and PPs functioning as predicates.

1

(50) a. (Copular-PS-1)

Ram doctor hai
rAma dOYktara hE
Ram doctor be.Prs.Sg

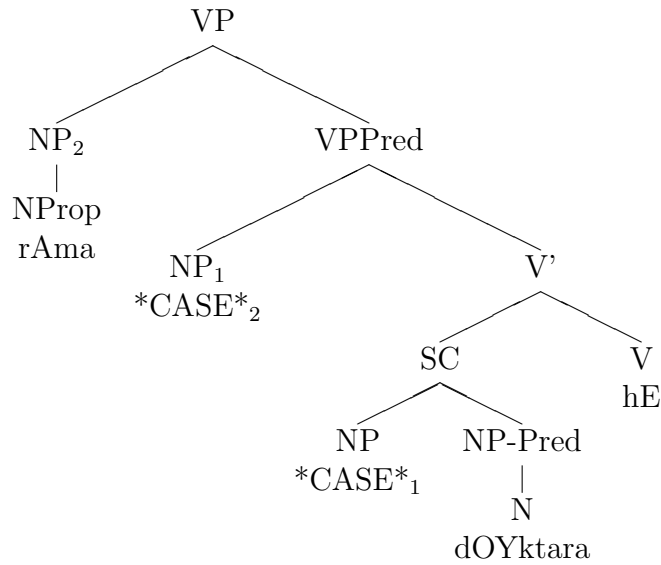
‘Ram is a doctor.’

Tree for (50a) (Copular-PS-1):

¹Certain adverbs may also function as non-verbal predicates. These are adverbs like *aaj* ‘today’.

- i. Ram kaa janmdin aaj hai
rAma kA janmdina Aja hE
Ram Gen birthday today be.Prs.Sg
‘Ram’s birthday is today.’

Other adverbs like *aksar* ‘often’ and *jaldii-se* do not function as non-verbal predicates. It seems that only adverbs that have the external syntax of NPs can function as non-verbal predicates. Temporarily, we can lump the adverbial non-verbal predicates with the NP non-verbal predicates.

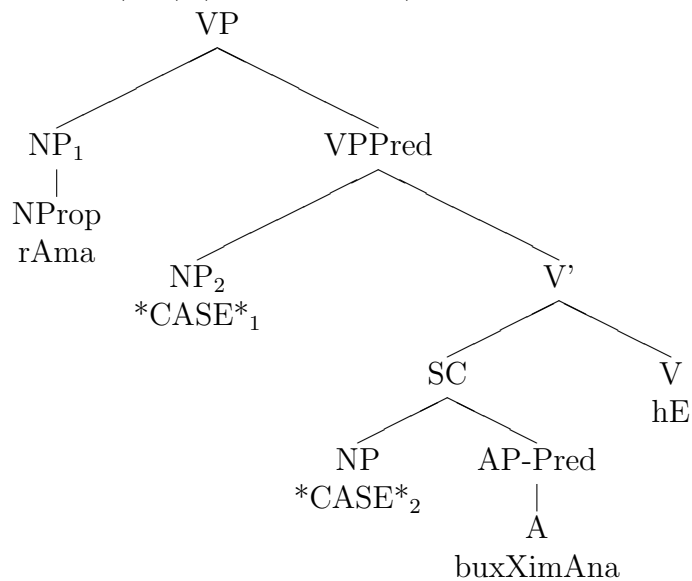


b. (Relation-PS-k1s-1)

Ram buddhimaan hai
 rAma buxXimAna hE
 Ram intelligent be.Prs.Sg

‘Ram is intelligent.

Tree for (50b) (Relation-k1s-1):

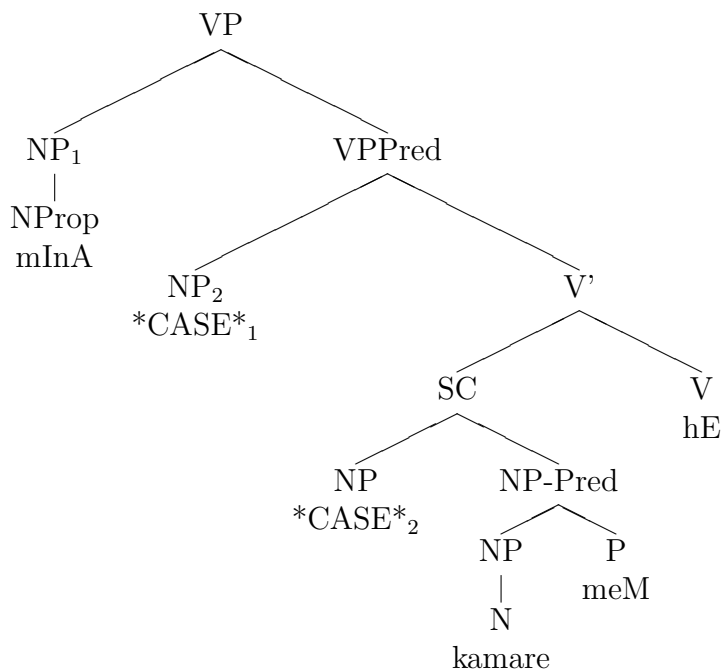


c. (Predicative-locative-PS-1)

Mina kamre me hai
 mInA kamare meM hE
 Mina room in be.Prs.Sg

‘Mina is in the room.

Tree for (50c) (Predicative-locative-PS-1):



These are analyzed as involving the main verb *be*, which is present for syntactic reasons. *be* does not contribute to the argument structure which is contributed by the noun, adjective, or postposition. The predicative nature of the NPs/APs/PPs is represented via the -PRED dash tag. This constituent then can take an argument and forms a small clause. We make a distinction between small clauses projected by predicative NPs/APs/PPs, represented by the node label SC/SC/SC respectively. The small clause combines with main verb *be* as a sister and the subject of the small clause raises for case reasons to the first argument position.

12.2 Adjectival Predicates with Eventive Meaning

The preceding cases involved stative readings of non-verbal predicates. We now turn to cases where non-verbal predicates yield a change of state reading. In these cases, the verb *ho* ‘be’ cannot appear in its present/past tensed form *hai/thaa* and instead must combine with aspectual morphology. Note that the cases at hand all involve adjectival small clauses.

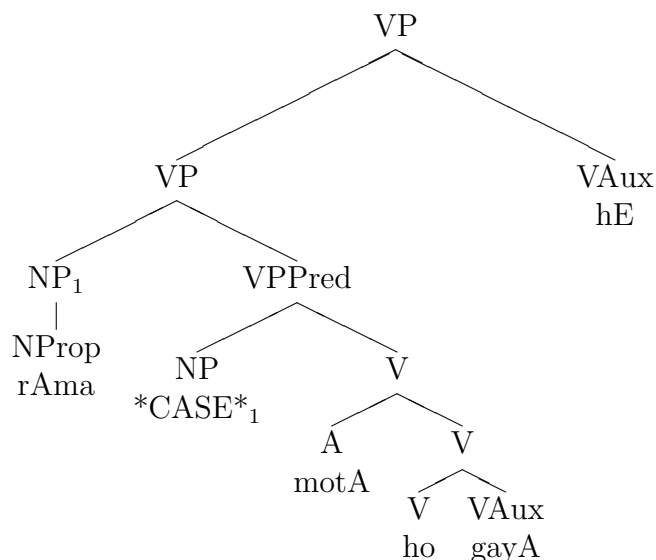
We handle such cases in exactly the same way as we handle stative non-verbal predicates. The subject of the small clause raises out to the specifier of the main verb. A fully articulated tree for such cases would be as follows.

(51) (AP-be-PS-1)

Ram motaa ho gayaa hai
 rAma motA ho gayA hE
 Ram fat.MSg be GO.Pfv.MSg be.Prs.Sg

‘Ram has become fat.’

Tree for (51)



The verb *kar* ‘do’ can also combine with adjectival small clauses. It introduces an additional argument, and the resulting meaning is causative.

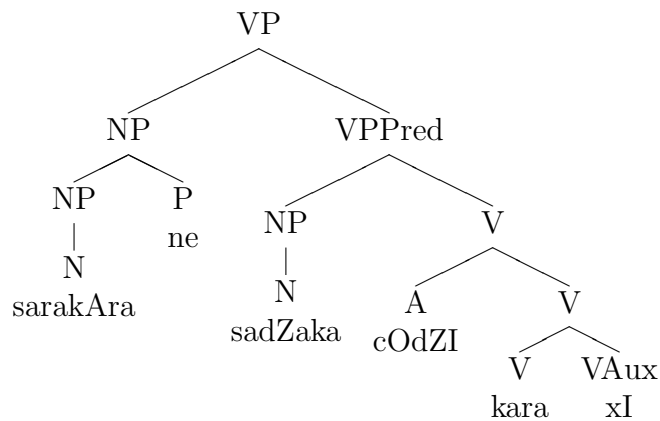
The treatment of these cases is similar to the case of adjectival small clauses with *ho* ‘be’, with the exception that the subject of the small clause ends up becoming the low argument. The high argument is contributed by *kar* ‘do’. Note that with *kar* ‘do’, we need the fully articulated structure to accommodate all of the arguments.

(52) (AP-do-PS-1)

sarkaar ne sarak cOri: kar dii
 sarakAra ne sadZaka cOdZI kara xI
 government Erg road.f wide do GIVE.Pfv.f

‘The government widened the road.’

Tree for (52) (AP-do-PS-1):



PP and NP small clauses cannot be complements of main verb *ho* ‘be’/*kar* ‘do’.

- (53) a. (Label:STAR-PP-be-PS-1)
 *Ram kamre me ho gayaa
 rAma kamare meM ho gayA
 Ram room in be GO.Pfv
 ‘*Ram became in the room.’
- b. (Label:STAR-NP-be-PS-1)
 *Ram doctor ho gayaa
 rAma dOYktara ho gayA
 Ram doctor be GO.Pfv
 ‘Ram became a doctor.’
- c. (Label:STAR-PP-do-PS-1)
 *Ram ne Ravi ko kamre me kiyaa
 rAma ne ravi ko kamare meM kiyA
 Ram Erg Ravi Acc room in do-Pfv
 ‘Ram caused Ravi to be in the room.’
- d. (Label:STAR-NP-do-PS-1)
 *Ram ne Ravi ko doctor kiyaa
 rAma ne ravi ko dOYktara kiyA
 Ram Erg Ravi Acc doctor do.Pfv
 ‘Ram made Ravi a doctor.’

12.3 Non-Verbal Nominal Predicates with *ban/banaa* ‘become/make’

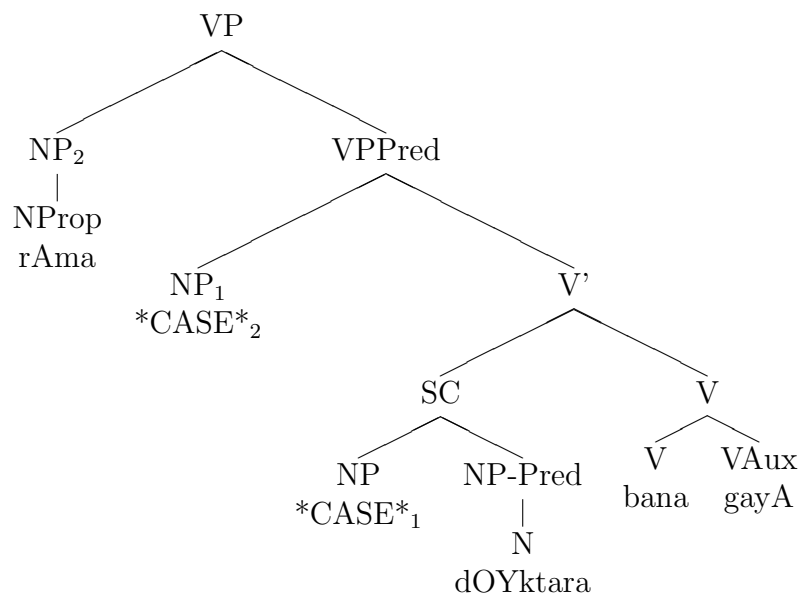
To obtain an eventive meaning, nominal predicates combine with the main verb *ban* ‘become’. The structure parallels that of the adjectival small clause complements of *ho* ‘be’ in (51).

(54) (NP-be-PS-1)

Ram doctor ban gayaa
 rAma dOYktara bana gayA
 Ram doctor become GO.Pfv

‘Ram became a doctor.’

Tree for (54) (NP-be-PS-1):



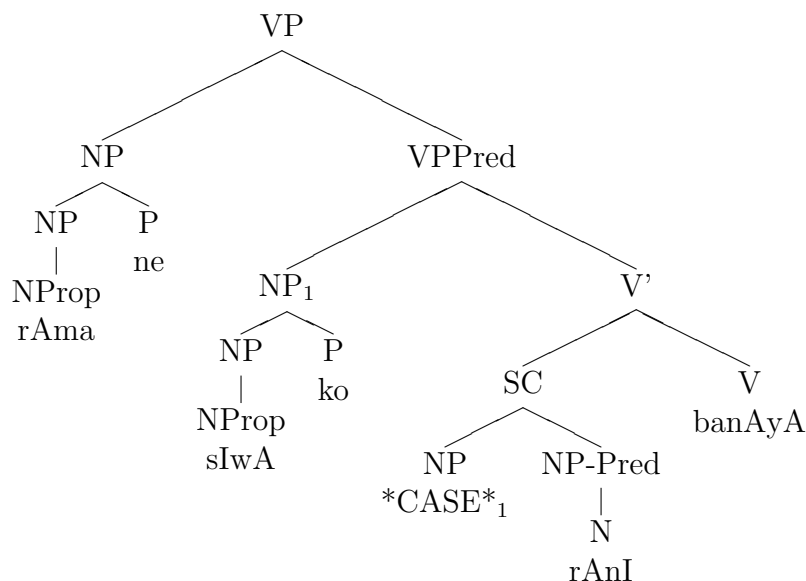
Just as the adjectival predicate combines with *kar* ‘do’, the transitive counterpart of *ho* ‘be’, the nominal predicate combines with *banaa* ‘make’, the transitive counterpart of *ban* ‘become’. Again, we analyze the construction using a small clause.

(55) (NP-do-PS-1)

Ram ne Sita ko raanii banaayaa/*banaayii
 rAma ne sIwA ko rAnI banAyA/*banAyI
 Ram Erg Sita Acc queen.f make.Pfv/make.Pfv.f

‘Ram made Sita a queen.’

Tree for (55) (NP-do-PS-1):



This example displays an important property of NP-Preds: they never trigger agreement. Note that the main verb surfaces with default agreement and does not agree with the feminine NP *rAnI*.

For completeness, we note that *ban/banaa* ‘become/make’ can only combine with nominal predicates and not with PP/AP predicates, at least not with the interpretations that we are interested in.

- (56) a. (Label:STAR-PP-become-PS-1)
 *Ram kamre me banaa
 rAma kamare meM banA
 Ram room in became.Pfv
 ‘Ram became in the room.’
 (has irrelevant interpretation: ‘Ram was made in the room.’)
- b. (Label:STAR-AP-be-PS-1)
 *Ram lambaa banaa
 rAma lambA banA
 Ram tall became.Pfv
 ‘Ram became tall.’
 (has irrelevant interpretation: ‘Ram was constructed tall.’)
- c. (Label:STAR-PP-make-PS-1)
 *Ram ne Ravi ko kamre me banaayaa
 rAma ne ravi ko kamare meM banAyA
 Ram Erg Ravi Acc room in do-Pfv
 ‘Ram caused Ravi to be in the room.’
 (has irrelevant interpretation: ‘Ram made Ravi in the room.’)
- d. (Label:STAR-AP-make-PS-1)

*Ram ne Ravi ko lambaa banaayaa

rAma ne ravi ko lambA banAyA

Ram Erg Ravi Acc doctor do.Pfv

‘Ram made Ravi tall.’

(irrelevant interpretation: ‘Ram constructed Ravi to be tall.’)

Chapter 13

Adjuncts: Instrument, Source, Location

13.1 Adverbs

aksar ‘often’

jaldii: perhaps some sort of nominal

jaldii-se

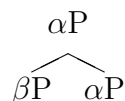
Chapter 14

Clausal Adjuncts

14.1 General

Adjuncts is the name for the following configuration: any element βP that combines with an αP to yield another αP is an adjunct on αP .

(57)



Adjuncts in the PS annotation cover two distinct classes. The first class corresponds to the general intuition that adjuncts are elements that are not lexically selected. The following is a list of such elements.

(58) Clear adjuncts: these are not lexically selected

- a. Adverbial Modifiers such as frequency, manner, temporal, locative, and reason adverbs
Adjoin to V' , $VPPred$, VP
- b. Negation, Incorporated Nouns
Adjoin to V
- c. Possessors, almost all modifiers to NP
Adjoin to NP
- d. $DegP$
Adjoin to NP , AP , VP etc.
- e. Moved XPs (these leave behind a trace)
Adjoin to any phrasal projection compatible with the word order
- f. Extraposed CPs
Right-Adjoin to VP

- almost all adjunction is left-adjunction. Right adjunction is found primarily with extraposed clauses (complement clauses, relative clauses, and result clauses).

One property of adjuncts is that when they are not present, there is no element in the syntactic representation that corresponds to them. This is different from the first/second arguments which are represented even when they are not overtly realized.

The second class of elements that are treated as adjuncts by the PS are lexically selected arguments that are not first/second arguments.

- (59) Selected elements which are not first/second arguments:
- a. Dative Goals and Experiencers
Start as adjoined to minimal VPPred
- even though these are structurally adjuncts, they have a designated location where they appear. As a result, their displacement involves traces.
 - b. Non-dative selected elements
(e.g. *Tim Bill-se mil-aa* ‘Tim Bill-with met’)
- these are treated as adjuncts and their displacement does not involve traces.

As with other adjuncts, when these elements are not overtly present, there is no covert syntactic representation that corresponds to them.

14.2 Participial Adjunct Clauses

Hindi/Urdu uses participial adjunct clauses productively to indicate temporal sequencing of events and causal relations between events.

These clauses are adjuncts on the main VP, VPPred, or V’.

Typically adjunct clauses will have a zero subject, which will be coreferential with an argument of the main clause. The zero subject is represented but the relationship between the PRO and its controller is not represented.

Adjunct clauses can appear in a variety of participial endings. The adjunct clauses in (61a-c) are treated as VP. When the participial ending is written separately from the verb, as in (61c), it is treated as an auxiliary (VAux) that is adjoined to the VP. When the participial ending is written with the verb, as in the first case below, its presence is indicated by the V+VAux label. The participial verb can be reduplicated; reduplication of the verb contributes a meaning of repetition.

The following morphosyntactic devices are used to mark participial adjunct clauses:

- (60) a. V + *kar*:
contributes the meaning ‘Having V’d, ...’
can be written as one word or two, *kar* is treated as VAux that adjoins to VP if written separately; if written with V, it is treated as V+VAux. V+VAux projects as V.
- b. V + V + *kar*:
contributes the meaning ‘Having repeatedly V’d, ...’
can be written as two words or three, *kar* is treated as VAux that adjoins to

VP if written separately; if written with V, it is treated as V+VAux. V+VAux projects as V.

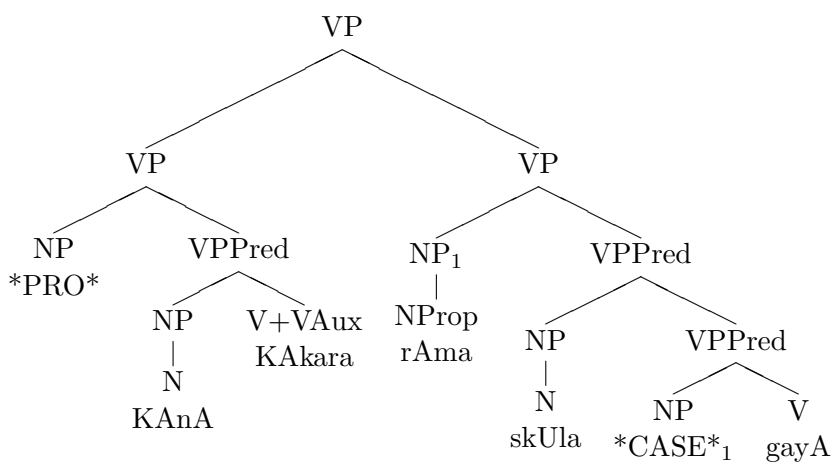
- c. *V-te + hue*:
contributes the meaning ‘While V-ing, ...’
hue is treated as VAux that adjoins to VP.
- d. *V-te + V-te*:
contributes the meaning ‘While repeatedly V-ing, ...’

(61) a. (Participial-Adjunct-Clauses-PS-1) *-kar*:

[khaanaa khaa-kar] Ram skuul gayaa
 KAnA KAkara rAma skUla gayA
 food eat-having Ram school go.Pfv.MSg

‘Ram went to school after eating lunch.’

Tree for (61a) (Participial-Adjunct-Clauses-PS-1)

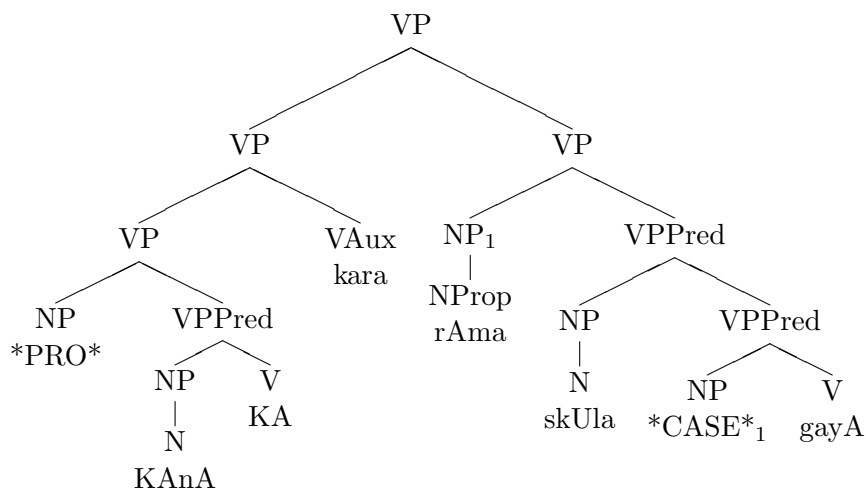


b. (Participial-Adjunct-Clauses-PS-1A) *kar* written separately:

[khaanaa khaa kar] Ram skuul gayaa
 KAnA KA kara rAma skUla gayA
 food eat having Ram school go.Pfv.MSg

‘Ram went to school after eating lunch.’

Tree for (61b) (Participial-Adjunct-Clauses-PS-1A)

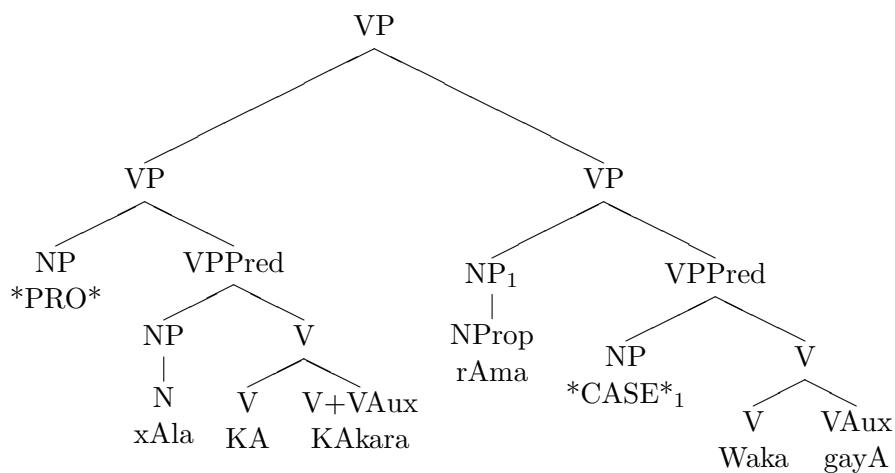


- c. (Participial-Adjunct-Clauses-PS-2) V-reduplication + *-kar*:

[daal khaa khaa-kar] Ram thak gayaa
 xAla KA KAkara rAma Waka gayA
 daal eat eat-having Ram tired go.Pfv.MSg

‘Ram has gotten tired from eating daal.’

Tree for (61c) (Participial-Adjunct-Clauses-PS-2)

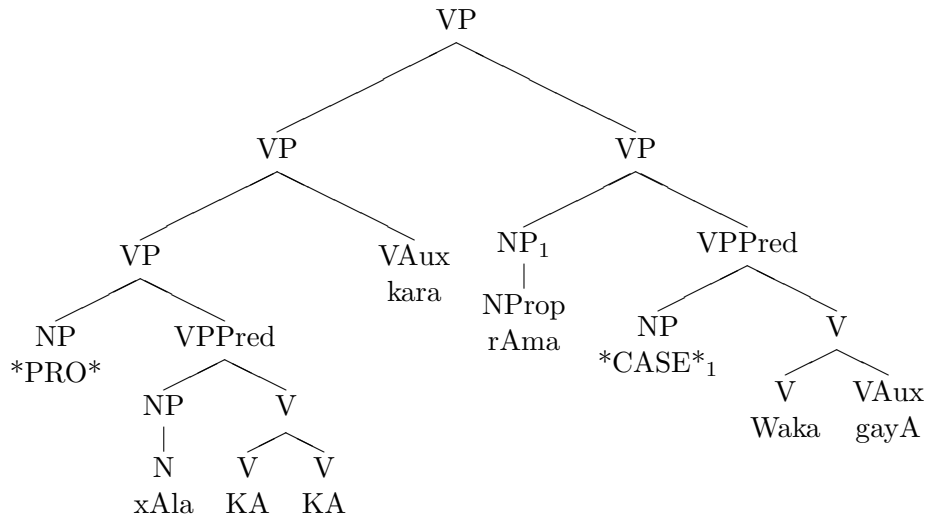


- d. (Participial-Adjunct-Clauses-PS-2A) V-reduplication + *kar* written separately:

[daal khaa khaa kar] Ram thak gayaa
 xAla KA KA kara rAma Waka gayA
 daal eat eat having Ram tired go.Pfv.MSg

‘Ram has gotten tired from eating daal.’

Tree for (61d) (Participial-Adjunct-Clauses-PS-2A)

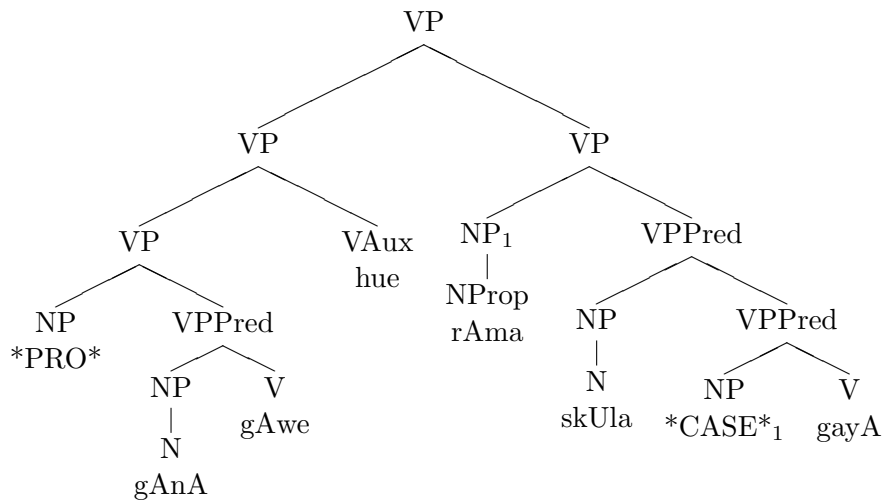


e. (Participial-Adjunct-Clauses-PS-3) *-te hue*:

[gaanaa gaa-te hue] Ram skuul gayaa
 gAnA gAwe hue rAma skUla gayA
 song sing-Impfv Ram school go.Pfv.MSg

‘Ram went to school singing a song.’

Tree for (61e) (Participial-Adjunct-Clauses-PS-3)

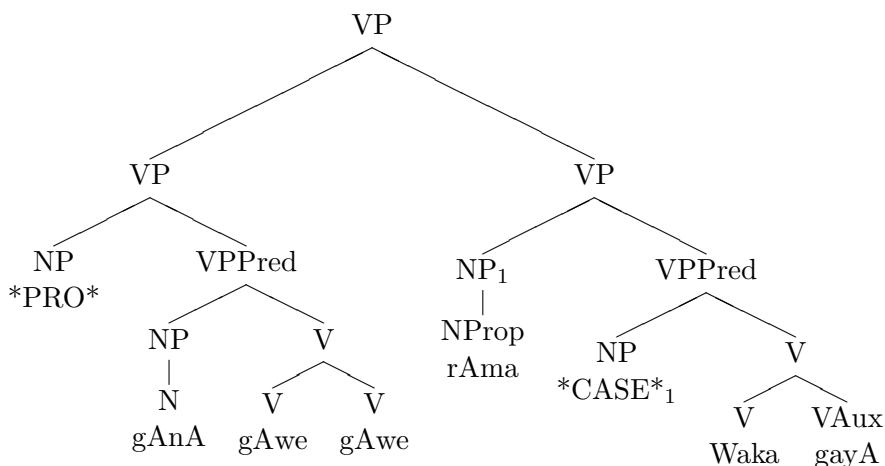


f. (Participial-Adjunct-Clauses-PS-3A) V-reduplication:

[gaanaa gaate gaate] Ram thak gayaa
 gAnA gAwe gAwe rAma Waka gayA
 song singing singing Ram tired go.Pfv.MSg

‘Singing a song, Ram got tired.’

Tree for (61f) (Participial-Adjunct-Clauses-PS-3A)



14.2.1 Gerunds

Non-finite adjuncts can also be based on the *-naa* forms of verbs. The adjunct clause can have a genitive subject or a silent subject. Such clausal adjuncts combine with NSTs like *pehle* ‘before’ and *baad* ‘after’ to give the meaning of before/after clauses.

Depending upon a number of parameters, *-naa* clauses may have an overt genitive subject or a *PRO* subject. Our treatment of *-naa* clauses is based on this aspect of *-naa* clauses. If they do not have a genitive subject, they are treated entirely like other VPs.

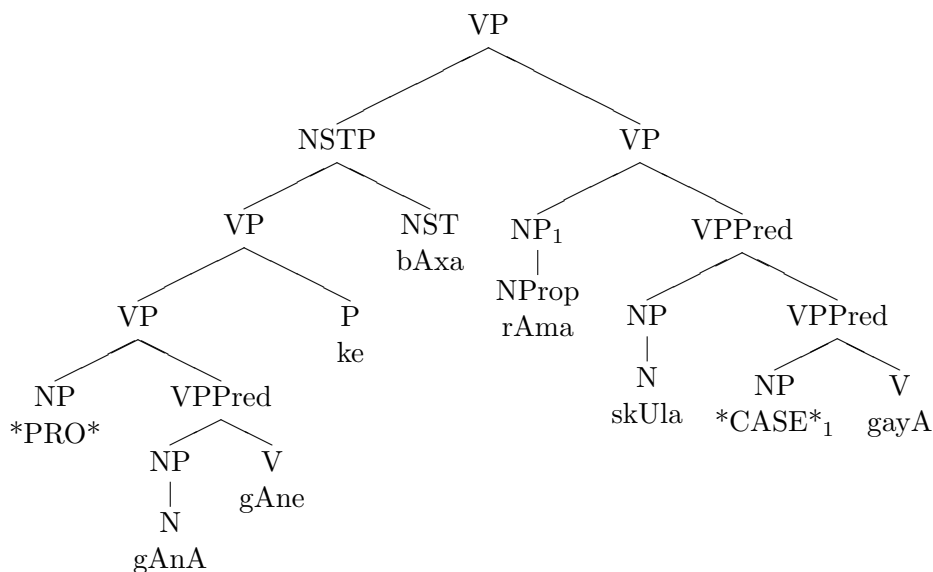
The treatment of (62), which does not have an overt genitive subject, is analogous to other types of adjunct clauses: it has a *PRO* subject, which is controlled by an argument of the main clause.

- (62) (Participial-Adjunct-Clauses-PS-4) infinitival + temporal modifiers: *-ke baad* ‘after’, *-se pehle* ‘before’ (this class allows for overt genitive subjects. This is consonant with its nominal nature.)

| | | | | | | |
|---------|--------------|---------|-------|------|--------|------------|
| [gaanaa | gaa-ne | ke | baad] | Ram | skuul | gayaa |
| gAnA | gAne | ke | bAxa | rAma | skUla | gayA |
| song | sing-Inf.Obl | Gen.Obl | after | Ram | school | go.Pfv.MSg |

‘Ram went to school after singing a song.’

Tree for (62) (Participial-Adjunct-Clauses-PS-4)



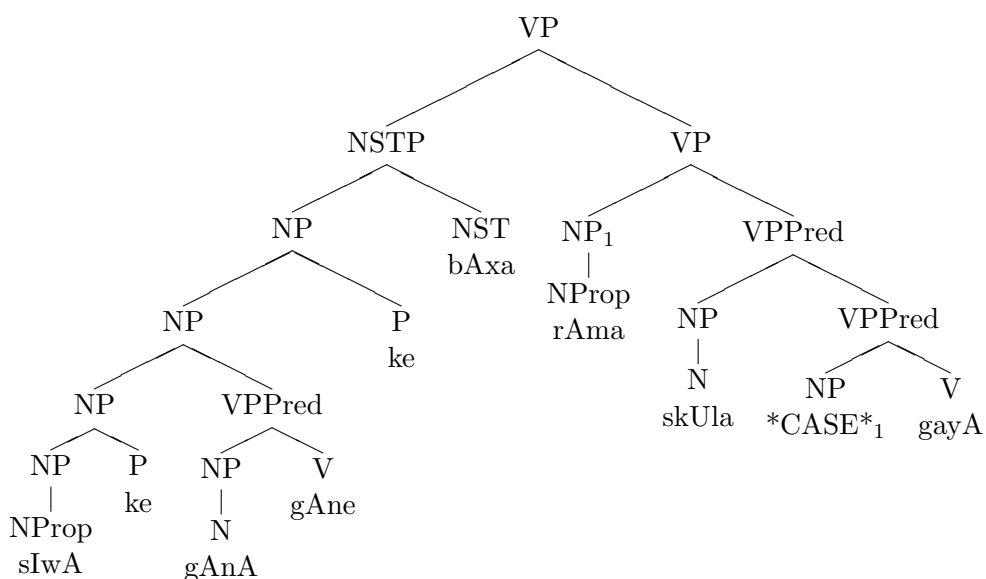
However, if a *-naa* clause has an overt genitive subject, we distinguish it from other VPs by making its top-level phrase label NP. Below the NP, it has the usual verbal projection. This NP label reflects the nominal nature of this clause, indicated by its allowing an overt genitive subject.

- (63) (Participial-Adjunct-Clauses-PS-41) infinitival with subject + temporal modifiers: *-ke baad* ‘after’, *-se pehle* ‘before’ (overt genitive subjects → nominal nature.)

[Siita ke gaanaa gaa-ne ke baad] Ram skuul gayaa
 sIwA ke gAnA gAne ke bAxa rAma skUla gayA
 song sing-Inf.Obl Gen.Obl after Ram school go.Pfv.MSg

‘Ram went to school after Sita’s singing a song.’

Tree for (63) (Participial-Adjunct-Clauses-PS-41)



We are aware that only marking those *-naa* clauses which have genitive subjects as NPs under-represents the scope of gerunds in Hindi. But we follow this regimen as the systematic ambiguity between infinitives and gerunds in Hindi makes a precise adjudication difficult. The genitive criteria is completely reliable - any *-naa* clause with a genitive subject is nominal. But some *-naa* clauses marked as VP could permit a nominal analysis.

Chapter 15

Negation

15.1 Sentential Negation

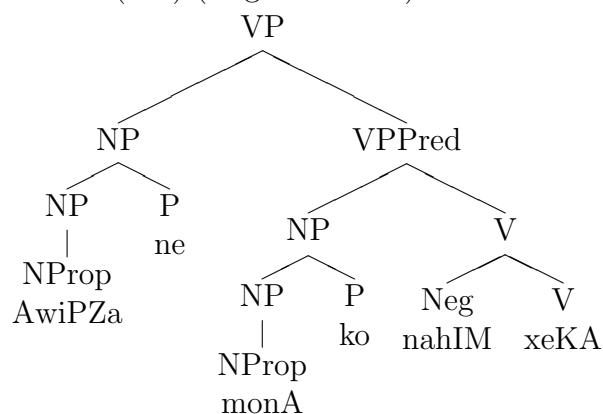
Negation appears immediately preceding the verbal complex. Sentential negation cannot appear discontinuous from the verbal complex.

(64) a. (Negation-PS-1)

Atif ne Mona ko nahĩ: dekh-aa
 AwiPZa ne monA ko nahIM xeKA
 Atif Erg Mona Acc neg see-Pfv

‘Atif did not see Mona.’

Tree for (64a) (Negation-PS-1)



b. (STAR-Negation-PS-2)

*Atif-ne nahĩ: Mona-ko dekh-aa
 AwiPZa ne nahIM monA ko xeKA
 Atif Erg neg Mona Acc see-pfv

The adjacency requirement is indicated by adjunction to V. The most unmarked order is when negation immediately precedes the main verb. However, negation can also follow the

verb and when a number of auxiliaries are present, negation can appear adjacent to any of them (to a first approximation). When negation immediately follows the verbal complex, it has an emphatic status.

(65) (SUMMARY-Negation-PS-3) ✓ indicates locations where *nahĩ:* can appear:

| | | | | | | | | |
|--------|--------|----------------|-------|----------------|----------|-----------------|-----------|----------------|
| Atif | skuul | ✓ ₁ | jaa | ✓ ₂ | rahaa | ?✓ ₃ | hai | ✓ ₄ |
| AwiPZa | skUla | | nahIM | jA | rahA | | hE | |
| Atif.m | school | | go | | Prog.Msg | | be.Prs.Sg | |

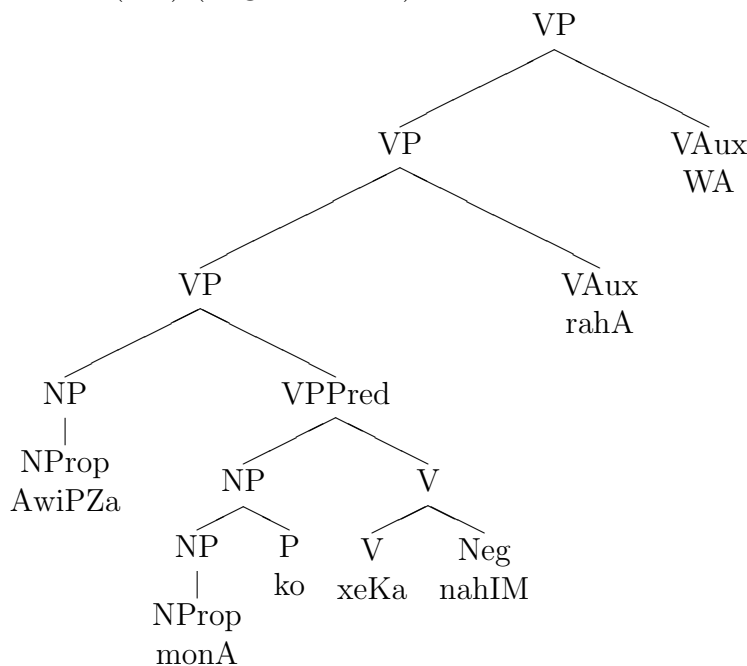
‘Atif is not going to school.’

a. (Negation-PS-2)

| | | | | | | |
|--------|------|-----|------|-------|-------|--------|
| Atif | Mona | ko | dekh | nahĩ: | rahaa | thaa |
| AwiPZa | monA | ko | xeKa | nahIM | rahA | WA |
| Atif | Mona | Acc | see | neg | Prog | be.Pst |

‘Atif wasn’t looking at Mona.’ (could be used in a situation where Atif has been asked to look at Mona but is stubbornly refusing to do so. He could be talking to her etc. but is just not *looking* at her.)

Tree for (65a) (Negation-PS-2)

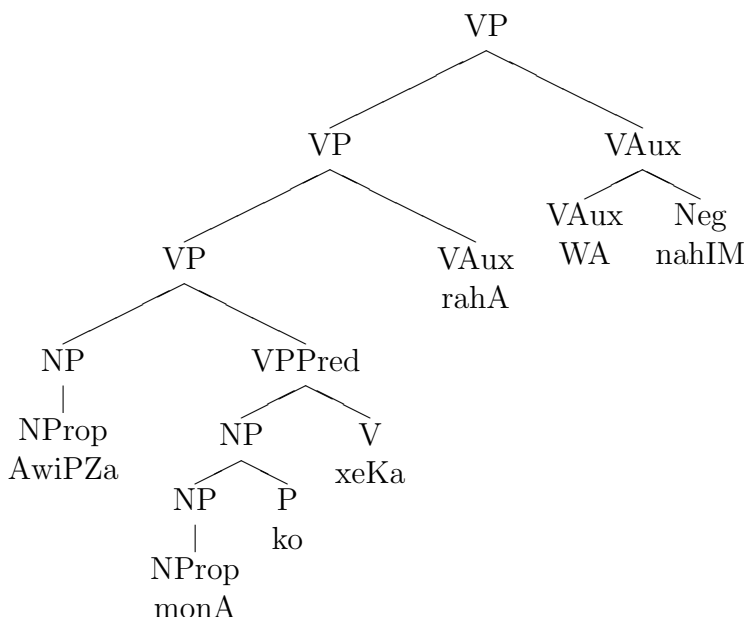


b. (Negation-PS-3)

| | | | | | | |
|--------|------|-----|------|-------|--------|-------|
| Atif | Mona | ko | dekh | rahaa | thaa | nahĩ: |
| AwiPZa | monA | ko | xeKa | rahA | WA | nahIM |
| Atif | Mona | Acc | see | Prog | be.Pst | Neg |

‘Atif wasn’t looking at Mona.’ (could be used in a situation where it has been claimed that Atif was looking at Mona.)

Tree for (65b) (Negation-PS-3)



ocr ***ocr The examples should be the same as the summary sentence. I guess the summary sentence should be changed. Check the "xiKa" and "xeKa" alternation and see what the matchup with DS is. ** When negation immediately precedes the main verb, it will be represented as a left adjunct of that verb. Thus $\sqrt{1}$ (in (65)) will be represented as an adjunct on *jaa* 'go'. In other cases, it is represented as a right adjunct on the verb that it immediately follows i.e. $\sqrt{2}$ is a right adjunct on *jaa* 'go', $\sqrt{4}$ will be a right adjunct on *hai*, and so on. This will connect with the treatment of constituent negation below and will also represent the fact that it is only the left adjunction case that has a non-emphatic/contrastive flavor. The position $\sqrt{3}$, which involves adjunction to the progressive auxiliary *rahaa*, is disfavored - this could be because there are no contextual alternatives that the progressive auxiliary could be contrasted with. Main verbs have many contextual alternatives and the tensed auxiliary has other tensed auxiliaries/the polarity of the sentence as a contextual alternative.

Negation differs from adverbs in that adverbs can only be left adjuncts to VP, VPPred, and V', while Negation can only be an adjunct to V.

15.2 Constituent Negation

There is also constituent negation which can appear on any major category. It seems to be limited to a contrastive 'A not B' usage, where both A and B are overtly expressed.

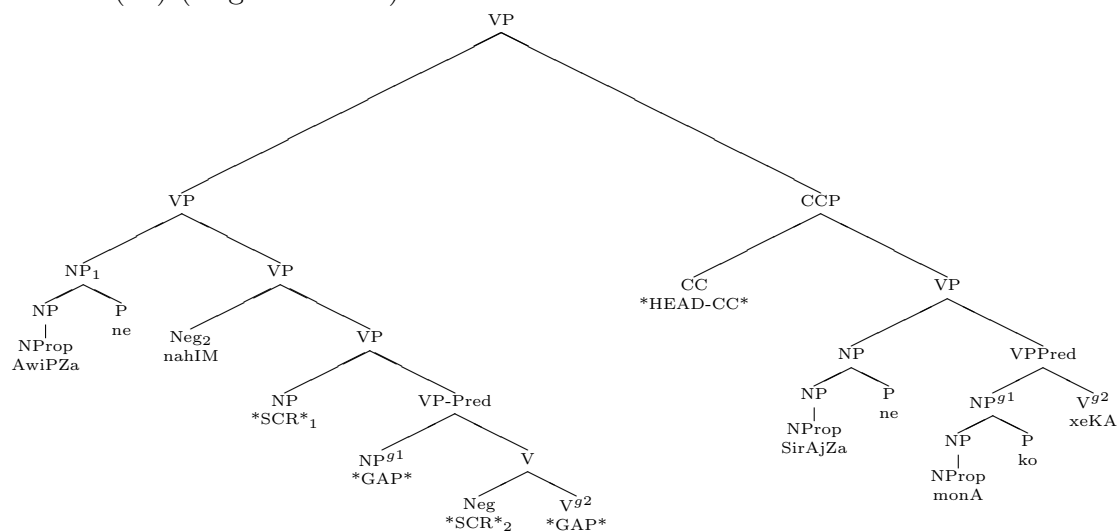
(66) (Negation-PS-4)

Atif ne nahī: Shiraz ne Mona ko dekh-aa
 AwiPZa ne nahIM SirAjZa ne monA ko xeKA
 Atif Erg neg Shiraz Erg Mona Acc see-pfv

Shiraz saw Mona, not Atif (saw Mona).'

We analyze these cases as involving ellipsis of the first clause. For a full discussion, see Chapter 29. We give the tree here, but the analysis is exactly as discussed in Chapter 29.

Tree for (66) (Negation-PS-4)



Chapter 16

Auxiliaries

The following is the maximally articulated auxiliary sequence possible in a Hindi clause.

(67) Main.Verb (Vector.Verb) (Passive.Aux) (Other.Aux) (Modal/Progressive) (Tensed.Aux)

The auxiliary system of Hindi can be described as follows.

| Position | Type | Subcategorizes.For | Morphological.Options |
|----------|-----------------------------|----------------------------|--|
| 1 | Main Verb | | Inf/Asp/Fut/Subjunctive |
| 2 | Vector Verb | Bare.Main.Verb | Inf/Asp/Fut/Subjunctive |
| 3 | Passive Aux | Perfective.Asp | Inf/Asp/Fut/Subjunctive |
| 4.1 | <i>jaa₁, rah</i> | Imperfective.Asp | Inf/Asp/Fut/Subjunctive |
| 4.2 | <i>jaa₂</i> | Indeclinable V- <i>ye</i> | Inf/Asp/Fut/Subjunctive |
| 4.3 | <i>kar</i> | Indeclinable V- <i>yaa</i> | Inf/Asp/Fut/Subjunctive |
| 5.1 | Modal <i>sak</i> | Bare.Main.Verb | Inf/Asp/Fut/Subjunctive |
| 5.2 | Modal <i>cuk</i> | Bare.Main.Verb | Perfective |
| 5.3 | Modal <i>rahaa</i> | Bare.Main.Verb | nothing, counts as Asp |
| 6.1 | Modal <i>par</i> | Inf | Inf/Asp/Fut/Subjunctive |
| 6.2 | Modal <i>caahiye</i> | Inf | nothing, counts as Asp |
| 7 | Aux <i>ho</i> | Inf/Asp | Inf, Asp, Fut, Subjunctive, Present/Past |

For some speakers, the *jaa* auxiliary in 4.2 can take a regular perfective complement that agrees with the subject.

We make a distinction between two kind of auxiliaries: auxiliaries that are either sensitive to argument structure or have an impact on argument structure, and auxiliaries that do not. One auxiliary that has a clear impact on argument structure is the passive auxiliary (Row 3). Vector verbs (Row 2) are also sensitive to argument structure - unaccusative vector verbs typically go with unaccusative main verbs and transitive vector verbs typically go with transitive main verbs. When there is a mismatch, the vector verb has an influence on the case on the subject (ergative or nominative).

The modal auxiliaries in Rows 5 and 6 and the temporal auxiliaries in Row 7 do not interact with the argument structure of the VP that they combine with. This is also the case with the auxiliaries in Row 4 though some of the cases are hard to test due to homophony between the passive auxiliary and the auxiliary *jaa* ‘go’.

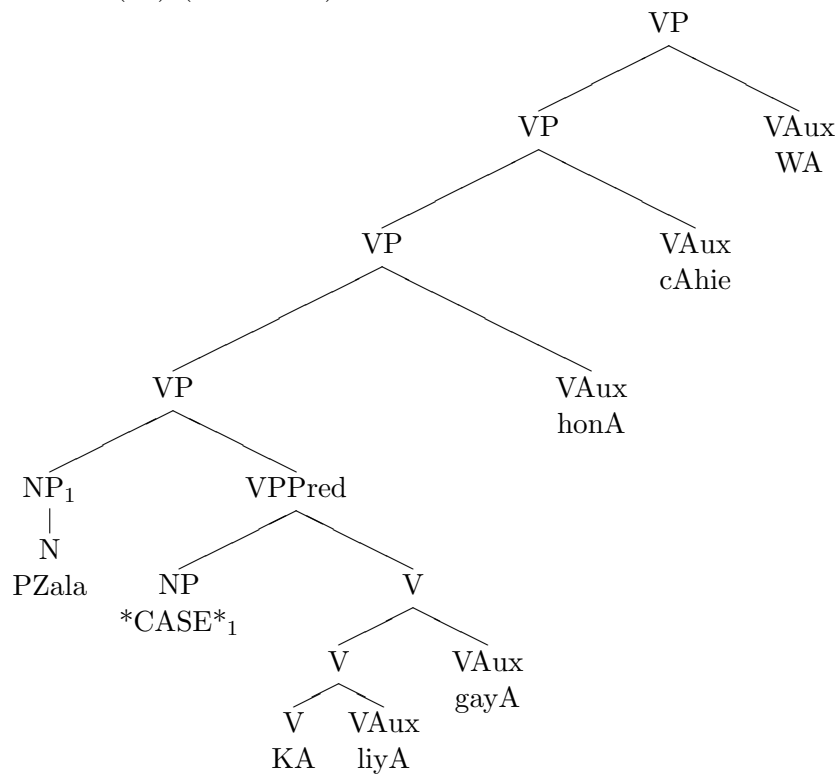
The argument structure difference is also reflected in the relative order of these auxiliaries. The argument structure sensitive auxiliaries (vector verbs, passive auxiliary) precede the rest. We represent this distinction structurally by adjoining the argument structure sensitive auxiliaries directly to the verb. The other auxiliaries are adjoined at the VP level.

(68) (Label:Aux-PS-1)

phal khaa li-yaa gayaa ho-naa caahiye thaa
 PZala KA liyA gayA honA cAhie WA
 fruit eat TAKE-Pfv GO.Pfv be-Inf should be.Pst

‘The fruit should have been eaten.’

Tree for (68) (Aux-PS-1):



Part V

Types of Clauses and Word Order in Clauses

Chapter 17

Passive Voice

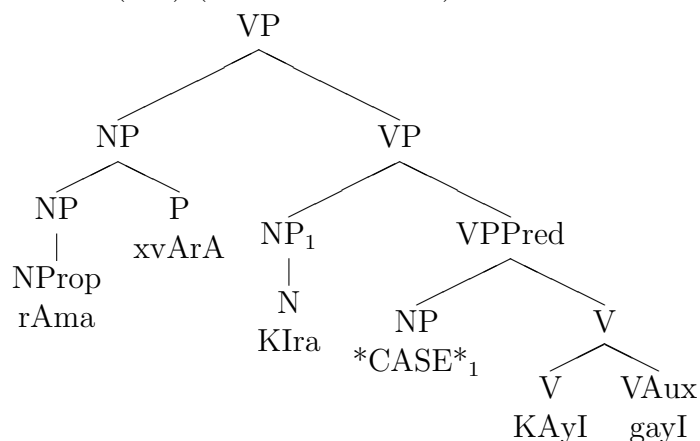
The treatment of passives has three components: the second argument promotes to become the first argument. This is represented by a trace in the second argument position. The external argument becomes optional and is realized as an adjunct on VP. Finally, there is typically a passive auxiliary. Since the passive auxiliary influences voice and argument structure, it is attached directly to the main verb. The analysis is exemplified below.

(69) a. (Relation-DS-k1-7)

Ram dwaaraa khiir khaayii gayii
 rAma xvArA KIra KAyI gayI
 ram by rice-pudding.f eat.Pfv.f GO.Pfv.f

‘Rice-pudding was eaten by Ram.’

Tree for (69a) (Relation-DS-k1-7):

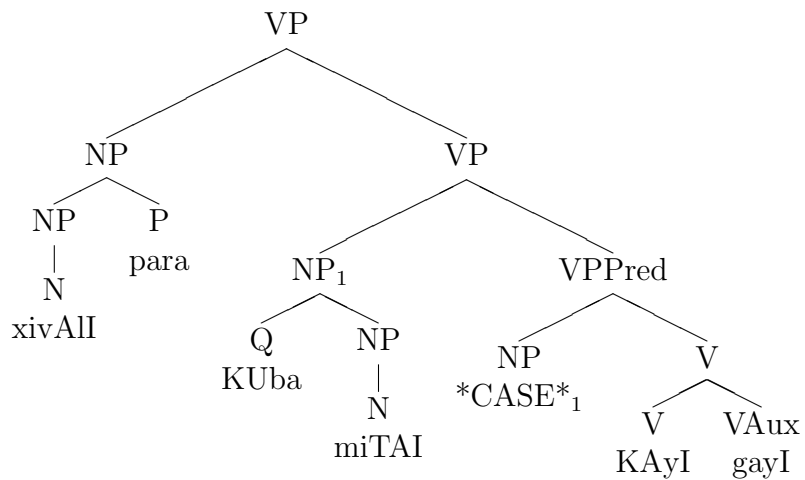


b. (Label:Relation-DS-k2-4)

Diwaalii par khuub miṭhaai khaayii gayii
 xivAlli para KUba miTAI KAyI gayI
 Diwali on lots sweets.f eat.Pfv.f GO.Pfv.f

‘Lots of sweets were eaten on Diwali’

Tree for (69b) (Relation-DS-k2-4):

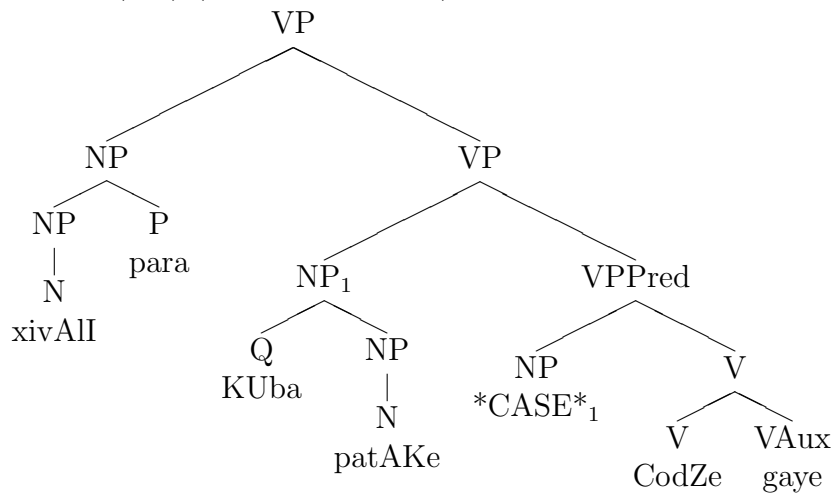


c. (Label:Relation-DS-k2-5)

Diwaalii par khuub paṭaakhe chore gaye
 xivAll para KUba patAKE CodZe gaye
 Diwali on lots crackers.MPl leave.Pfv.MPl go-Pfv.MPl

‘Lots of crackers were burst on Diwali’

Tree for (69c) (Relation-DS-k2-5):



Chapter 18

Causativization

The *-vaa* causative in Hindi-Urdu is very productive and applies on almost all transitive verb stems. Interestingly, it does not result in any increase in arity. This can be seen in the following schematic representation.

- (70) a. Transitive Verb: X Y V, interpretation: X V'd Y
Causative: A X-se Y V-*vaa*
Interpretation: A had Y V'd by X
- b. Ditransitive Verb: X W-Dat Y V
Causative: A X-se W-Dat Y V
Interpretation: A had Y V'd W-Dat by X

One place where this pattern seems to break down is with a class of verbs that have been called *ingesto-reflexive*. They denote physical or metaphorical ingestion of a substance or idea. Ingesto-reflexives come in pairs: a transitive version *V* and a ditransitive *V-aa*. Here are some examples.

- (71) a. *khaa* 'eat' / *khil-aa* 'feed'
b. *pii* 'drink' / *pil-aa* 'cause to drink'
c. *dekh* 'see' / *dikh-aa* 'show'
d. *siikh* 'learn' / *sikh-aa* 'teach'
e. *parh* 'read' / *parh-aa* 'teach'

For morphophonological reasons, the *-vaa* causative of both a transitive ingesto-reflexive and a ditransitive ingesto-reflexive sometimes ends up looking the same; the two causatives still have distinct argument structures.

- (72) *-vaa* causatives of *parh* 'read' / *parh-aa* 'teach' = *parhvaa*
- a. *-vaa* causative of *parh* 'read'
- Ram ne bacce se kitaab parhvaayii
rAma ne bacce se kiwAba paDZvAyI
Ram Erg child Instr book.f read.CAUS.Pfv.f

‘Ram had the book read by the child.’

b. *-vaa* causative of *paṛh-aa* ‘teach’

Ram ne teacher se bacce ko itihaas paṛhvaayaa
 rAma ne tIcara se bacce ko iwihAsa paDZvAyA
 Ram Erg teacher Instr child Dat book.f read.CAUS.Pfv.MSg

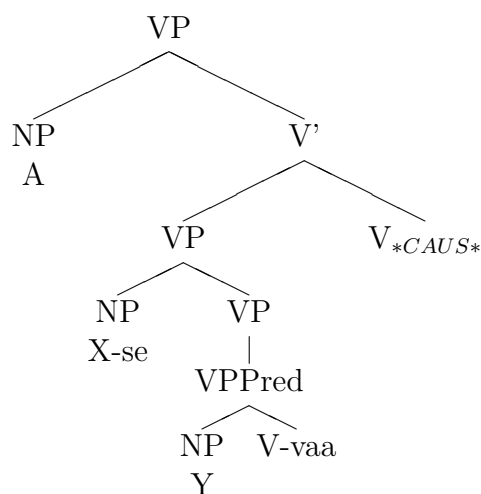
‘Ram had the child taught history by the teacher.’

It is not the case that *-vaa* causatives based on all ingesto-reflexive roots are ambiguous. When the *-vaa* causative is unambiguous, it is based on the ditransitive variant.

18.1 Initial Non-Binary Branching Analysis

The PS treatment initially adopted for *-vaa* causatives embedded a passive-like VP under a VP headed by an abstract causative head which introduced the external argument. The intermediate agent was adjoined to the embedded VP. The lower VP was passive-like in having no external argument; however, the object did not promote to the first position, as schematized below.

(73) *-vaa* Causatives: Previous analysis
 A X-se Y V-*vaa*



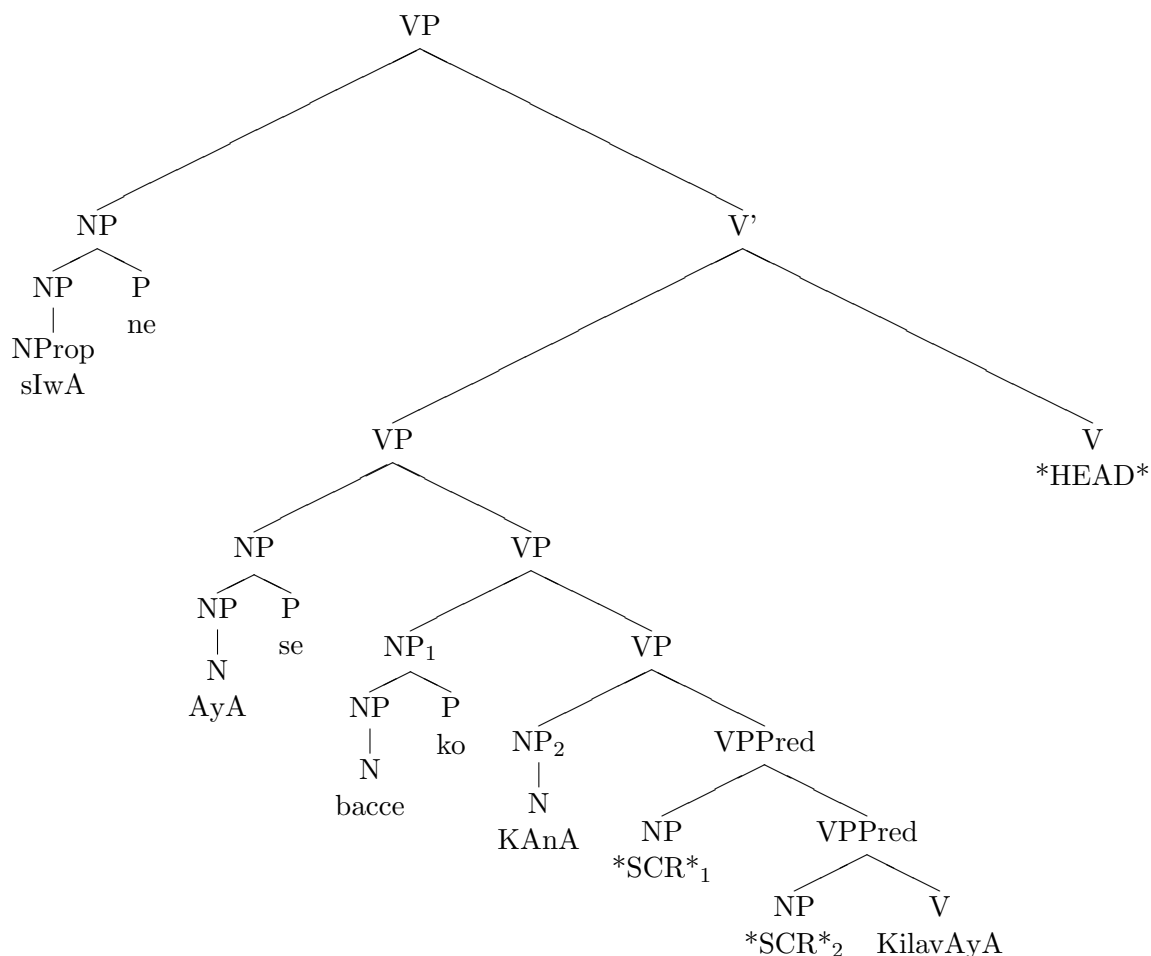
Under this analysis, the tree for the example sentence in (74) looked as follows:

(74)

(Label:Relation-PS-pk1-2)

Sita ne aayaa se bacce ko khaanaa khilvaayaa
 sIwA ne AyA se bacce ko KAnA KilavAyA
 Sita erg maid by child Dat food eat.Caus.Pfv

'Sita had the child fed through the maid.'

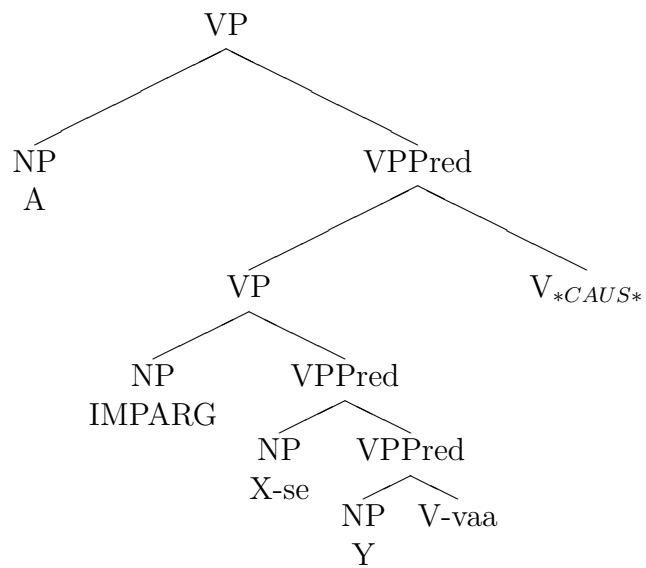


In this treatment, the lack of change in arity in the causativized version of the verb was reflected in the projection of VP above VPPred, indicating the presence of a (silent) agent of the embedded VP. Because the specifier of VP was not projected, this analysis resulted in non-binary branching, since the daughter of VP, VPPred, had more than one daughter.

18.2 Revised Binary Branching Analysis

The revised analysis eliminates this confound by uniformly projecting a null implicit argument in the specifier position of the lower VP, as schematized below. Note that the instrument phrases, if present, adjoin to VPPred, below the implicit argument, in accordance with our general convention of attaching adjuncts as low as possible within the clause.

(75) -vaa Causatives: Non-unary branching analysis



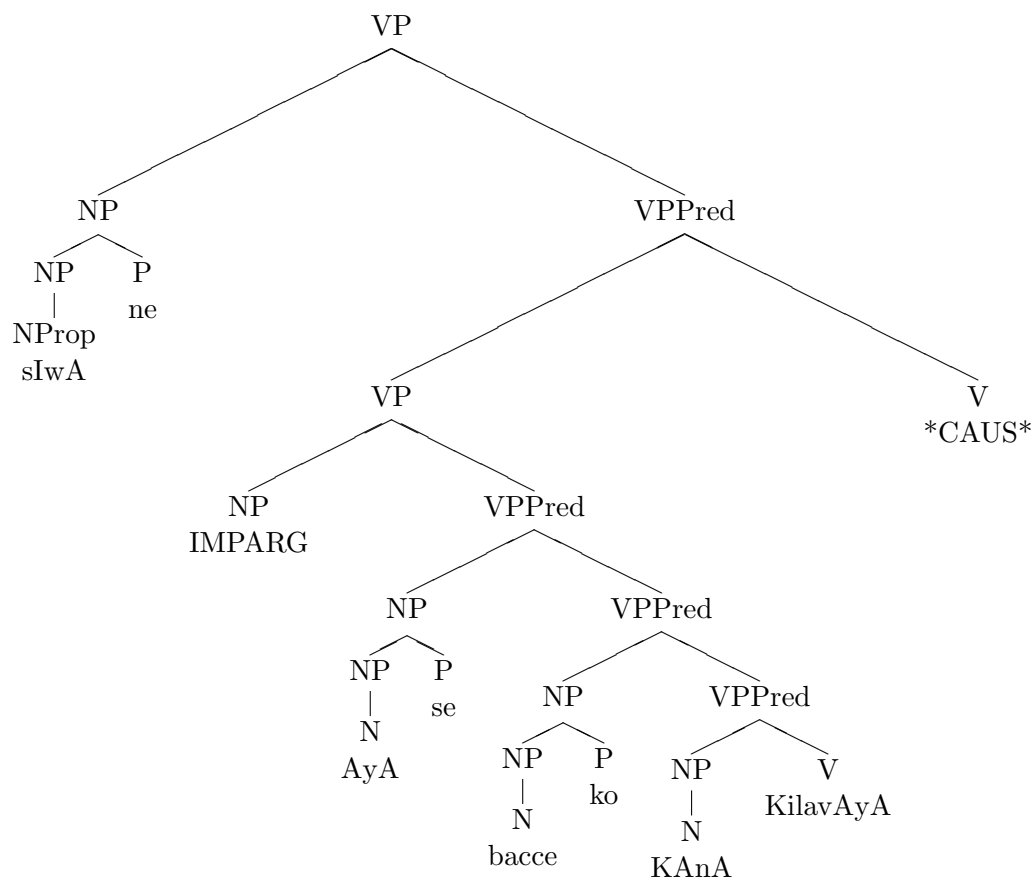
The tree for the example sentence (76) above, repeated below, now looks as follows:

(76) (Relation-PS-pk1-2)

Sita ne aayaa se bacce ko khaanaa khilvaayaa
 sIwA ne AyA se bacce ko KAnA KilavAyA
 Sita erg maid by child Dat food eat.Caus.Pfv

‘Sita had the child fed through the maid.’

Tree for (76) (sent-Relation-PS-pk1-2):



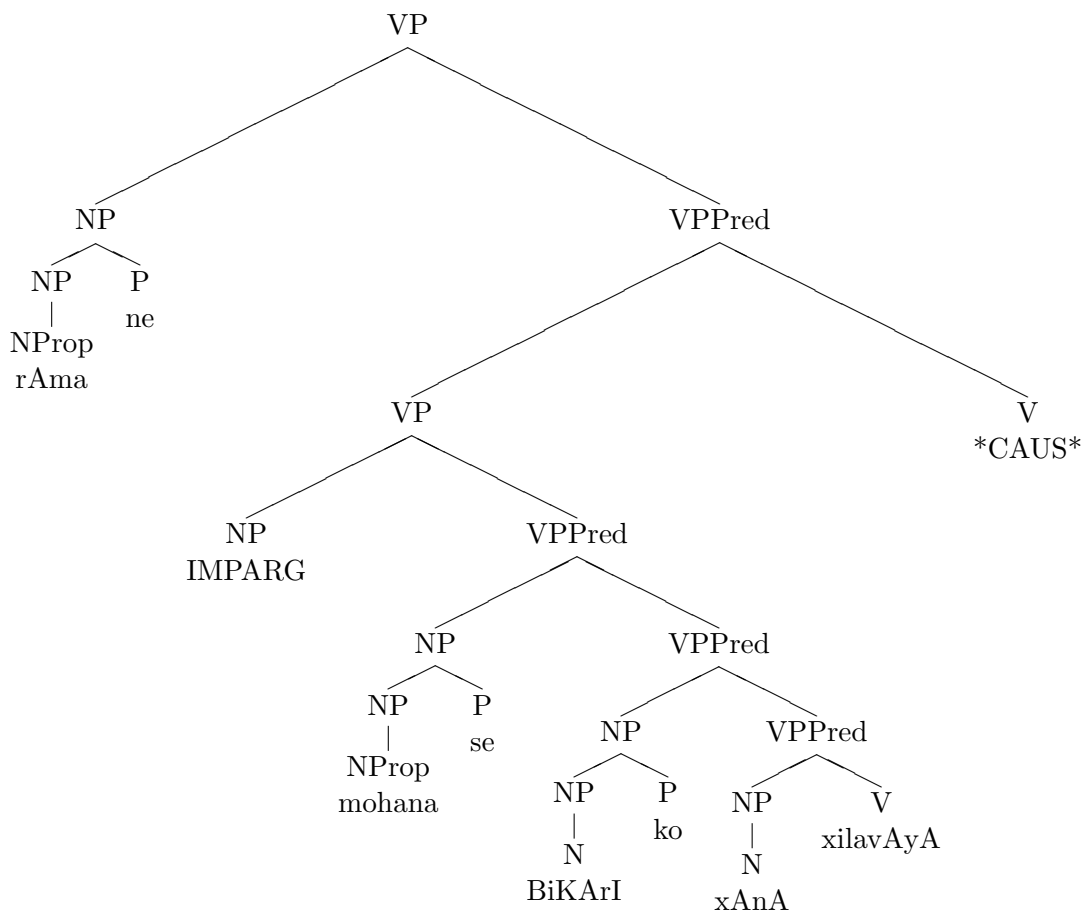
A further example is in (77a).

(77) a. Relation-DS-pk1-3

rAma ne mohana se BiKArI ko xAna xilavAyA
 Ram erg Mohan by beggar acc food caused-to-give

‘Ram made Mohan give the alms to the beggar.’

Tree for (77a) (Relation-DS-pk1-3):

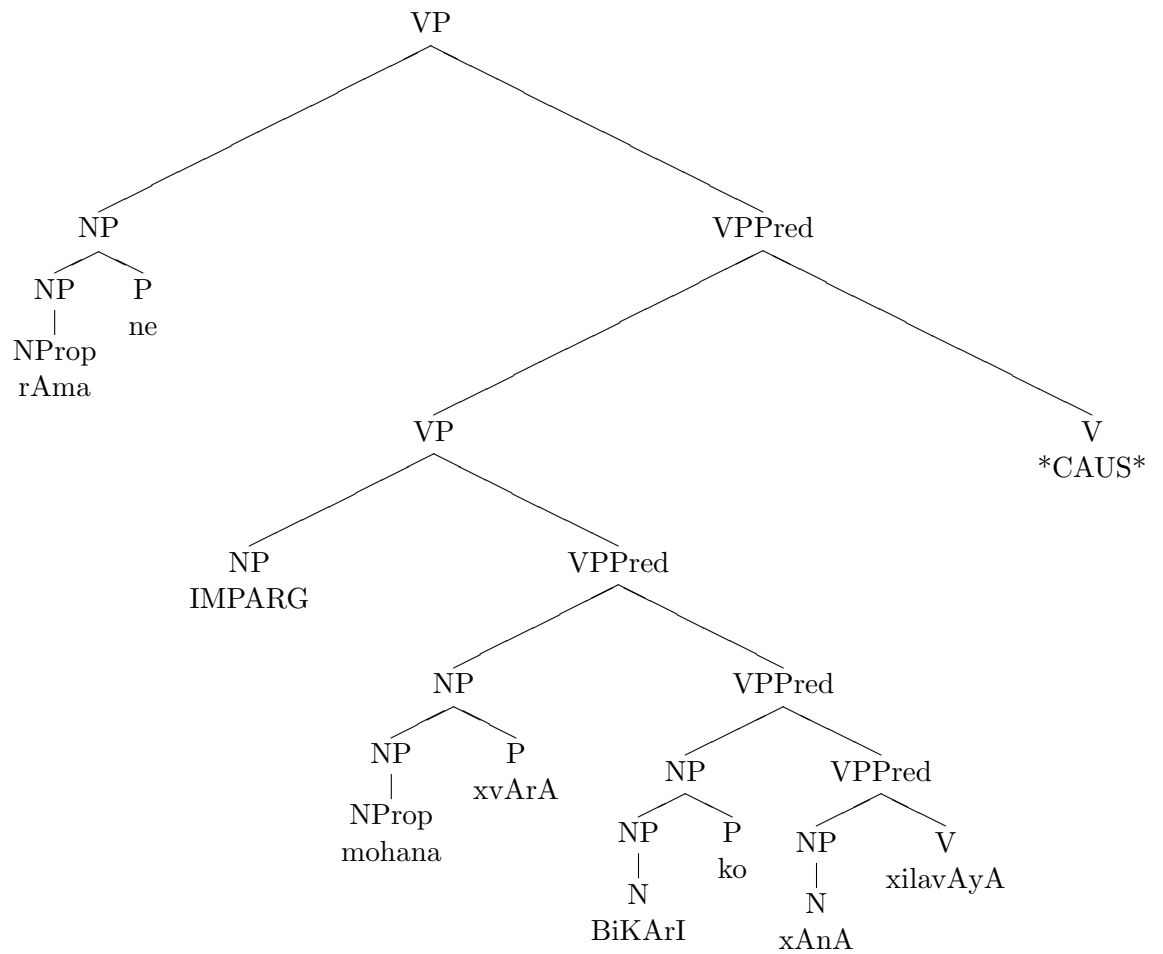


b. (Label: sent-Relation-PS-pk1-31)

Ram ne Mohan dwaaraa bhikhaarii ko daanaa dilvaayaa
 rAma ne mohana xvArA BiKArI ko xAna xilavAyA
 Ram erg Mohan by beggar Dat grain give.Caus.Pfv

‘Ram made Mohan give the alms to the beggar.’

Tree for (77b) (Label: sent-Relation-PS-pk1-31):

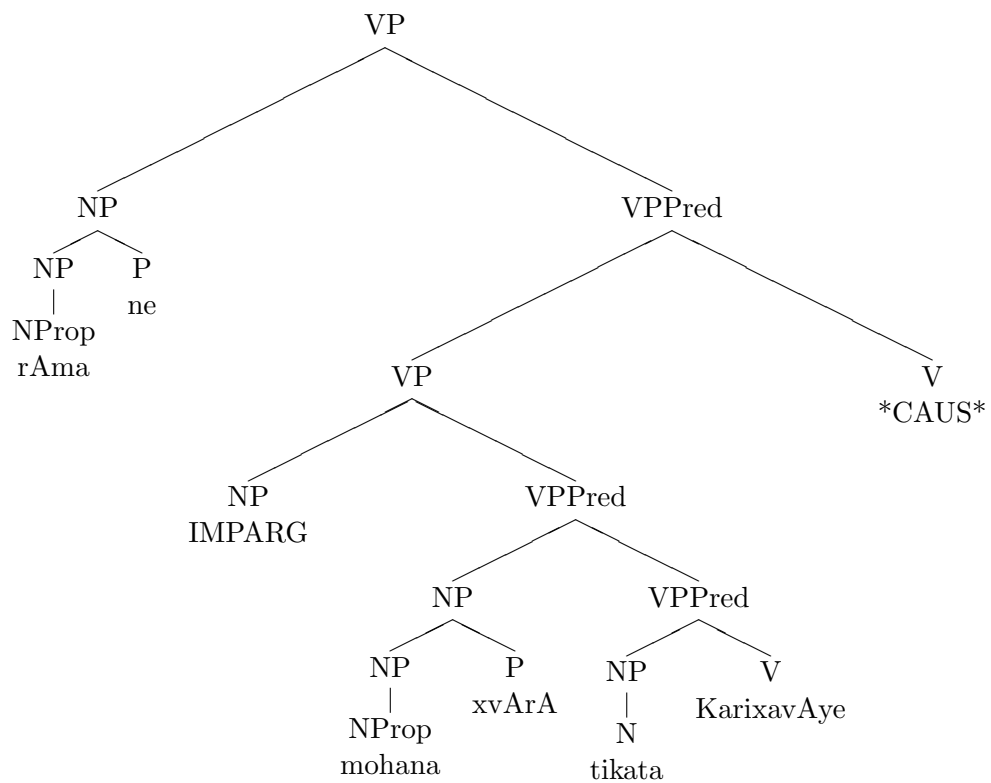


c. (Label:Relation-PS-jk1-2)

Ram ne Mohan dwaaraa tikaṭ kharidvaaye
 rAma ne mohana xvArA tikaṭa KarixavAye
 Ram erg Mohan by ticket buy.Cause.Pfv

‘Ram made Mohan buy tickets for Raja.’

Tree for (77c) (Relation-PS-jk1-2):

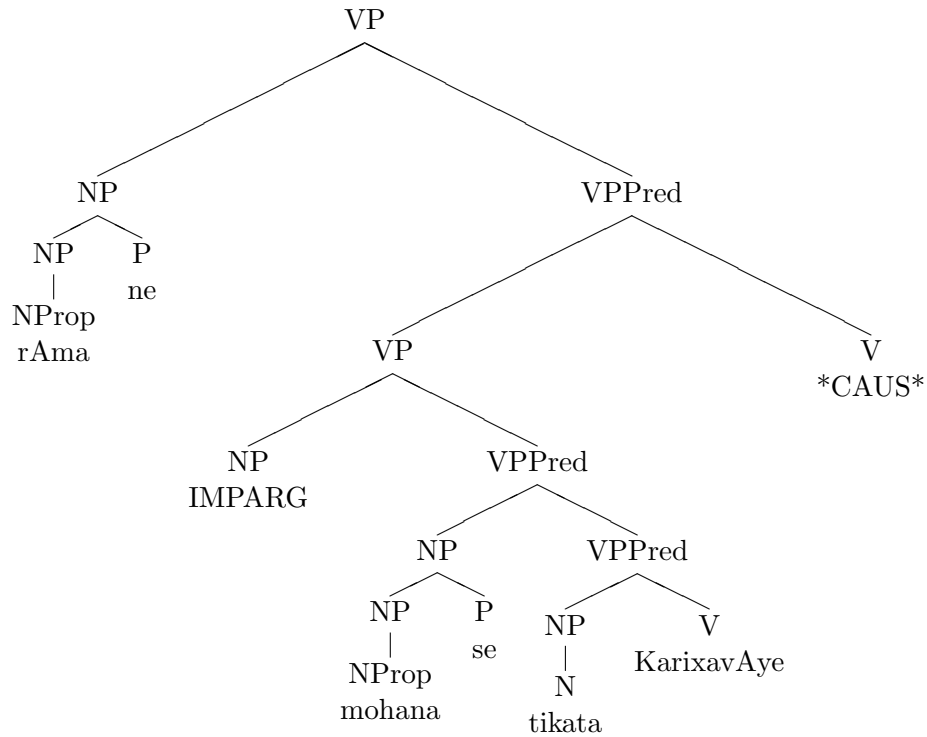


d. (Label:Relation-PS-jk1-21)

Ram ne Mohan se ṭikaṭ kharidvaaye
 rAma ne mohana se tikata KarixavAye
 Ram erg Mohan INSTR ticket buy.Cause.Pfv

‘Ram made Mohan buy tickets for Raja.’

Tree for (77d) (Relation-PS-jk1-21):



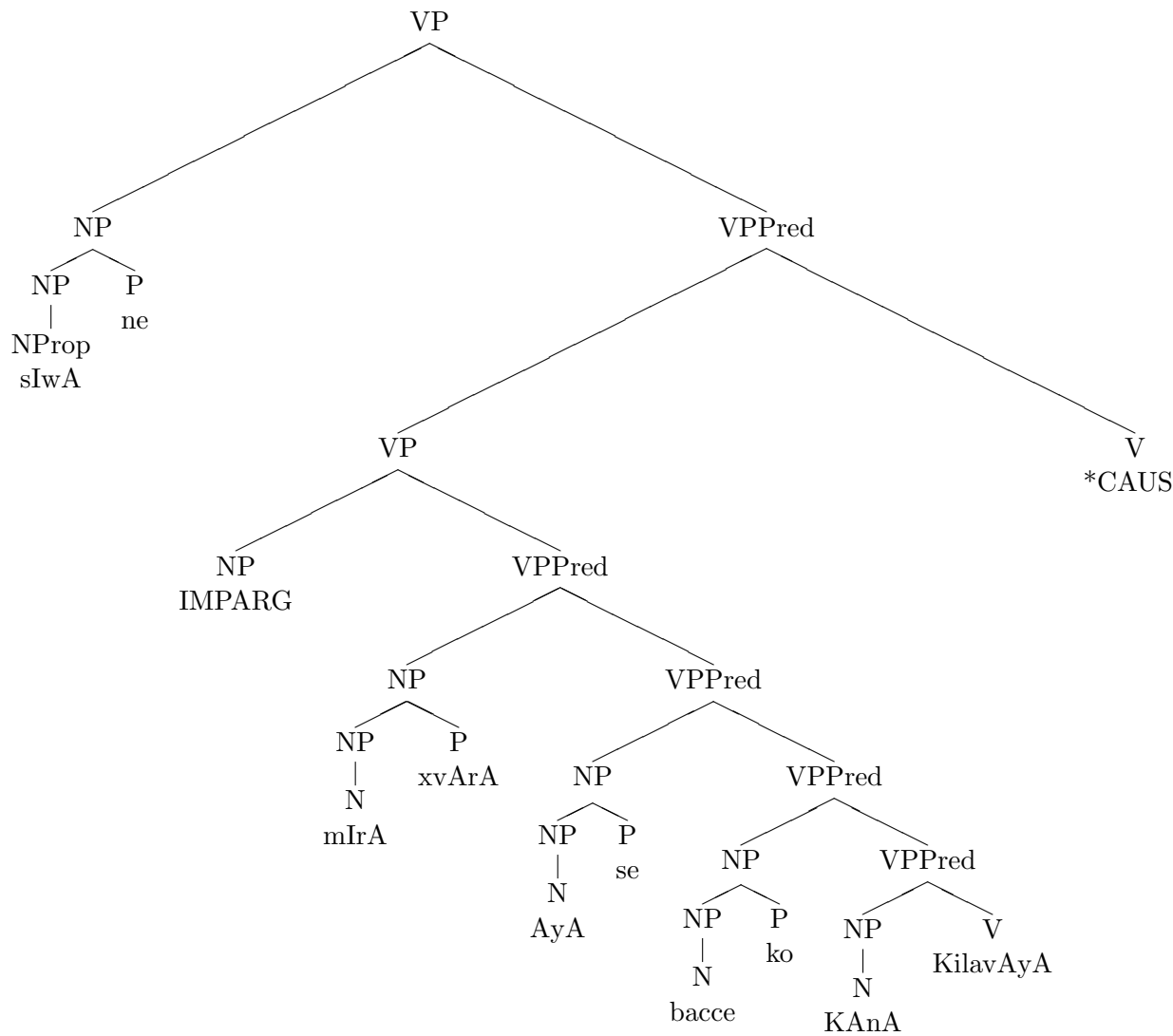
Instrumental phrases can be recursed. When this happens, typically one of the instrumental phrases is realized with *dwaaraa* ‘through’ and the other through *se*, the instrumental marker. The same marker may also be repeated though this is dispreferred.

(78) a. (Label:Relation-mk1-2)

Sita ne Mira dwaaraa aayaa se bacce ko khaanaa khilvaayaa
 sIwA ne mIrA xvArA AyA se bacce ko KAnA KilavAyA
 Sita erg Mira through maid by child Dat food eat.Caus.Pfv

‘Sita had the child fed through the maid through (the agency of) Mira.’

Tree for (78a) (Relation-mk1-2):

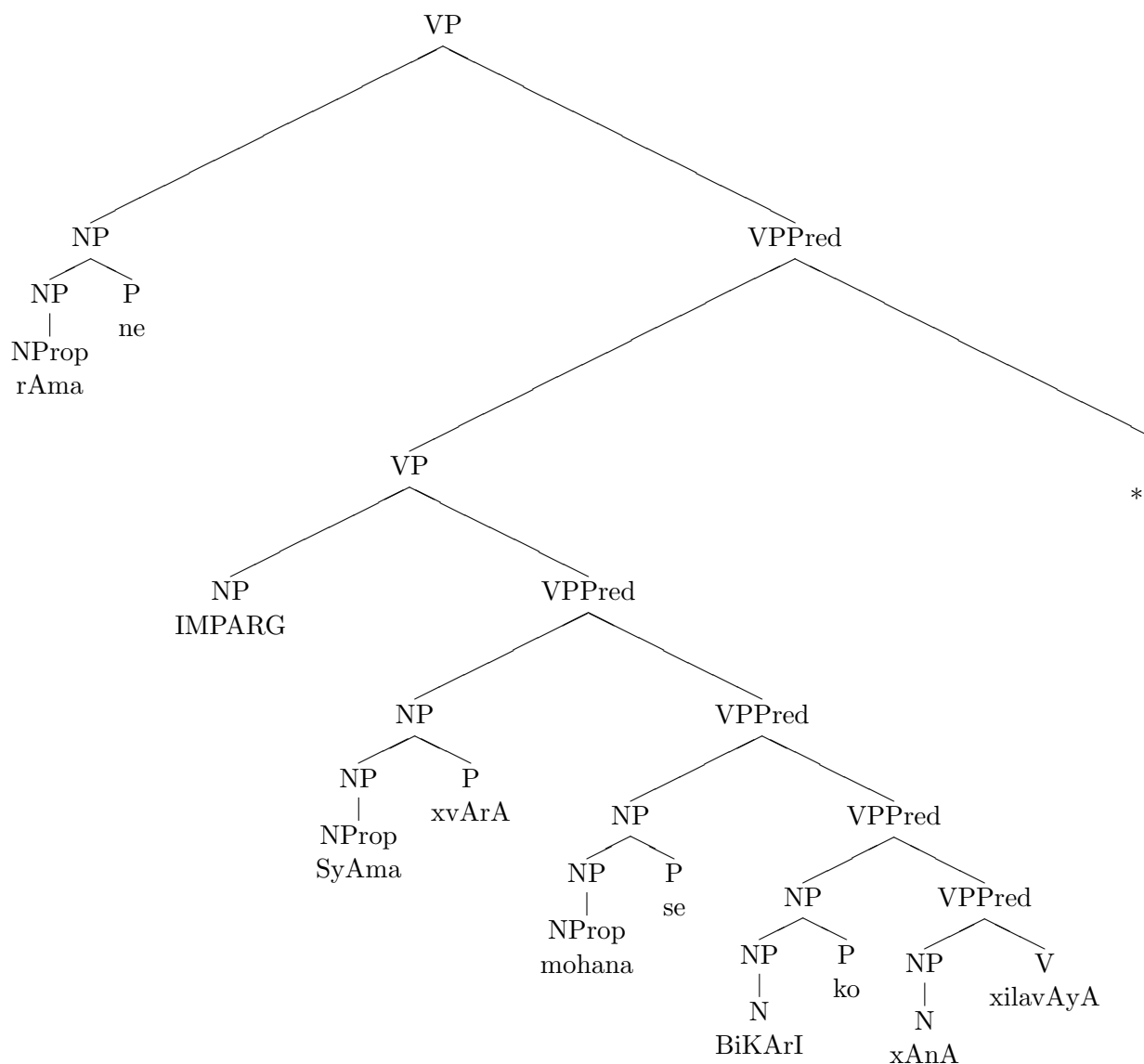


b. (Label:Relation-PS-mk1-3)

Ram ne Shyam dwaaraa Mohan se bhikhaarii ko daanaa dilvaayaa
 rAma ne SyAma xvArA mohana se BiKArI ko xAna xilavAyA
 Ram erg Shyam through Mohan by beggar Dat grain give.Caus.Pfv

‘Ram made Mohan give the alms to the beggar through the agency of Shyam.’

Tree for (78b) (Relation-PS-mk1-3):



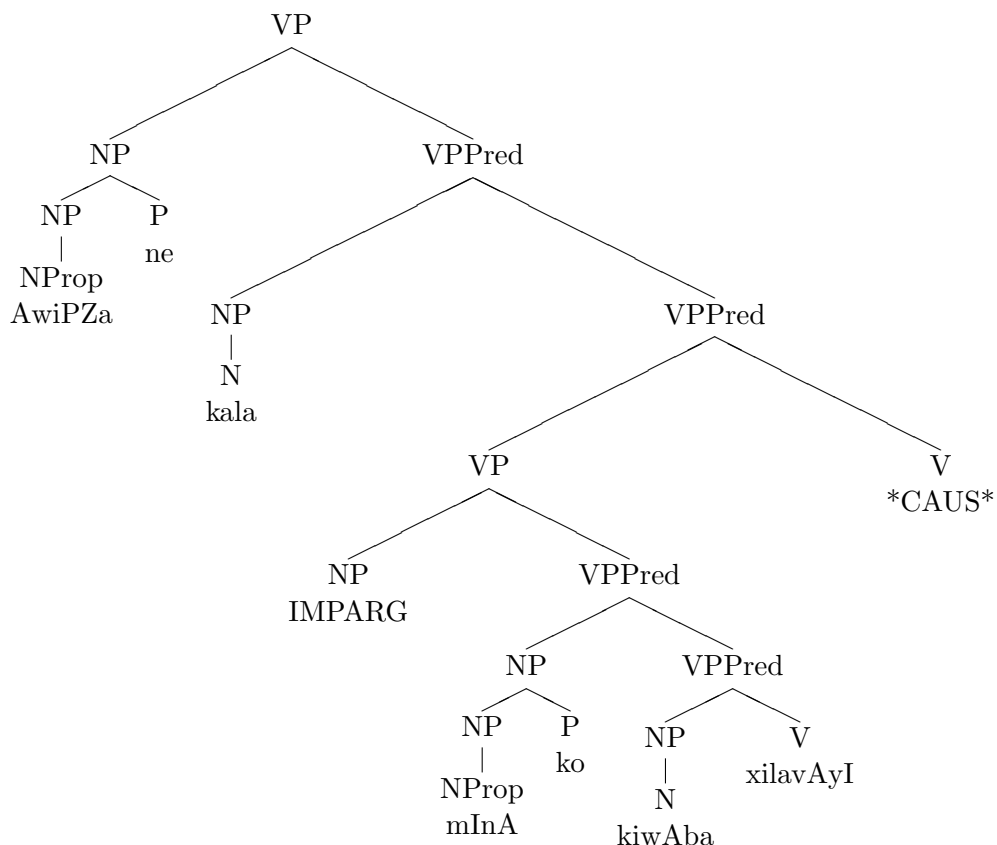
Interactions with adjuncts

(79) a. (Label:Causative-PS-1)

Atif ne kal Mina ko kitaab dilvaayii
 AwiPZa ne kala mInA ko kiwAba xilavAyI
 Atif Erg yesterday Mina Dat book.Sg give.Caus.Pfv.F.Sg

‘Atif caused Mina to buy a book yesterday.’

Tree for (79a) (Causative-PS-1):



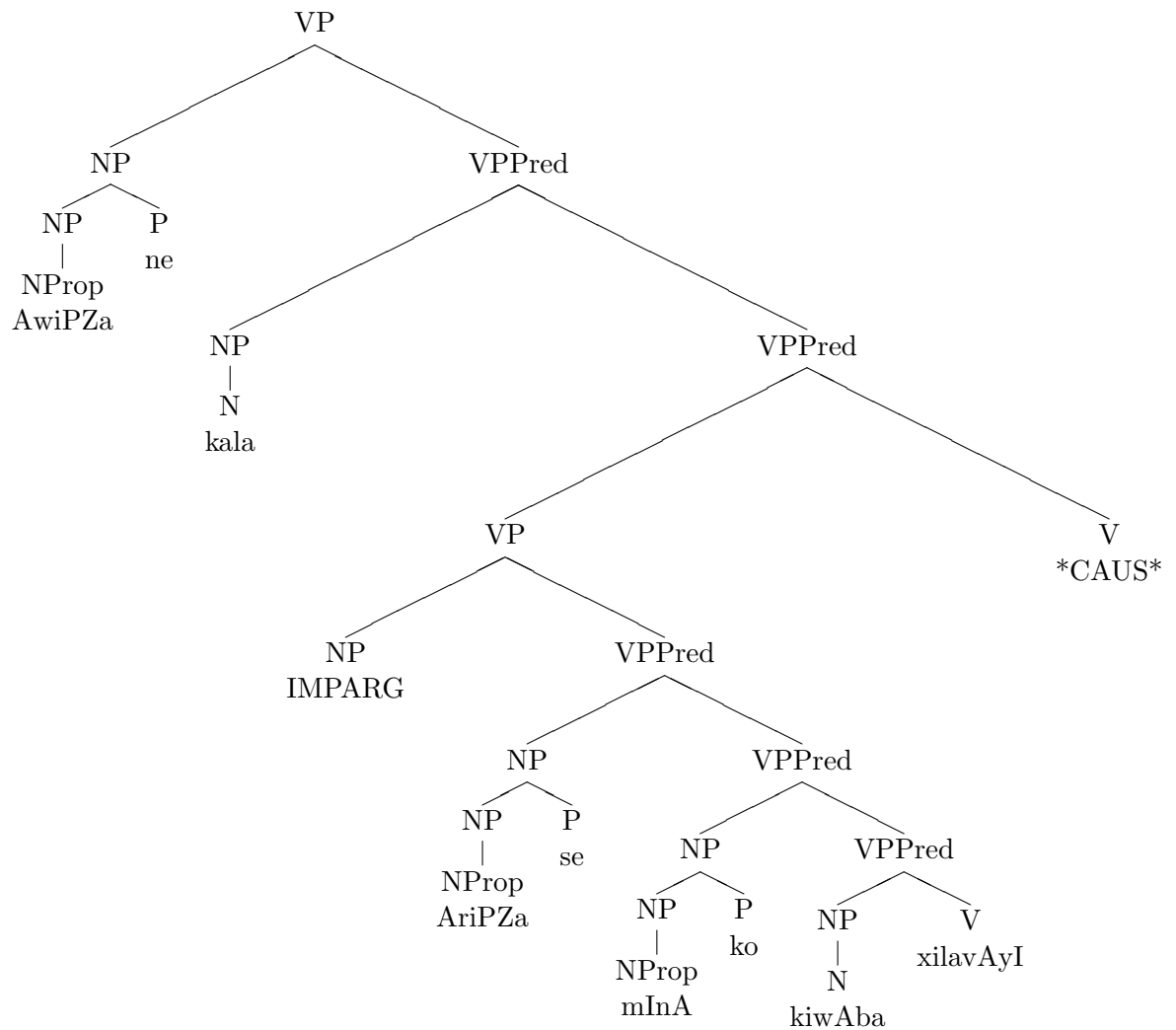
Note that the adverb is attached as high as possible consistent with the word order, hence at the upper VPPred and not at the VP complement of CAUS.

b. (Label:Causative-PS-2)

| | | | | | | | | |
|--------|-----|-----------|--------|-------|------|-----|----------|--------------------|
| atif | ne | kal | Arif | se | Mina | ko | kitaab | dilvaayiii |
| AwiPZa | ne | kala | Arif | se | mInA | ko | kiwAba | xilavAyI |
| Atif | Erg | yesterday | AriPZa | instr | Mina | Dat | book.FSg | give.Caus.Pfv.F.Sg |

‘Atif caused Arif to make Mina buy a book yesterday.’

Tree for (79b) (Causative-PS-2):



18.3 Transitivity Alternations that we do Not Handle

We do not represent the following alternations:

- (80) a. Transitivity Alternations of Unergatives
 (*nahaa* ‘bathe’, *nahlaa* ‘make someone bathe’)
- b. Ditransitivization of Ingesto-reflexives
 (*dekh* ‘see’, *dikhaa* ‘show’)

The transitivity alternation of unaccusatives is represented; the intransitive has its unique argument moving to the first argument position. In the corresponding transitive, an external argument would occupy the first position and the internal argument would stay in the second position.

We also do not represent the spray-load alternation.

- (81) a. Ram ne truck bhuuse se bharii
 rAma ne traka BUse se BarI
 Ram Erg truck hay with fill.Pfv.f
 ‘Ram filled the truck with hay.’
- b. Ram ne bhuusaa truck me bharaa
 rAma ne BUa traka meM BarA
 Ram Erg hay truck in fill.Pfv.MSg
 ‘Ram filled hay on the truck.’

Here the verb form stays the same but the argument structure changes. These two distinct argument structures would be given essentially the same representation in the phrase structure.

We do handle all of the above examples; it’s just that our representations do not explicitly represent the systematic relationships that hold between the pairs.

Chapter 19

Clausal Modifiers of Nouns

Hindi has the following clausal modifiers of nouns:

- (82)
- a. English style finite headed relative clauses, immediately postnominal
 - b. English style finite headed relative clauses, extraposed
 - c. Correlatives, typically clause initial, can be associated with more than one NP, and can be discontinuous from these NPs
 - d. Extraposed relative clauses, finite
 - e. Participial relative clauses, prenominal
 - f. Infinitival *vaalaa* relative clauses, prenominal

Correlative constructions are also used to construct comparatives, conditionals, *when*, *until*, and *since* clauses.

An important question is whether the above ‘constructions’ need to be given any kind of special status. Syntactically they seem to be just plain old correlatives.

Another question concerns the category of relative clauses.

On the surface, relative clauses for the most part do not involve a C-layer. The relative phrase is typically fronted but not always. Here is a case where a C-element appears:

- (83) vo kitaab [jo₁ ki tumhē t₁ pasand hai]
vo kiwAba jo ki wumheM pasanxa hE
that book Rel that you.Dat like be.Prs.Sg
‘The book that you like.’

When there is a C-element, it always follows the relative pronouns. There is no consensus about the treatment of this C-element. We could assume that it is in C⁰ and the relative phrase has moved to [Spec,CP] or else we could assume that relative pronouns have an optional realization with a suffixal *-ki*. Interestingly this suffixal C⁰ only appears in headed relative clauses, perhaps indicating the subordinate nature of such clauses. For now, we will

treat all relative clauses as CPs. In most cases the C^0 head will be empty. Given the fact that relative pronouns always follow the C-element, if it is present, one option would be to maintain uniformity between the cases where the C-element and those where it is not, by assuming that fronted elements move to [Spec,CP] even when such movement is string vacuous in the case of subject relatives. We will not adopt this option; rather, we show movement to Spec, CP only when such movement is visible on the surface—movement of subject relative pronouns to Spec,CP when the C is overt, and movement of object relative pronouns.

19.1 English Style Headed Relative Clauses

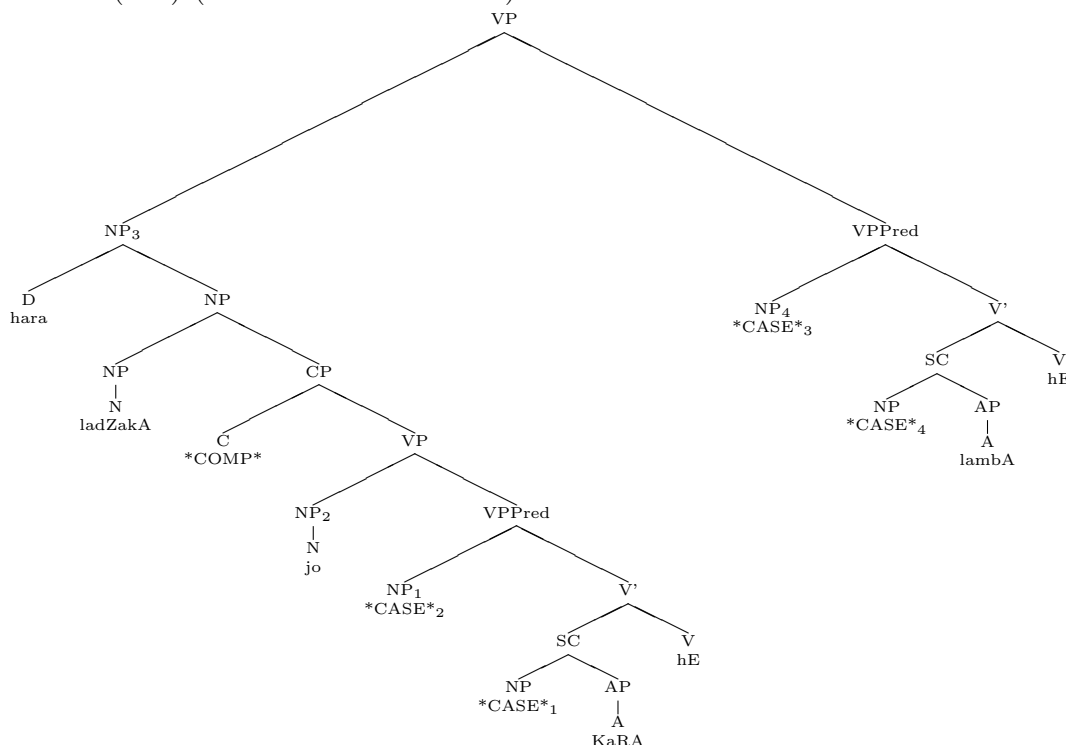
There will be two kinds of relative clauses: NP-internal and extraposed relative clauses. As discussed, extraposition will be represented by a trace adjoined to the source NP.

(84) NP-Internal Relative Clause:

a. (Relative-Clauses-PS-1)

har laṛkaa [jo khaṛaa hai] lambaa hai
 hara ladZakA jo KadZA hE lambA hE
 every boy who standing.MSg be.Prs.Sg tall.MSg be.Prs.Sg
 ‘Every boy who is standing is tall.’

Tree for (84a) (Relative-Clauses-PS-1)

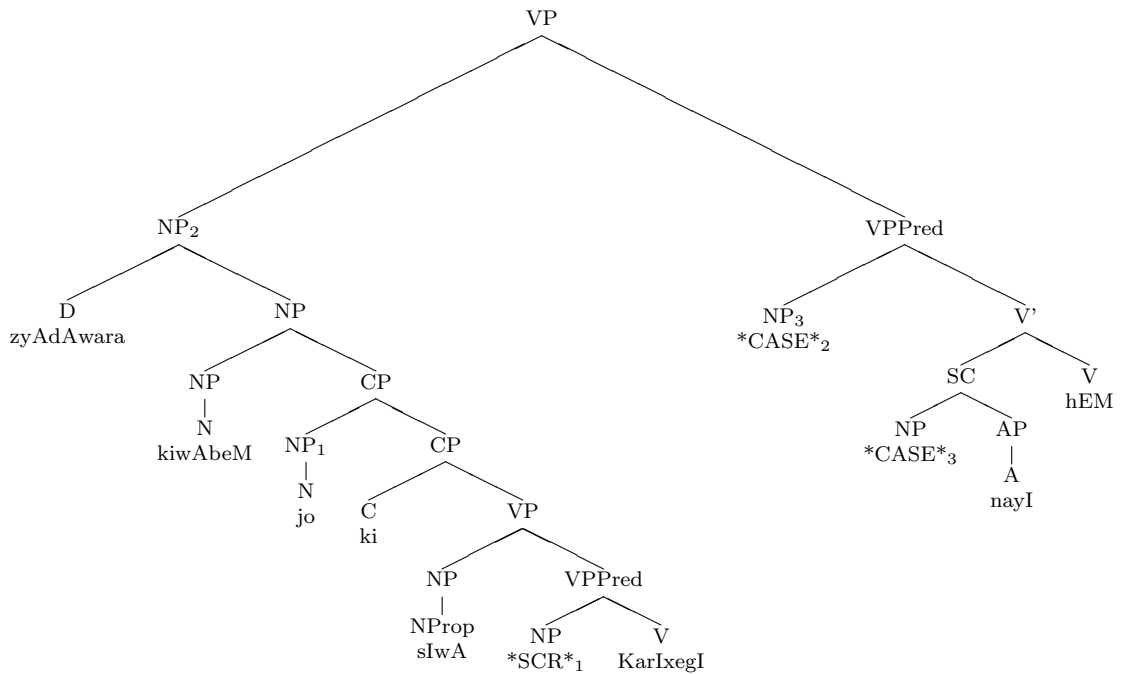


b. (Relative-Clauses-PS-2)

zyAxaatar kitaabē [jo₁ (ki) Sita t₁ khariid-egii] nayii hĒ
 jyAxAwara kiwAba jo (ki) sIwA KarIxeGI naI hogI
 most books.f Rel that Sita buy-Fut.F new are

‘Most books that Sita will buy are new.’

Tree for (84b) (Relative-Clauses-PS-2)

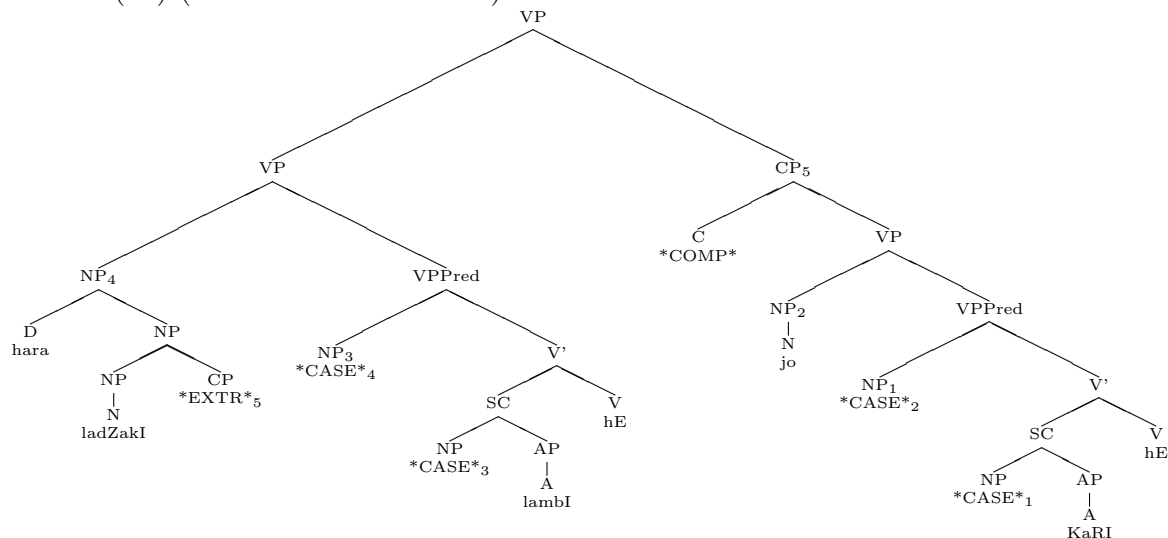


(85) (Relative-Clauses-PS-3) Extraposed Relative Clauses

har lar̄kii lambii hai [jo khar̄ii hai]
 hara ladZakI lambI hE jo KadZI hE
 every girl tall is Rel standing.f is

‘Every girl is tall who is standing.’

Tree for (85) (Relative-Clauses-PS-3)



19.2 Correlatives

Correlatives appear to the left of the NP that they modify. If immediately left adjacent to the NP they modify, they are attached to the NP. Otherwise they are adjoined to the finite clause they appear in. The semantic relationship between the demonstrative phrase and a non-adjacent correlative that modifies it is indicated by movement. The correlative starts off adjoined to the demonstrative phrase that it modifies and is moved from there to its surface position.¹

We will treat correlatives as CPs with a silent C^0 head. Unlike relative clauses, we will display displacement of the relative phrase only when forced to do so by word order. The optional movement of the relative phrase will be to a position below the silent C^0 head.

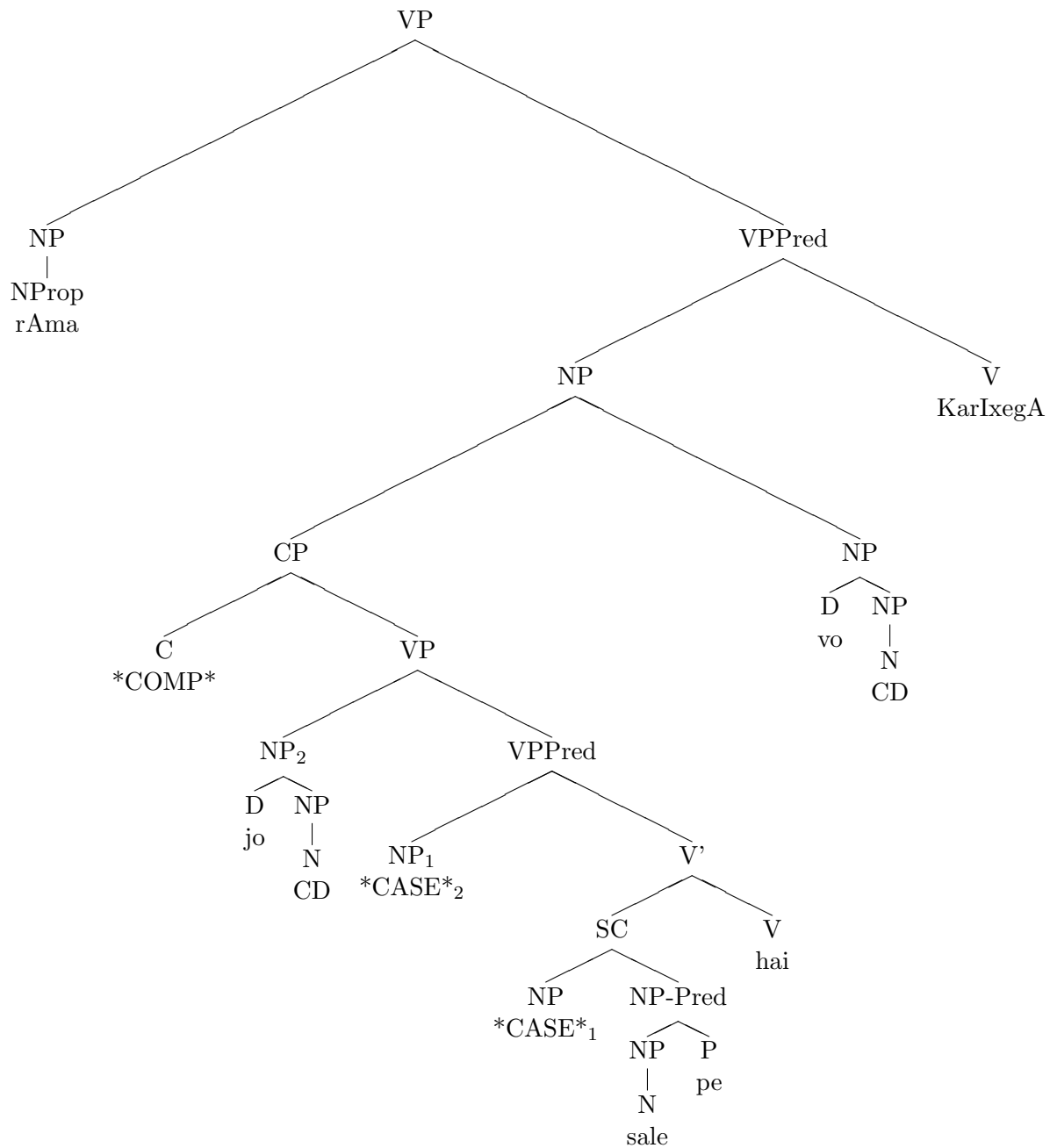
- (86) a. (Relative-Clauses-PS-4) adjacent correlative:

Ram [[jo CD sale pe hai] [vo CD]] khariid-egaa
 rAma jo CD sela pe hE vo CD KarIxeGAA
 Ram Rel CD sale on is that CD buy-Fut.MSg

‘The CD that is on sale, Ram will buy that CD.’

Tree for (86a) (Relative-Clauses-PS-4)

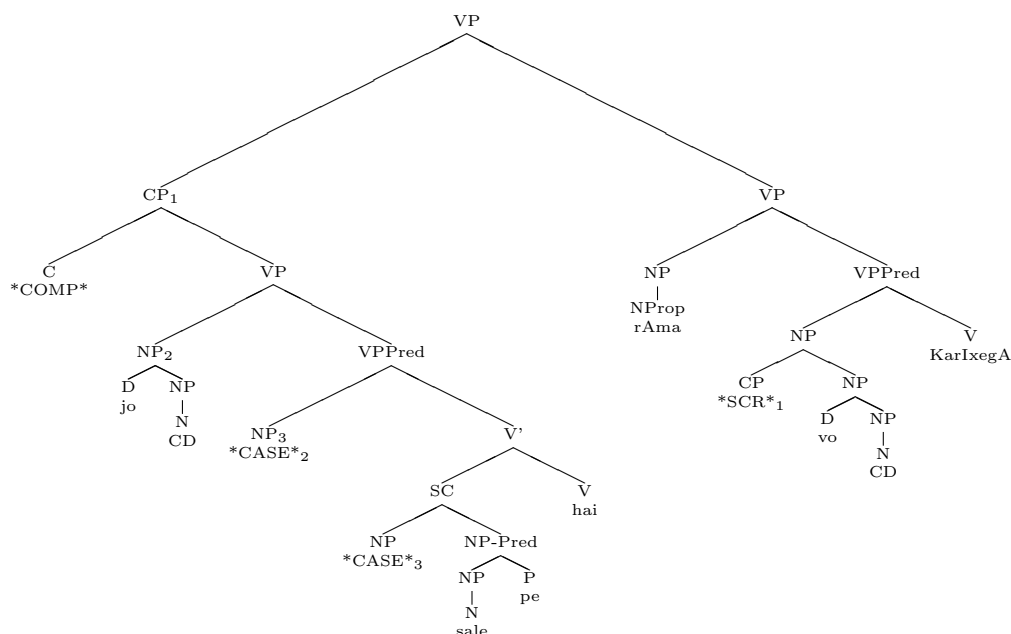
¹The semantic relationship between a multi-head correlative and its demonstrative phrases cannot be represented by movement. It needs a more abstract semantic representation which we omit in the interest of simplicity.



The CP is directly adjoined to the NP it modifies.

- b. (Relative-Clauses-PS-5) non-adjacent correlative:
 [jo CD sale pe hai] [Ram vo CD khariid-egaa]
 jo CD sela pe hE rAma vo CD KarIxegA
 Rel CD sale on is Ram that CD buy-Fut.MSg
 'The CD that is on sale, Ram will buy that CD.'

Tree for (86b) (Relative-Clauses-PS-5)



When the CP and the demonstrative phrase that it restricts are non-adjacent, we indicate the relationship between them via movement. We assume that the CP always originates as a left-adjunct on the NP it modifies; in cases where they are non-adjacent on the surface, there will be a *SCR* trace in the CP's position of origin.

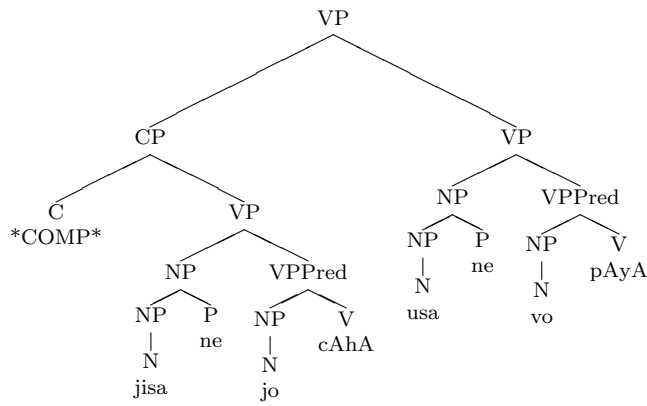
Multi-head correlatives constitute one class of cases where it is not feasible to indicate the relationship between the CP and the demonstrative phrase with a movement relation. In these cases we could indicate the relationship between them via semantic coindexation, but we choose not for this treebank for reasons of simplicity.

- c. (Relative-Clauses-PS-6) multi-head correlative:

[jis ne jo caah-aa] [us ne vo paa-yaa]
 jisa ne jo cAhA usa ne vo pAyA
 Rel Erg Rel want-Pfv Dem Erg Dem get-Pfv

‘Who desired what, he got that.’ (Everyone got what they desired.)

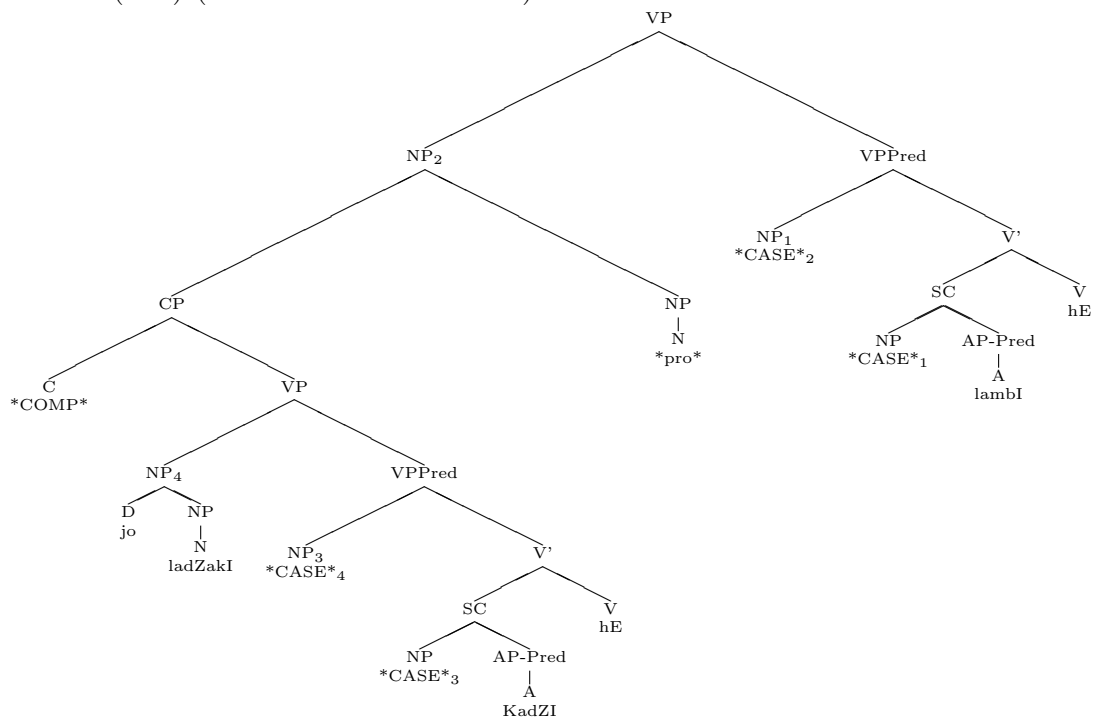
Tree for (86c) (Relative-Clauses-PS-6)



d. (Relative-Clauses-PS-41) correlative with null demonstrative phrase:

[jo laṛkii kharii hai] lambii hai
 jo ladZakI KadZI hE lambI hE
 Rel girl standing.f be.Prs.Sg tall.f be.Prs.Sg
 ‘The girl who is standing is tall.’

Tree for (86d) (Relative-Clauses-PS-41)



19.3 Participial Relative Clauses

A past participial relative clause targets ‘lower’ arguments that have moved upwards through passivization or other syntactic processes (other than scrambling): subjects of unaccusatives, objects of transitives. A past participle relative clause is not possible with an unergative verb.

(87) Participial relative clauses

- a. (Participial-Relative-Clauses-PS-1) Transitive: these are passive structures, we will use our passive analysis for these.

[khaa-ye hue] phal kacche the
 KAye hue Pala kacce We
 eat-Pfv.Mpl be.Pfv.MPl fruit unripe.MPl be.Pst.MPl

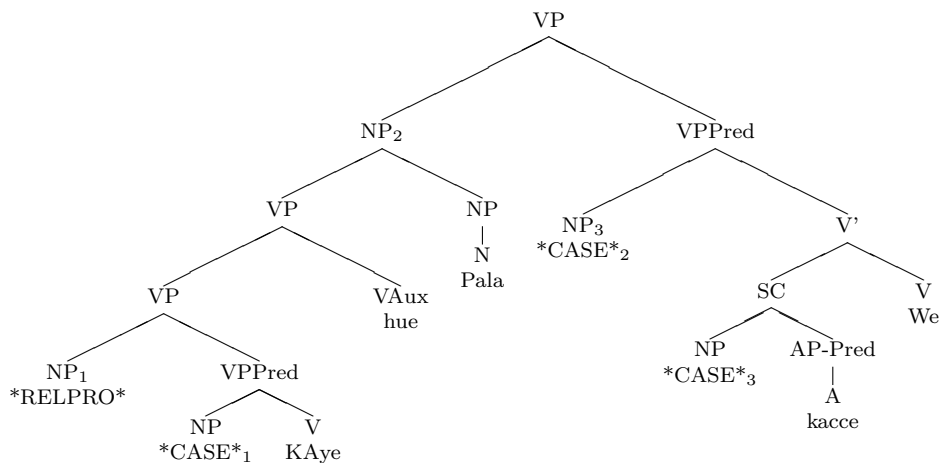
‘The eaten fruits were unripe.’

- b. (Participial-Relative-Clauses-PS-1) Unaccusatives:

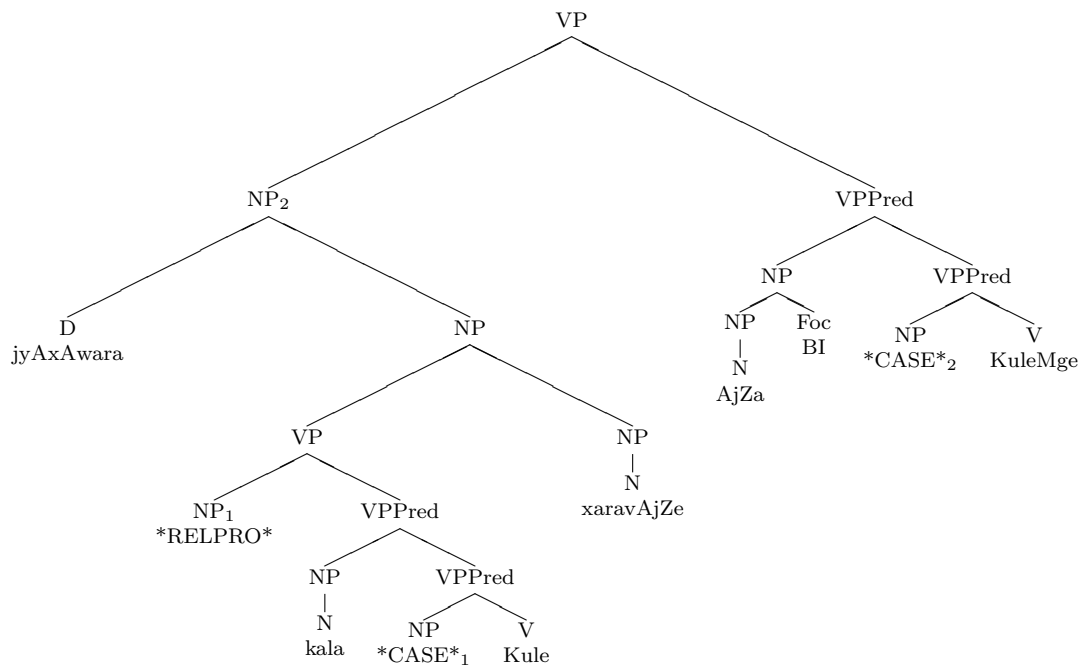
zyAxaaatar [kal khul-e] darwaaze aaaj bhii khulenge
 jyAxAwara kala Kule xaravAjZe AjZa BI KuleMge
 most yesterday open-Pfv.MPl door.MPl today also open.Fut

‘Most doors that opened yesterday will also open today.’

Tree for (87a) (Transitive (Passive) Participial Relative Clause):



Tree for (87b) (Unaccusative Participial Relative Clause):



Participial relatives based on the habitual and progressive participles are systematically on the highest argument. They do not raise any special issues; their structure would be parallel to the above trees.

19.4 *vaalaa* Modifiers of Nouns

vaalaa allows NPs and APs and infinitivals to be attached to a noun and receive a wide range of interpretations.

Note: VAALAA is tagged as a postposition in the entire treebank.

When *vaalaa* is attached to an infinitival clause, one of the arguments of the infinitival clause, typically the subject, is relativized over. This is indicated by putting a *RELPRO* in that position. A *RELPRO* is only indicated if the *vaalaa* is attached to an infinitival. When *vaalaa* attaches to an NP/AP, there is no *RELPRO*.

- (88) a. (Vaalaa-Clauses-PS-1) infinitival + *vaalaa*
- | | | | |
|-------------|------------|------------|-------------------|
| [Dilli jAne | vaalaa] | laṛkaa | naachegaa |
| xilli jAne | vAIA | ladZakA | nAcegA |
| Delhi | go-Inf.Obl | VAALAA.MSg | boy dance.Fut.MSg |
- ‘The boy who goes/is about to go to Delhi will dance.’
- b. (Vaalaa-Clauses-PS-3) XP + *vaalaa*
- | | | | | |
|---------------|--------|--------|--------|------------|
| piine | vaalaa | paanii | gandaa | thaa |
| pIne | vAIA | pAnI | ganxA | wA |
| drink.Inf.Obl | VAALAA | water | dirty | be.Pst.MSg |

‘The drinking water was dirty.’

- c. (Vaalaa-Clauses-PS-21) XP + *vaalaa*

Ram vaalii kitaab nahiin bikii
 rAma vAII kiwAba nahIM biki
 Ram VAALAA.f book.f

‘The ‘Ram’ book didn’t sell.’

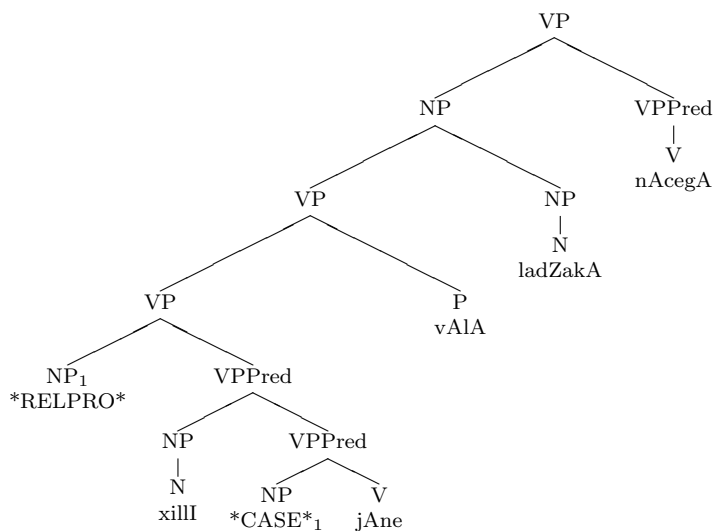
- d. (Vaalaa-Clauses-PS-22) XP + *vaalaa*

laal vaalii kitaab nahiin biki
 lAla vAII kiwAba nahIM biki
 red VAALAA.f book.f

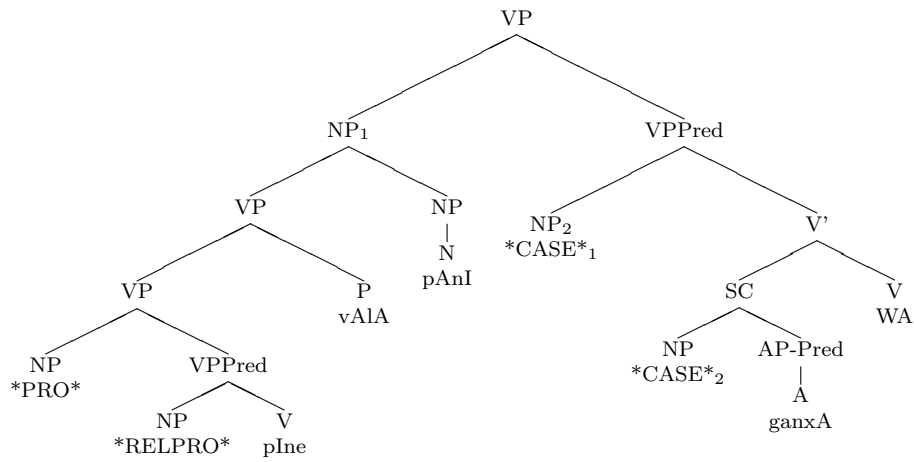
‘The red book didn’t sell’

The infinitival versions will have a null subject.

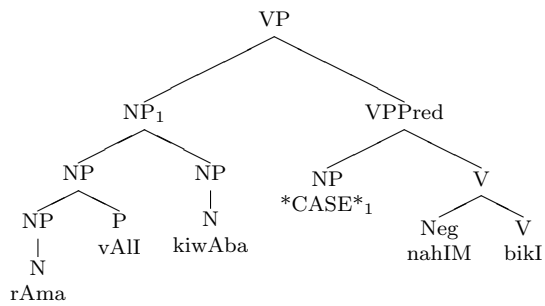
Tree for (88a) (VAALAA Infinitival Clause with subject relativization):



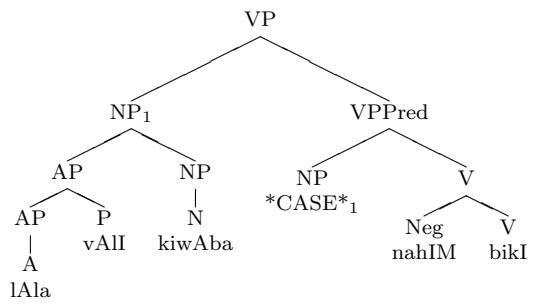
Tree for (88b) (VAALAA Infinitival Clause with non-subject relativization):



Tree for (88c) (VAALAA plus NP):



Tree for (88d) (VAALAA plus AP):



Chapter 20

Light Verb Constructions

20.1 Linguistic Description

In a Light Verb Constructions (LVCs), an event noun appears with a light verb, often *ho* ‘be’ or *kar* ‘do’. The event noun supplies the event description as well as the internal arguments if any, and the light verb supplies the clausal structure, including clausal case marking. In the case of *kar* ‘do’, the light verb also introduces the external argument. We assume that the event noun along with its internal arguments (non-agentive arguments) originate in the internal position below V' .

There are two issues that together determine the phrase structure:

- Does the light verb construction have an external argument? We have two cases.
 1. If there is no such argument realized, the support verb is usually *ho* ‘be’, and we have case-driven movement of a nominal to the first position (as in the passive and unaccusative constructions).
 2. If there is such an argument, the support verb is usually *kar* ‘do’. The first position is now filled by the external argument. Our phrase-structure does not explicitly indicate the relation of the external argument to the event nominal.
- Is the internal argument of the event noun realized with clausal or nominal syntax, i.e., with clausal case marking and/or agreement or with genitive? The facts fall neatly into two cases: ****ocr These facts should be separated out as linguistic description!****
 1. If the internal argument of the event noun is realized as a genitive, the event noun itself participates in clausal syntax: it can get clausal case marking and/or show agreement with the verb. We analyze these cases by showing the event noun moving for case to the second position. If there is no external argument, it continues on to the first position. If the internal argument is separated from the event noun (say, by an adverb), then we analyze this as genitive scrambling, which is independently attested. In all these cases, sentential negation must follow the event noun, as expected by our analysis.

2. If the internal argument does not have genitive marking, it must get case in the verbal projection. In these cases we assume it is only the internal argument of the event noun that leaves the event noun NP for a case position, namely the second position. If there is no external argument, it continues on to the first position. The remaining event noun can (but does not have to) incorporate into the verb (adjoin to V). In these cases, sentential negation may precede the event noun. If sentential negation follows the event noun in this case, then we assume the event noun stays in its original position.

20.2 A Typology of Light Verb Constructions

The above classification gives us the following four cases, which we now illustrate.

- (89) a. no external argument, internal argument of event nominal gets nominal case
 b. no external argument, internal argument of event nominal gets clausal case
 c. external argument, internal argument of event nominal gets nominal case
 d. external argument, internal argument of event nominal gets clausal case
 (nominal case = genitive)

Note that some event nouns (*udghaatan* ‘inauguration’) only allow for their internal argument to receive the nominal case marking, some only for clausal case marking (*kshamaa* ‘forgiveness’), and some allow both (*corii* ‘theft’). This is a lexical property.

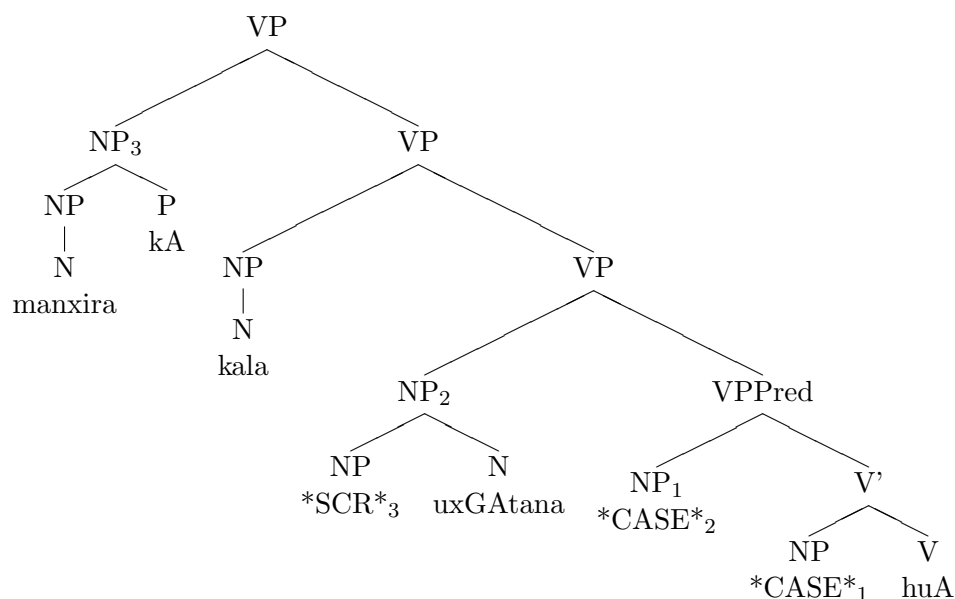
20.2.1 No External Argument, Internal Argument has Nominal Case

- (90) (N-predicates-PS-3) NP + be:

| | | | | |
|---------|-----|-----------|--------------|------------|
| mandir | kaa | kal | udghaatan | huaa |
| manxira | kA | kala | uxGAtana | huA |
| temple | Gen | yesterday | inauguration | be.Pfv.MSg |

‘The temple was inaugurated yesterday.’

Tree for (90) (N-predicates-PS-3):



The NP is an inner argument of V and the internal argument of the event nominal branches from N. That the event nominal NP is a proper argument of the verb is shown by the fact that it triggers agreement. Moreover negation follows it. This is represented by the fact that the NP headed by *udghaṭan* ‘inauguration’ ends up in [Spec,VP], the inner argument position.

- (91) (N-predicates-PS-5)
NP+*be*: genitive, no external argument:

kal [gehenō kii chorii] huii
kala gahanoM kI corI huI
yesterday jewels Gen.f theft.f be.Pfv.f

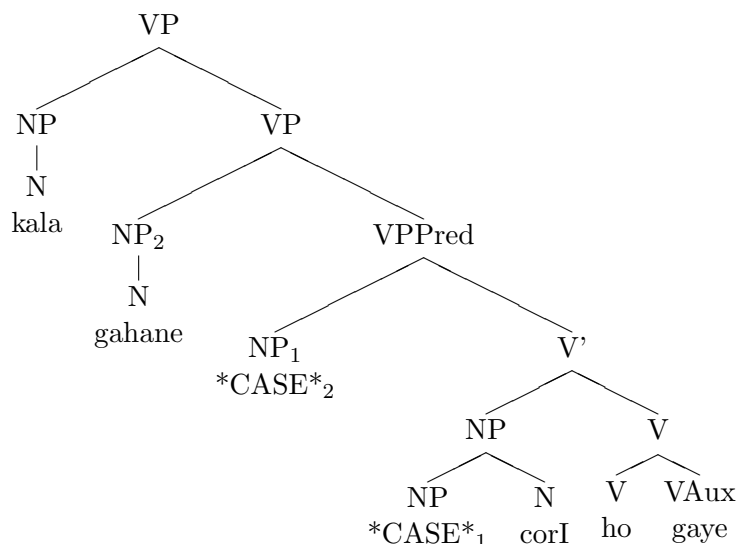
‘Yesterday the jewels got stolen.’ (Literally: Yesterday, there was a theft of jewels.)
Tree is similar to (90).

20.2.2 No External Argument, Internal Argument has Clausal Case

- (92) (N-predicates-PS-6)
NP + *be*: no genitive, no external argument:

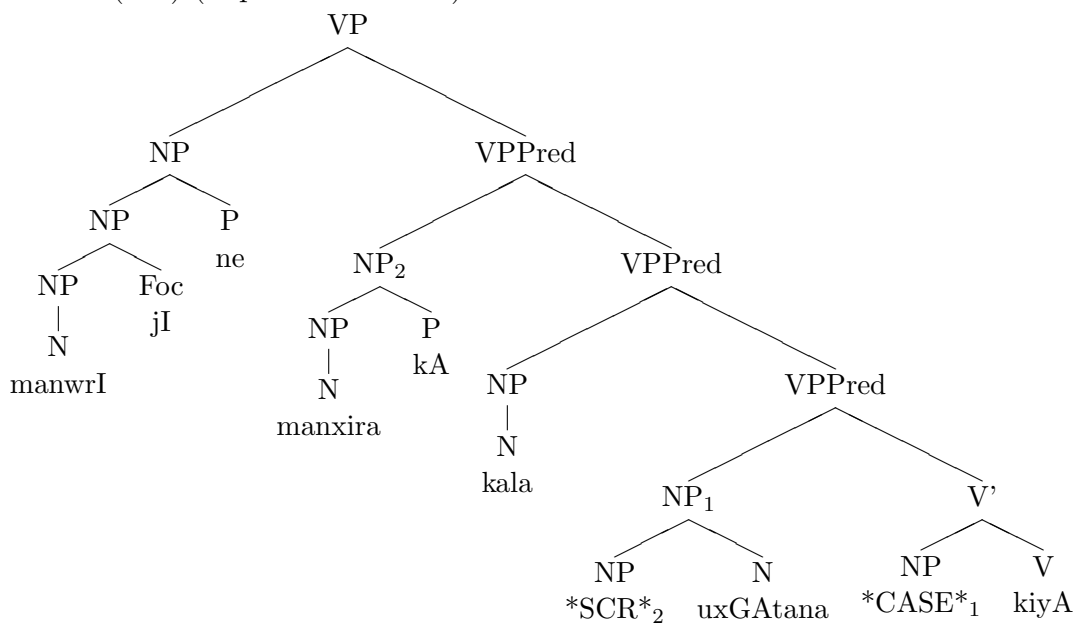
kal gehenē chorii ho gaye
kala gahane corI ho gaye
yesterday jewels-MPl theft.f be GO.Pfv.MPl

‘Yesterday the jewels got stolen.’
Tree for (92) (N-predicates-PS-6):



20.2.3 External Argument, Internal Argument has Nominal Case

- (93) a. (Label:N-predicates-PS-4) NP + do:
 mantrii jii ne mandir kaa kal udghaatan kiyaa
 manwrI jI ne manxira kA kala uxGAtana kiyA
 minister-Hon Erg temple Gen yesterday inauguration do.Pfv.MSg
 ‘The minister inaugurated the temple yesterday.’
 Tree for (93a) (N-predicates-PS-4):



- b. (N-predicates-PS-7)
 NP + *kar*: genitive, external argument:
 Atif ne kal [gehenō kii chori] kii
 AwIPZa ne kala gahanoM kI corI kI
 Atif Erg yesterday jewels Gen.f theft.f do.Pfv.f

‘Atif stole the jewels yesterday.’

Tree is similar to (93a).

20.2.4 External Argument, Internal Argument has Clausal Case

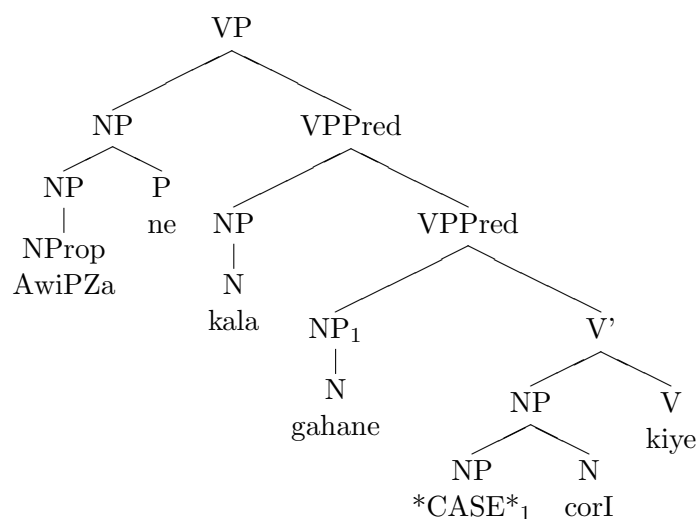
(94) (N-predicates-PS-8)

NP + *kar*: no genitive, external argument:

Atif ne kal gehenē chorii kiye
 AwiPZa ne kala gahane corI kiye
 Atif Erg yesterday jewels.f theft.f do.Pfv.f

‘Atif stole the jewels yesterday.’

Tree for (94) (N-predicates-PS-8):



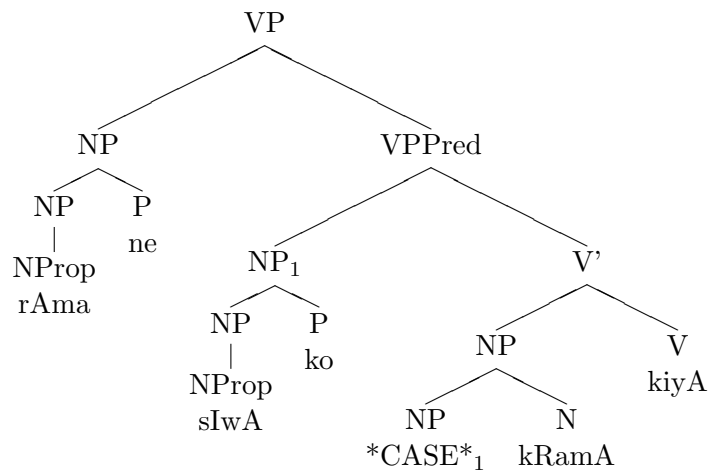
(95) (Label:N-predicates-PS-15)

NP + *kar*: no genitive, external argument:

Ram ne Sita ko shamaa kiyaa
 rAma ne sIwA ko kRamA kiyA
 Ram Erg Sita Acc forgiveness do.Pfv

‘Ram forgave Sita.’

Tree for (95) (N-predicates-PS-15):



this tree is similar to 94.

In these cases, the internal argument of the event nominal obligatorily moves out of the NP headed by the event nominal. Negation can precede the event nominal indicating that the event nominal can incorporate into the verb.

20.3 Dative Arguments associated with Event Nominals

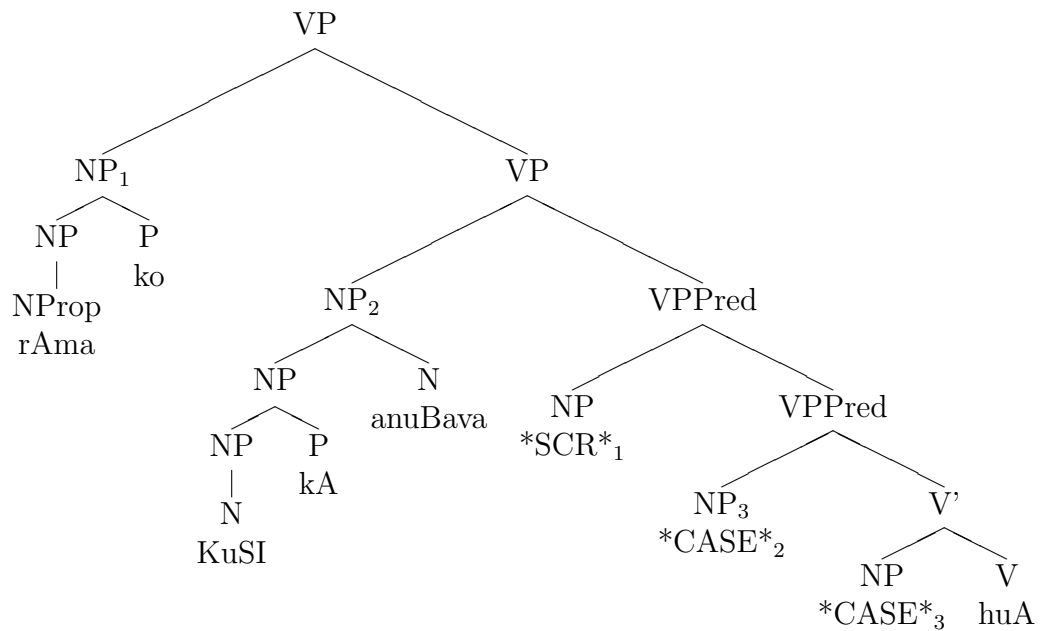
Some event nouns also allow for a dative experiencer argument.

- (96) a. (N-predicates-PS-9) dative argument and an internal argument with nominal case:

Ram ko khushii kaa anubhav huaa
 rAma ko KuSI kA anuBava huA
 Ram Dat happiness.f Gen experience be.Pfv.MSg

‘Ram experienced happiness.’

Tree for (96a) (N-predicates-PS-9):

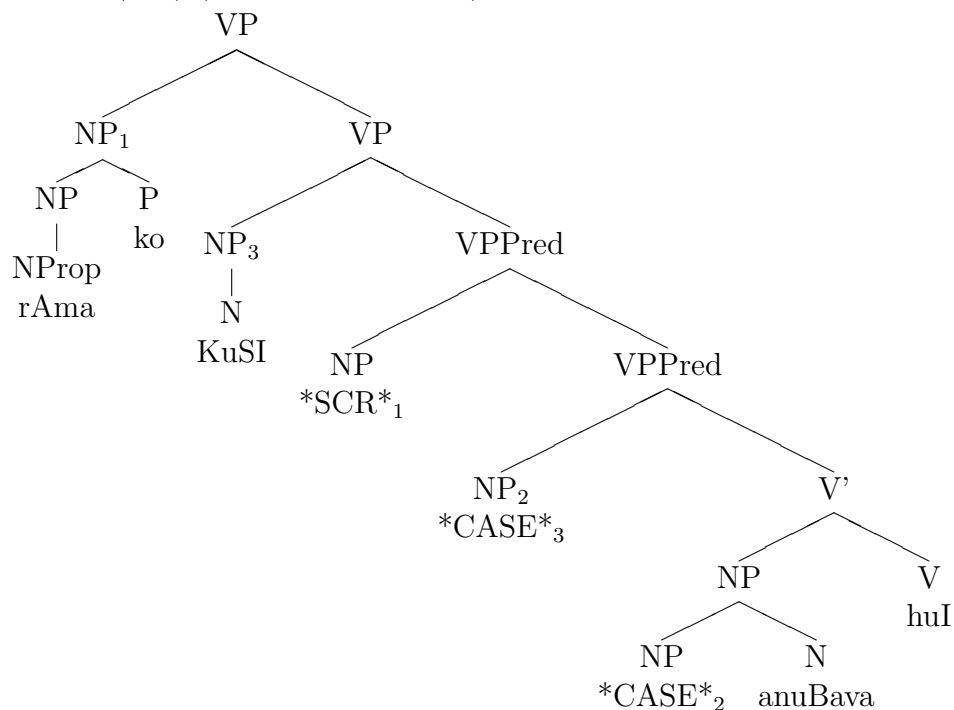


- b. (N-predicates-PS-10) dative argument and an internal argument with clausal case:

Ram ko khushii anubhav hui
 rAma ko KuSI anuBava huI
 Ram Dat happiness.f experience be.Pfv.f

‘Ram experienced happiness.’

Tree for (96b) (N-predicates-PS-10):

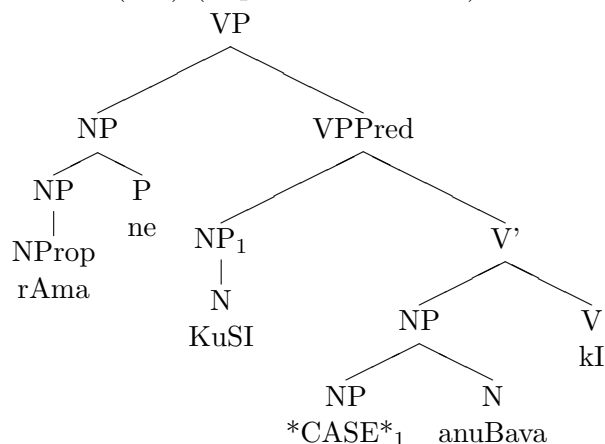


- c. (Label:N-predicates-PS-11) internal argument has clausal case, instead of the dative there is an ergative external argument:

Ram ne khushii anubhav kii
 rAma ne KuSI anuBava kI
 Ram Erg happiness.f experience do.Pfv.f

‘Ram experienced happiness.’

Tree for (96c) (N-predicates-PS-11):



This tree is similar to (94). Negation can precede *anuBava* ‘experience’.

- (97) a. (Label:N-predicates-PS-12) dative argument and internal argument with nominal case:

Ram ko Atif kii **yaad** **aa-ii**
 rAma ko AtiPZa kI yAxa AI
 Ram Dat Atif Gen.f memory.f come-Pfv.f

‘Ram thought of Atif.’

Tree is similar to (96a).

- b. (Label:N-predicates-PS-13) dative argument and internal argument with clausal case:

Ram ko **Atif** yaad **aa-ya**
 rAma ko AwiPZa yAxa AyA
 Ram Dat Atif.M memory.f come-Pfv.MSG

‘Ram thought of Atif.’

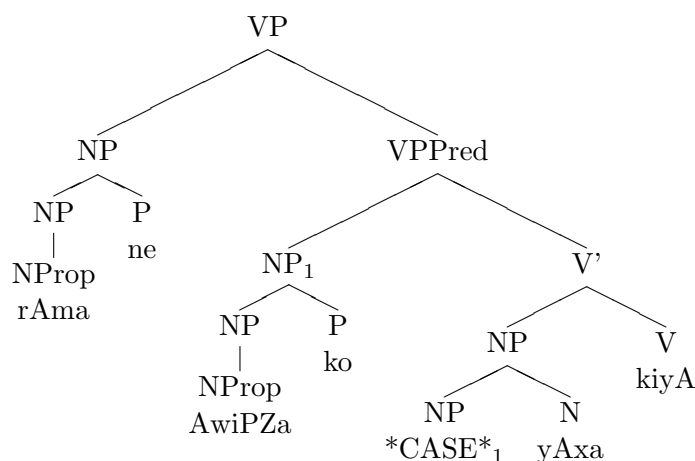
Tree is similar to (96b).

- c. (Label:N-predicates-PS-14) internal argument has clausal case, instead of the dative there is an ergative external argument

Ram ne Atif ko yaad kiyaa
 rAma ne AwiPZa ko yAxa kiyA
 Ram Erg Atif Acc memory.f do.Pfv

‘Ram thought of Atif.’

Tree for (97c) (N-predicates-PS-14):



This tree is similar to 94. Negation can precede *yAxa* ‘memory’.

1

20.4 Light Verb Constructions with Foreign Eventive Expressions

ocr Need more text here!

- (98) unhone sab ko shaadii kaa kard email kiyaa
 (99) sab ko shaadii kaa kard email kiyaa jaa chukaa hai
 (100) Ram khabar kal announce karegaa
 (101) Game finish/khatam ho chukaa hai
 (102) Game start/shuruu ho gayaa hai
 (103) paanii freeze ho gayaa

These are treated in essentially the same way as the cases above.

20.5 Placement of Negation

In general, sentential negation immediately precedes a verbal head. The only exceptions to this generalization are found with a subclass of NP+V predicates. With most NP+V predicates, we find the following pattern, where the negation cannot precede the NP.

¹The issue of where the dative is being licensed is being set aside here. It is clear that the NP by itself cannot license a dative in Hindi. But the N can license a secondary genitive which seems to play the same role as that played by the dative i.e. experiencer. So it is possible that the possibility of the dative is licensed within the NP but the actual dative is only licensed by verbal structure. Additional complications arise when there is a *kar* ‘do’ and an ergative subject. In such cases, a dative is not possible - the ergative argument expresses essentially the meaning that the dative would have. For now, we plan to not represent the second argument as having an NP-internal source.

(104) (*Negation) NP (Negation) V

But with a subclass of NP+V predicates, negation can optionally precede the NP, yielding the following pattern.

(105) (Negation) NP (Negation) V

This subclass of NP+V predicates happens to be exactly the subclass of NP+V predicates whose NPs have the option of not licensing case on their arguments. Some such nouns can license genitive but have the option not too - these are the cases in §??. These nouns allow a negation to precede exactly when they do not license a genitive. Other nouns lack the ability to license genitive on their arguments altogether - these are the ones in §??.

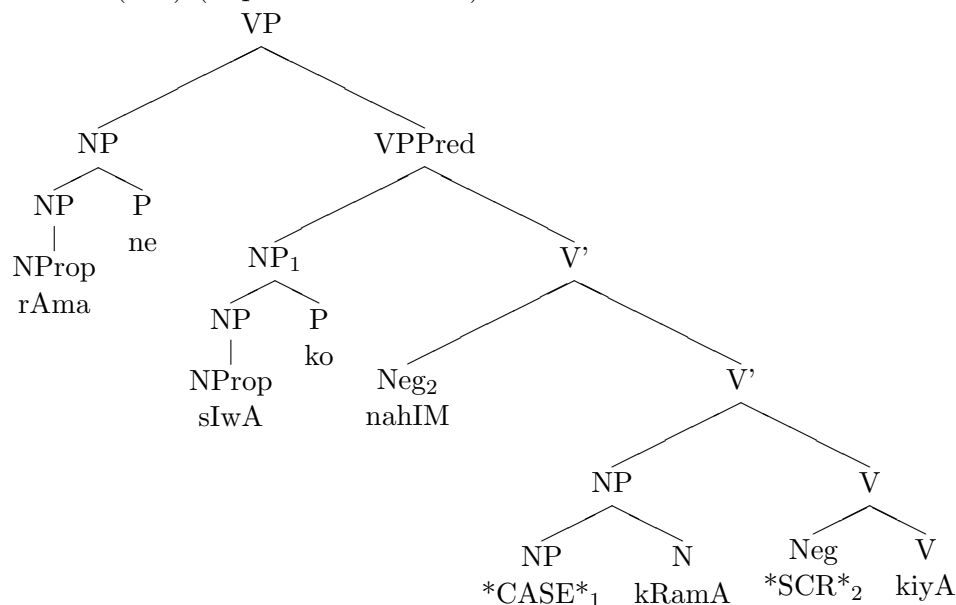
(106) (Label:N-predicates-PS-16)

NP + *kar*: no genitive, external argument:

Ram ne Sita ko nahī: shamaa kiyaa
 rAma ne sIwA ko nahIM kRamA kiyA
 Ram Erg Sita Acc Neg forgiveness do.Pfv

‘Ram did not forgive Sita.’

Tree for (106) (N-predicates-PS-16):



We could analyze the cases of negation preceding the N in an NP+V predicate as involving incorporation of the N into the V. Incorporation is only an option when an NP that does not license case on its internal argument is the sister to a verb.

It is quite possible that incorporation also takes place in the absence of negation. But it is only in the presence of a negation preceding the N that an incorporation analysis is forced on us.

However, for simplicity we treat these case as involving a scrambled negation. In general though, scrambling of negation seems to be restricted in the language to cases like these.

20.6 Pronominal Predicates with Extraposed Clause

There is nothing here yet.

20.7 Non-Finite Complement Clauses

****ocr All embedded infinitivals project to NP.****

20.8 Small Clauses

****ocr Do we still need this?****

20.9 Finite Complement Clauses

Nothing here.

20.10 Result Clauses

Nothing here.

20.11 Conditionals

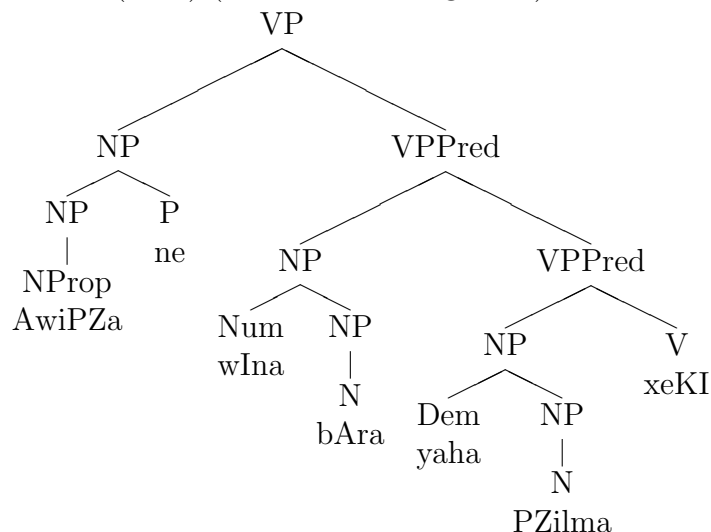
Chapter 21

Local Scrambling

As indicated earlier, adjuncts will appear adjoined to V' or to VP. If an adjunct appears between the lower argument and the verb, it indicates that the lower argument has scrambled over the adjunct. This will be indicated by the trace *SCR*.

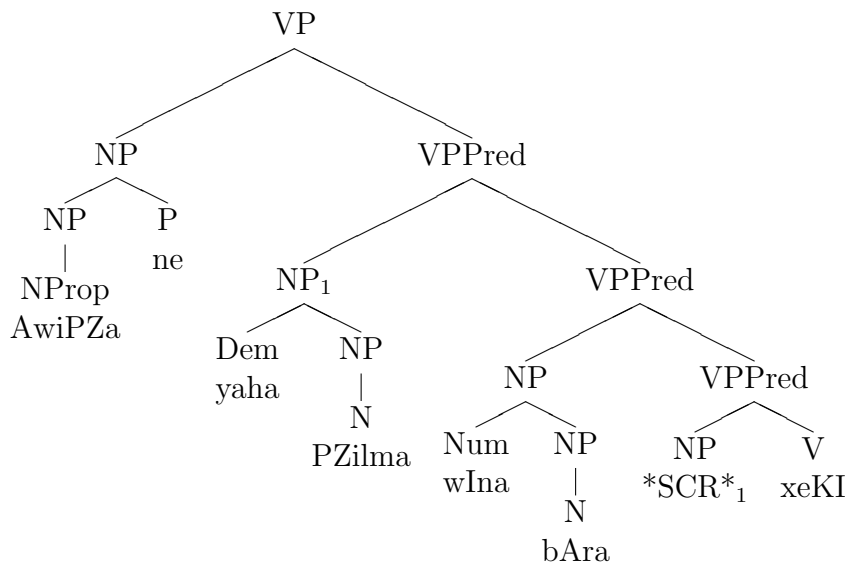
- (107) a. (Local-Scrambling-PS-1)
 Atif ne tiin baar yeh film dekh-ii
 AwiPZa ne wIna bAra yaha PZilma xeKI
 Atif-Erg three times this film.f see-Pfv.f
 'Atif saw this film three times.'

Tree for (107a) (Local-Scrambling-PS-1):



- b. (Local-Scrambling-PS-2)
 Atif ne yeh film tiin baar dekh-ii
 AwiPZa ne yaha PZilma wIna bAra xeKI
 Atif-Erg this film.f three times see-Pfv.f
 'Atif saw this film three times.'

Tree for (107b) (Local-Scrambling-PS-2):



Any deviation of elements from canonical order is indicated via traces. This includes the first and the second argument, dative arguments, as well as NP possessors. The general rule of thumb is that if an element can be interpreted in a position compatible with its surface order, then a trace is not necessary. This is the case with most adverbs where it does not matter for our purposes whether they are adjoined to VP, VPPred or V'. First and second arguments, dative arguments, possessors, degree modifiers of AP, however, can only be introduced in designated structural locations and this is indicated via traces.

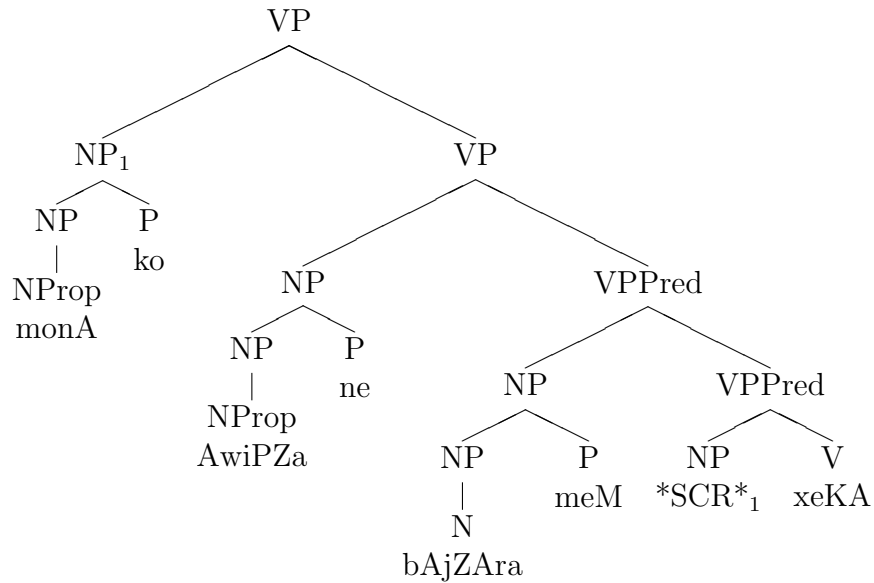
****ocr Why do we need the trace if adjuncts can adjoin to V'??****

(108) a. (Local-Scrambling-PS-3): non canonical order of second argument: trace

Mona ko Atif ne bazaar me dekh-aa
 monA ko AwiPZa ne bAjZARA meM xeKA
 Mona Acc Atif Erg market in see-Pfv

‘Atif saw Mona in the market.’

Tree for (108a) (Local-Scrambling-PS-3):

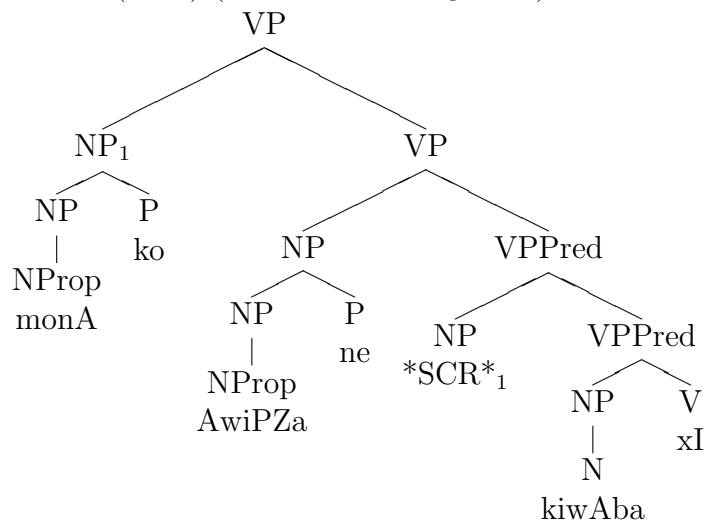


- b. (Local-Scrambling-PS-4): non canonical order of dative argument: trace

Mona ko Atif ne kitaab di-i
 monA ko AwiPZA ne kiwAba xI
 Mona Dat Atif Erg book.f give-Pfv.f

‘Atif gave a book to Mona.’

Tree for (108b) (Local-Scrambling-PS-4):

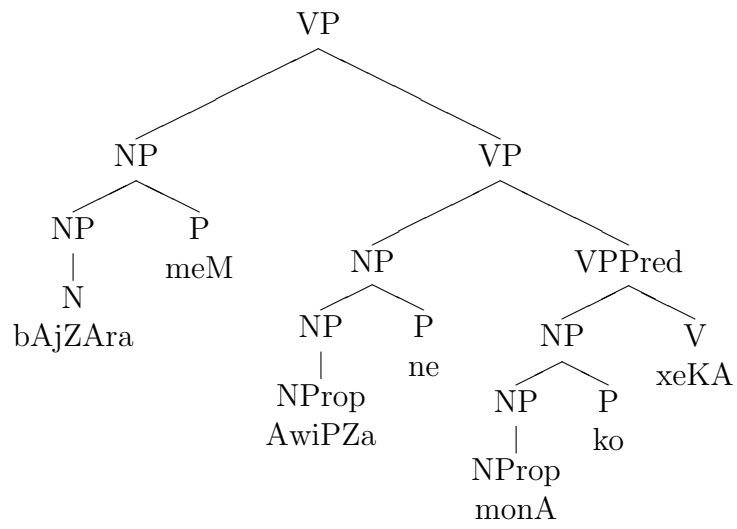


- c. (Local-Scrambling-PS-5): variant placement of adverbs: no traces

bazaar me Atif ne Mona ko dekh-aa
 bAjZAra meM AwiPZA ne monA ko xeKA
 market in Atif Erg Mona Acc see-Pfv

‘Atif saw Mona in the market.’

Tree for (108c) (Local-Scrambling-PS-5):



Chapter 22

Cross-clausal Scrambling

Reorderings that cross clausal boundaries always involve *SCR* traces. The traces are left as follows: traces of the embedded first argument is left in its structural position. Traces of non-structurally represented arguments/adjuncts are left in the adjoined to VP position. Traces of the second argument are left in two positions: the low position and the adjoined to VP position: this is because the evidence from reconstruction suggests that while long scrambled NPs reconstruct, they do not need to do so to their base position.

****ocr The above discussion comes a little bit out of left field, there has been no discussion of the properties of scrambling. What reconstruction effects (binding?)? Also, what about the dative argument – presumably it behaves like the second argument?****

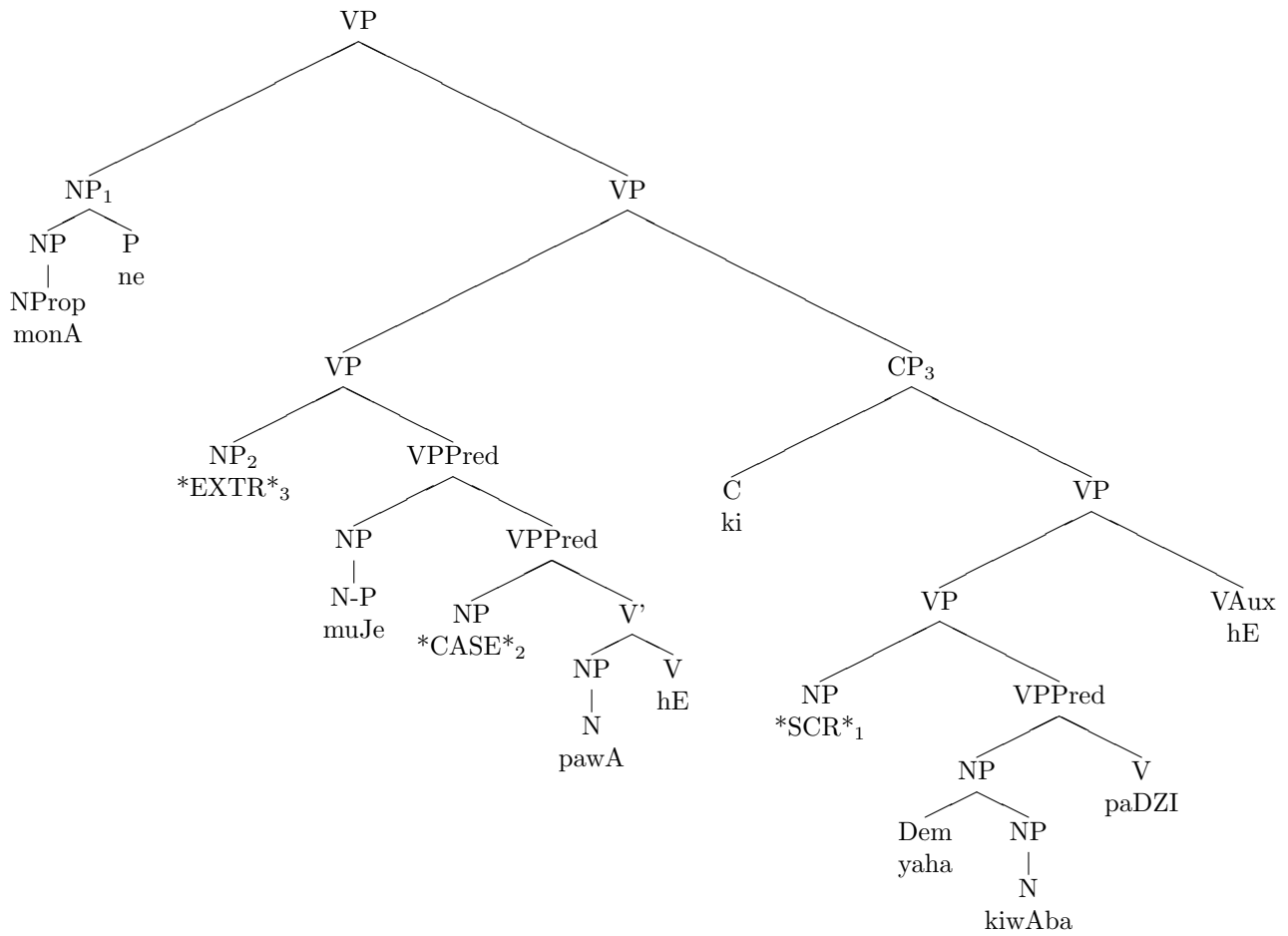
****ocr Can we also have a simple example with embedded infinitival complement (as opposed to extraposed finite complement)?****

(109) a. (Long-Scrambling-PS-1) embedded subject:

Mona ne, mujhe pataa hai [ki yeh kitaab paṛh-ii hai]
monA ne muJe pawA hE ki yaha kiwAba paDZI hE
Mona Erg me.Dat known is that this book.f read-Pfv.f be.Prs.Sg

‘Mona, I know has read this book.’

Tree for (109a) (Long-Scrambling-PS-1):

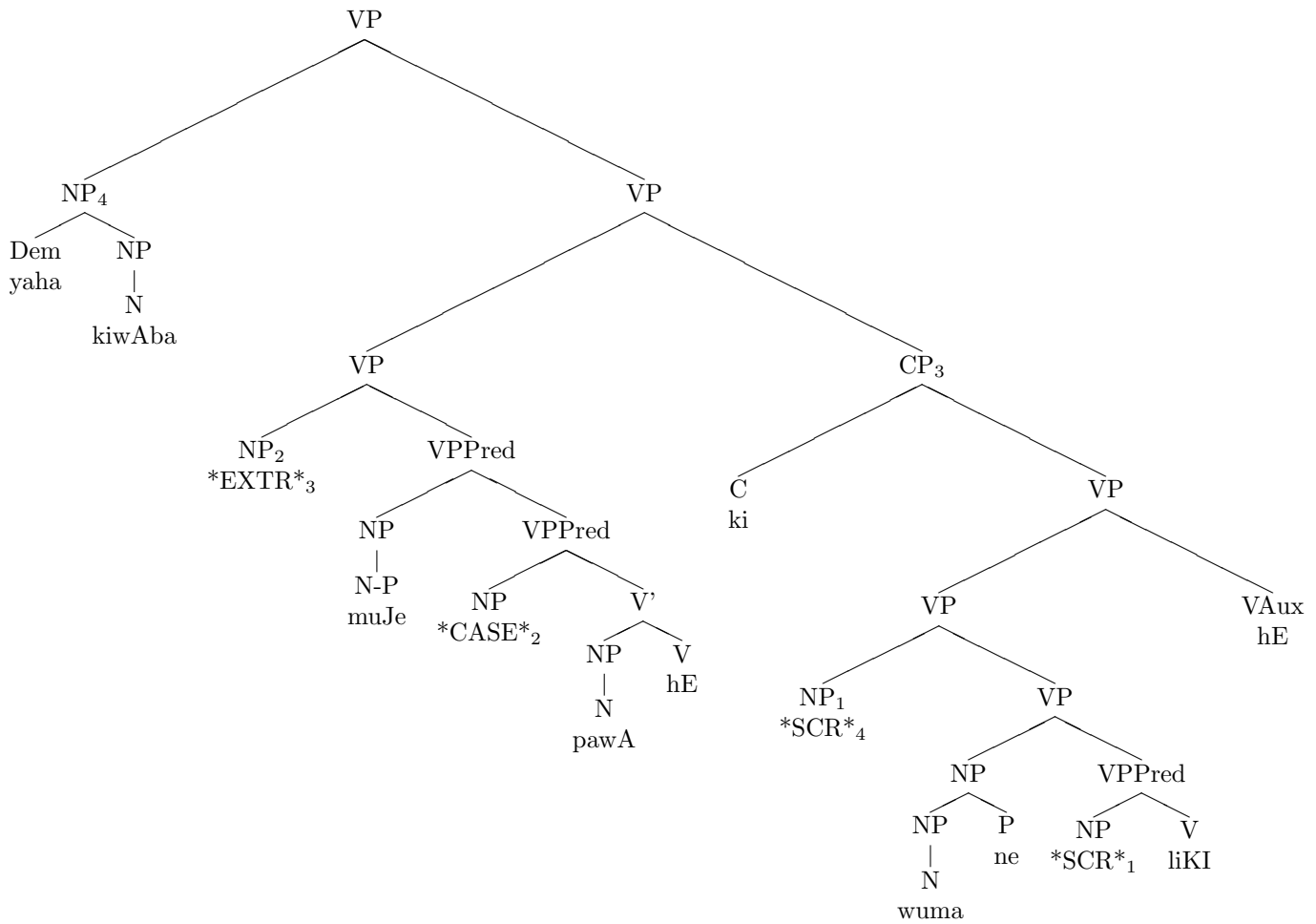


b. (Long-Scrambling-PS-2) embedded object:

yeh kitaab mujhe pataa hai [ki wuma ne likh-ii hai]
 yaha kiwAba muJe pawA hE ki wuma ne liKI hE
 this book.f me.Dat known is that you Erg write-Pfv.f be.Prs.Sg

‘This book, I know that you have written.’

Tree for (109b) (Long-Scrambling-PS-2):

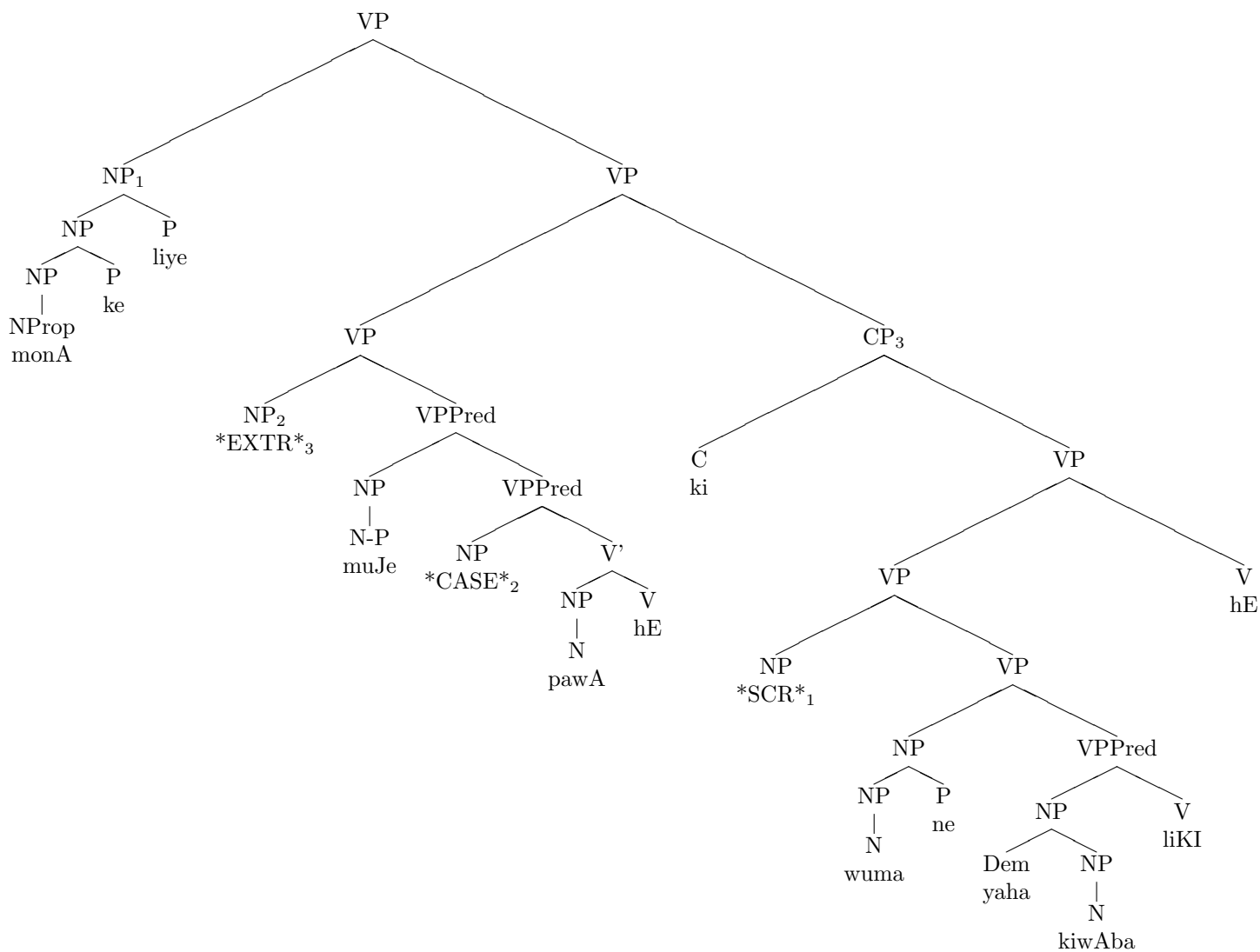


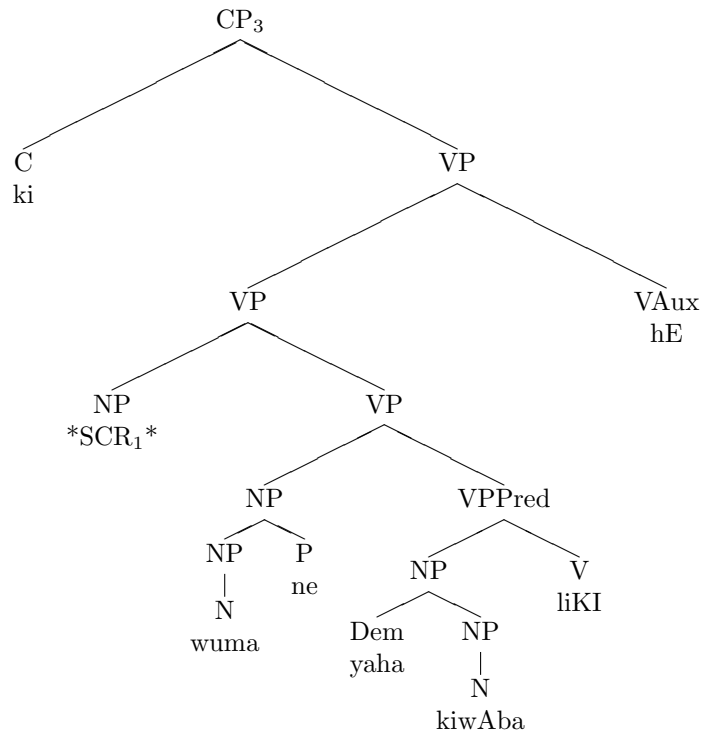
c. (Long-Scrambling-PS-3) embedded adjunct:

Mona ke liye mujhe pataa hai [ki tum ne yeh kitaab likh-ii
 monA ke liye muJe pawA hE ki wuma ne yaha kiwAba liKI
 Mona Gen for me.Dat known is that you Erg this book.f wrote-Pfv.f
 hai]
 hE
 be.Prs.Sg

‘For Mona, I know that you wrote this book.’

Tree for (109c) (Long-Scrambling-PS-3):





Chapter 23

‘Rightward’ Scrambling

is tough.

Part VI

Noun Phrases and Other Phrases

Chapter 24

Postpositional Phrases

24.1 Linguistic Description

24.1.1 True Postpositions (P)

Postpositions follow a noun phrase or a clause and relate it to its governor. A postposition can play the role of a case marker and/or a preposition in other languages.

****ocr Add NP and VP examples****

Focus particles can separate an NP from its postposition.

(110) (Label:Structure-of-PP-PS-1)

Ram tak ne/Ram hii ne
rAma waka ne/rAma hI ne
Ram even Erg/Ram Foc Erg

‘Even/Only Ram...’

Moreover, Ps can be shared across a coordination.

(111) (Label:Structure-of-PP-PS-2)

Ram aur Ramesh ne
rAma Ora rameSa ne
Ram and Ramesh Erg

‘Ram and Ramesh’

The placement of focus particles with respect to case markers is semantically significant in the case of coordination.

(112) a. (Label:Structure-of-PP-PS-3)

laṛkiyō ne aur laṛkō ne bhii yeh kah-aa
 ladZakiyoM ne Ora ladZakoM ne BI yaha kahA
 girls Erg and boys Erg also this say-Pfv

Reading 1: Girls said this. Boys also said this. ([NP-K] and [NP-K-Foc])

Reading 2: Someone else said this. Girls and boys said this also. ([NP-K and NP-K]-Foc)

(Reading 2 is less prominent than Reading 1)

However, if the two conjuncts share a postposition, then the scope of the focus marker extends over the whole conjunction.

(113) a. (Label:Structure-of-PP-PS-4)

laṛkiyō aur laṛkō ne bhii yeh kah-aa
 ladZakiyoM Ora ladako ne BI yaha kahA
 girls and boys-Erg-also this say-Pfv

* Unavailable Reading 1: Girls said this. Boys also said this. ([NP] and [NP-K-Foc])

Reading 2: Someone else said this. Girls and boys said this also. ([NP and NP]-K-Foc)

Finally there is also the issue of obliqueness that plays a role in creating certain surface ambiguities. Obliqueness refers to the fact that an NP that combines with a postposition appears in special oblique form that is distinct from the form that it would have without the postposition, which is sometimes called the direct form. Obliqueness is marked on the head as well as on any modifiers.

(114) a. (Label:Obliqueness-PS-1) direct form:

Ram ke lambe beṭe
 rAma ke lambe beṬe
 Ram Gen.MPl tall.MPl son.MPl
 ‘Ram’s sons’

b. (Label:Obliqueness-PS-2) oblique form:

Ram ke lambe beṭe ko
 rAma ke lambe beṬe ko
 Ram Gen.Obl tall.Obl son.Sg.Obl Dat
 ‘Ram’s tall son’

One important issue here is that Hindi orthography does not always represent the NP and the postposition as two separate words. In general, proper names and full NPs are separated from the postposition. The situation with pronouns tends towards writing them as one word with the postposition. But with the exception of a few forms like *unhone* ‘they.Erg’ and *mujhe* ‘me.Dat’, there are no good linguistic grounds for analyzing pronoun-postposition

sequences as lacking syntactic structure. Indeed Urdu orthography does not treat pronoun-postposition sequences as forming a single word.¹

(115) Pronouns + Postpositions:

- a. morphologically complex: datives: *mujhe* ‘me.Dat’, *hamē* ‘we.Dat’, *tujhe* ‘you.Dat’, *tumhē* ‘you.Pl/Hon.Dat’, *use* ‘s/he.Dat’, *unhē* ‘they.Dat’,
genitives: *meraa* ‘my’, *hamaaraa* ‘ours’, *teraa* ‘yours’, *tumhaaraa* ‘yours.Pl/Hon’
ergatives: *unhone* ‘they.Erg’
- b. should be separated into two words:
us ne ‘s/he Erg’, *sab ko* ‘all Dat’,
(and all others)

24.1.2 Spatio-Temporal Nouns (NST)

There are several complex postpositions that consist of the genitive followed by a semantically contentful head. There are two classes.

- The heads that contribute relational meanings cannot be modified. Moreover they cannot stand by themselves (i.e. without the NP-Genitive).

ocr Need full examples

- (116) a. partitive: *mē se* ‘in from’
b. the oblique genitive + X class with relational meaning:
ke dwaaraa ‘by’
ke liye ‘for’
ke baare me ‘about’

- The locative/temporal heads can be modified by elements like *thiik* ‘exactly’ and measures of time/distance. They can also stand on their own without a genitive complement.

ocr Need to check if these examples are treated as NST in DS!

ocr Need full examples

- (117) a. temporal/locative:
ke paas: near
ke u:par/ni:ce/saamne/aage/piice:
above/below/ in front of/ ahead/behind
ke baad: after
ke andar/baahar ‘inside/outside’
ke bagal mē ‘next to’
..... (many others)
- b. the instrumental + X class:
se pehle ‘before’

¹What is the status/category of *saa*?

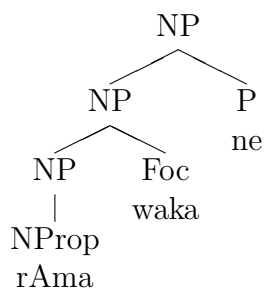
24.2 PS Analysis

24.2.1 True Postpositions (P)

In PS, postpositions are represented syntactically but they do not head their own projection. Instead, they adjoin to an NP or, sometimes, to a VP. The reasons for representing postpositions syntactically rather than only morphologically include the facts about focus particles and about coordination.

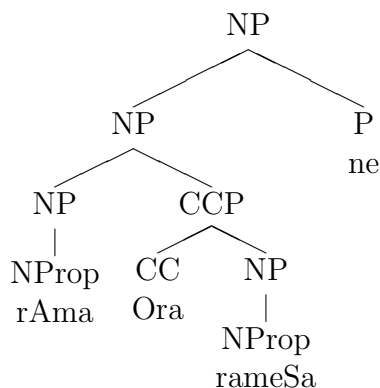
As seen in (110), focus particles can separate an NP from its postposition.

Tree for (110) (Structure-of-PP-PS-1):



As seen in (111), Ps can be shared across a coordination. This can be straightforwardly represented by using our standard analysis of coordination (Chapter 27) on the NPs, and the P adjoining to the resulting maximal NP node.

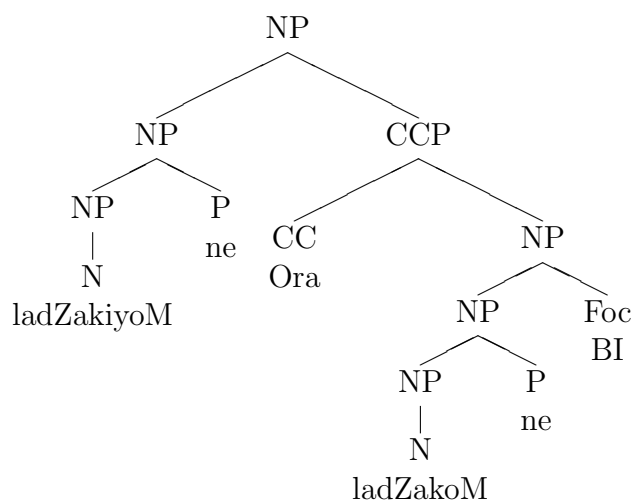
Tree for (111) (Structure-of-PP-PS-2):



Finally, we saw that the placement of focus particles with respect to case markers is semantically significant. This is straightforwardly represented in our analysis.

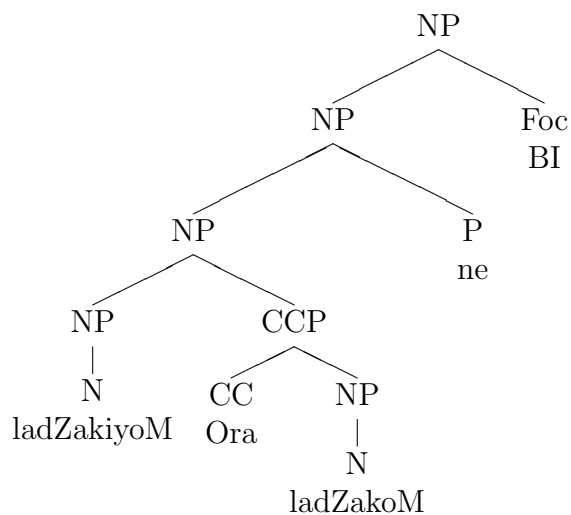
Structure for Reading 1 of (112a):

Tree for (112a) (Structure-of-PP-PS-3):



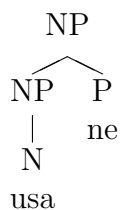
Structure for Reading 2 of (113a):

Tree for (113a) (Structure-of-PP-PS-4):



Concerning the issue of orthography, at the phrase level we have a uniform treatment of these items as NPs since the P (of spelled separately) is adjoined.

(118) a. usa ne (s/he Erg)



b. usane (s/he.Erg)

NP
|
N+P
usane

24.2.2 Spatio-Temporal Nouns (NST)

We now turn to the cases of complex postpositions. One question here is whether one should go for a $[[NP P_1] X]$ or a $[NP [P_1 X]]$ structure. For the first type, we ...

****ocr What do we do? We do P recursion****

For the NST cases, we take the fact that the NP and the genitive marker can be omitted as indication that the NST has its own projection to NSTP, and the genitive NP is an (optional) complement of that NST.

****ocr Need trees for the examples****

NSTP
/ \
NP NST
/ \
NP P

This structure makes it easy to analyze cases like the following:

(119) *mere paas* 'my near'

NSTP
/ \
NP NST
| pA_sa
N
mere

These are cases of pronouns that have synthetic single-word genitives. If the structure is $[[Pron Gen] X]$, the fact that $[Pron Gen]$ surfaces as the same single word as it does in ordinary genitive environments makes sense. It also allows for the cases where the X element can be modified and is able to stand on its own.

Chapter 25

Nominal Phrases

25.1 Structure of Noun Phrases

****ocr This needs work.****

How much structure do we want within the NP? There is a fairly strict word order with respect to modifiers - how do we want to represent this?

(120) Possessors Demonstratives Numerals Adjective Noun

Since we are tree-banking, we don't have to build in the strict ordering - representing what we find will be enough.

While all of the above modifiers can appear on a given noun, none of them have to. Bare nouns can also appear in argument positions. So we will need to allow for non-branching projections like $[_{NP}[_{N}\dots]]$.

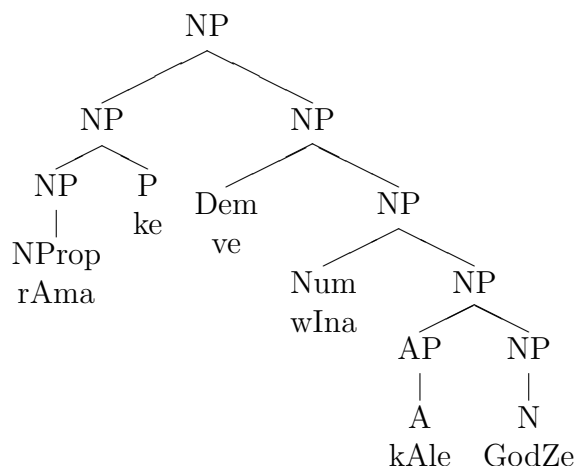
A straightforward option that we adopt is to attach all of the NP modifiers as adjuncts to the NP.

(121) (Label:Structure-of-NPS-1)

| | | | | | |
|------|---------|--------|-------|-----------|----------|
| Ram | ke | ve | tiin | kaale | ghore |
| rAma | ke | ve | wIna | kAle | GodZe |
| Ram | Gen.MPl | Dem.Pl | three | black.MPl | horses.M |

'Those three black horses of Ram's'

Tree for (121) (Structure-of-NPS-1):



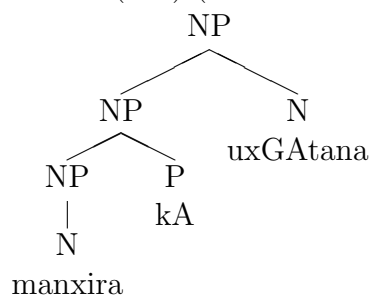
Only internal arguments of nouns combine with N directly i.e. not as adjuncts to NP. At this point, we have only identified the arguments of nouns that participate in NP+V predicates as being internal arguments of nouns. This includes cases like *mandir kaa udghaṭan* ‘temple Gen inauguration’, *Ram kii yAxa* ‘Ram Gen.f memory’, as well as cases of internal arguments that move out of the NP for case reasons such as *saikil corii* ‘cycle theft’ and *Sita ko kshamaa* ‘Sita Acc forgiveness’.

(122) (Label:Structure-of-NP-Comp-PS-1)

mandir kaa udghaṭan
 manxira kA uxGAtana
 temple Gen inauguration

‘The inauguration of the temple’

Tree for (122) (Structure-of-NP-Comp-PS-1):



(90), (97a), and (106) are illustrative examples. For further details, see §20, the NP+V predicates section.

Internal arguments of N aside, other genitive marked NPs such as possessors are treated as adjuncts on NP. Since in general the argument/adjunct distinction is difficult to make with nouns, we will simplify and treat all genitive NPs as adjuncts. The only exception will be in light verb constructions where the larger context supports treating the NP as an internal argument.

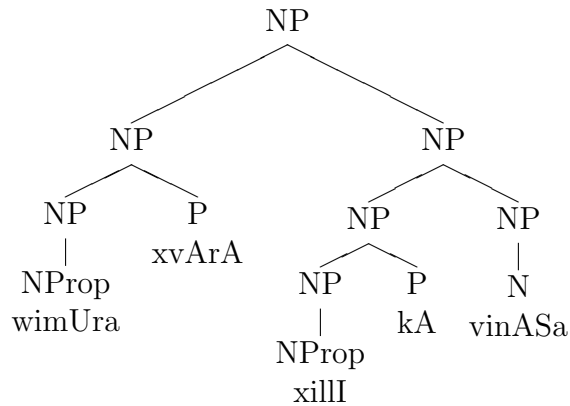
Commonly we find only one genitive marked NP licensed by a noun and this genitive PP is adjoined to the NP.

(123) (Label:Structure-of-NPS-2)

Timur dwaaraa Dilli kaa vinaash
 wimUra xvArA xilli kA vinASa
 Timur by Delhi Gen destruction

‘The destruction of Delhi by Timur’

Tree for (123) (Structure-of-NPS-2):



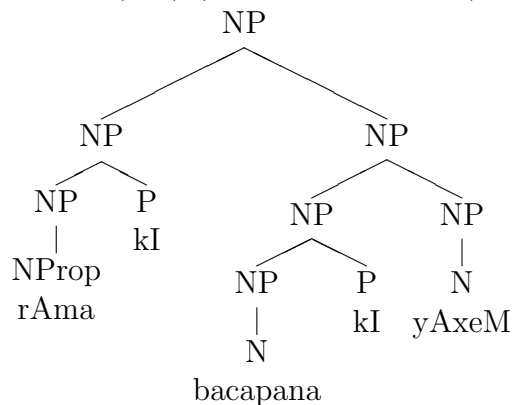
A single NP can also support more than one genitive marked NP modifier.

(124) (Label:Structure-of-NPS-3)

Ram kii bachpan kii yaadē
 rAma kI bacapana kI yAxeM
 Ram Gen.f childhood.M Gen.f memories.f

‘Ram’s memories of childhood’

Tree for (124) (Structure-of-NPS-3):



If we treat a particular NP as an internal argument of N, its appearance in a non-canonical position will be marked by traces. But if the NP in question is treated as a possessor, then

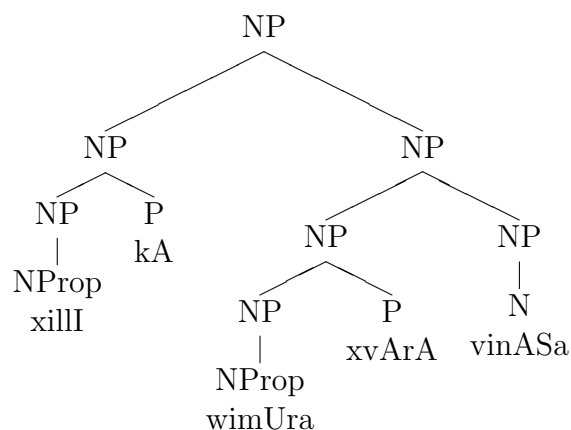
its appearance in non-canonical position does not require traces. So if we were to treat *Dilli kA* ‘Delhi Gen’ as an internal argument of *vinaash* ‘destruction’, then we would need to indicate its displacement via traces. But all else being the same, we treat genitive modifiers as adjuncts and therefore the following representation does not involve traces.

(125) (Label:Structure-of-NPS-4)

Dilli kA Timur dwaaraa vinaash
 xillI kA wimUra xvArA vinASa
 Delhi Gen Timur by destruction

‘The destruction of Delhi by Timur’

Tree for (125) (Structure-of-NPS-4):



Possessor NPs can be extracted quite freely from their NP. This movement leaves a trace. The location of the trace depends upon whether the extracted NP is an internal argument of N or an adjunct on NP.¹

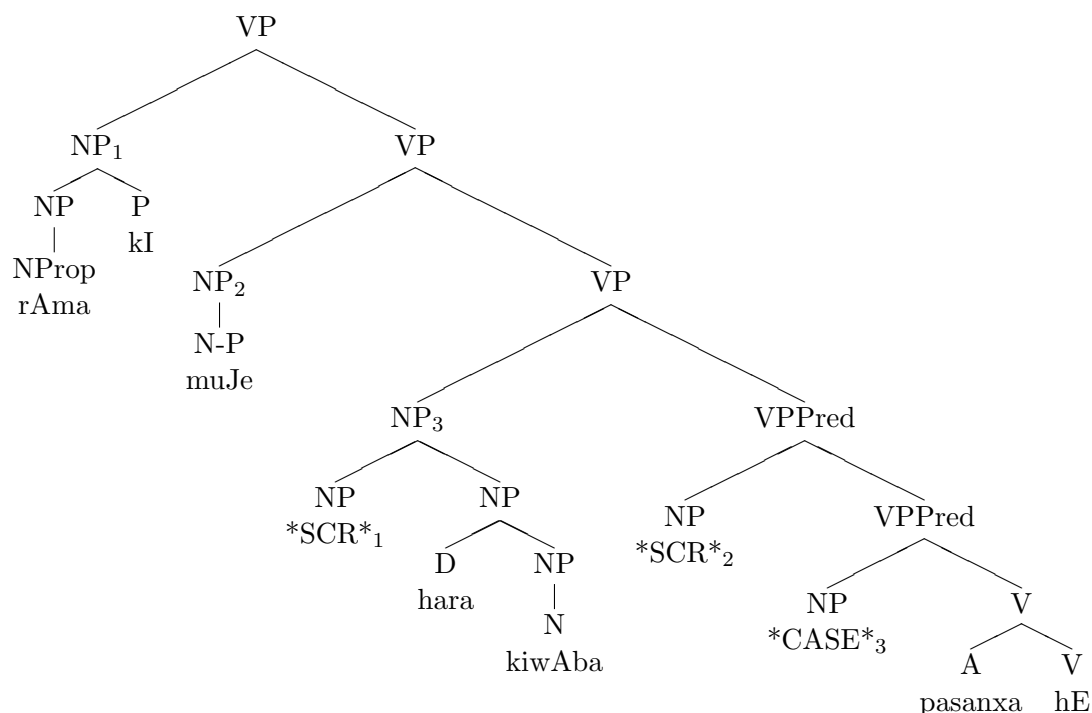
(126) (Label:Poss-Extraction-PS-1)

Ram kii mujhe har kitaab pasand hai
 rAma kI muJe hara kiwAba pasanxa hE
 Ram Gen.f me.Dat every book.f like is

‘I like every book by Ram.’

Tree for (126) (Poss-Extraction-PS-1):

¹This is one of the sources of non-projectivity within PDG.



2

25.2 Empty NPs

Any noun phrase can be replaced by a null anaphor. Furthermore, an NP with a determiner or genitive can also be a pro.

We represent this as *pro* for the whole thing, even if a genitive is present.

25.3 Noun Compounds

Various types of compound nouns also occur in Hindi-Urdu.³

²Treatment of dative subjects of the form: NP-Dat NP pasand hai is not entirely clear. I will insert *pasand* ‘like’ in the sister-to-V position.

³It should be noted that we use of the term ‘compound’ here in a largely descriptive fashion, and *not* in the context of any well-developed theory of compounds; we do not wish to espouse any particular views about morphology. This is somewhat problematic, since, while in some cases—as in the case of compound names—the choice of compound treatment is intuitively clear, other cases which the DS treats as compounds are not, from the PS’ perspective, as obviously subject to being analyzed as compounds, as opposed to nominal appositives (see next section), or adjectivally modified nouns, as in the case of *xAna rAsi* ‘donation amount’ in (1).

- (1) (Sentence-type-DS-1)
 Apa **xAna rAsi** para Cuta kA xAvA kara leM
 you **donation amount** on exception of claim do imp
 ‘You claim (tax) exception on the donated amount’

(127) Compound nouns

a. (Conjunct-verb-DS-1) Compound N:

usane apanA Ora piSAca kA **vriwwamwa varnana** kiya
 usane apanA Ora piSAca kA **vriwwAMwa varNana** kiya
 he-Erg own and devil of **narration description** do.Pfv.MSg

‘He told his own story and the story of the ghost.’

b. (Relation-DS-rad-2) Compound N:

mastara sahaba, kya kala skula kula he
mAstara sAhaba, kA kala skUla Kula hE
master hon what tomorrow school open be.Prs.Sg

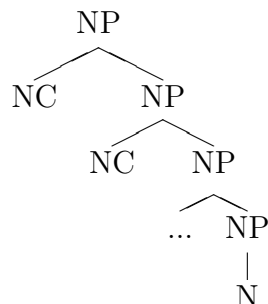
‘Teacher, is the school open tomorrow?’

c. (Relation-DS-fragof-1): Compound NProp:

BAkaPA (mAovAxi) ke **rAmabacana yAxava** ko giraPZawAra kara
 BAkaPA (mAovAxi) ke **rAmabacana yAxava** ko giraPZawAra kara
 Bakapa (Maosist) of **Rambacana Yadav** Dat arrest do
 liyA gayA
 liyA gayA
 TAKE.Pfv GO.Pfv.MSg

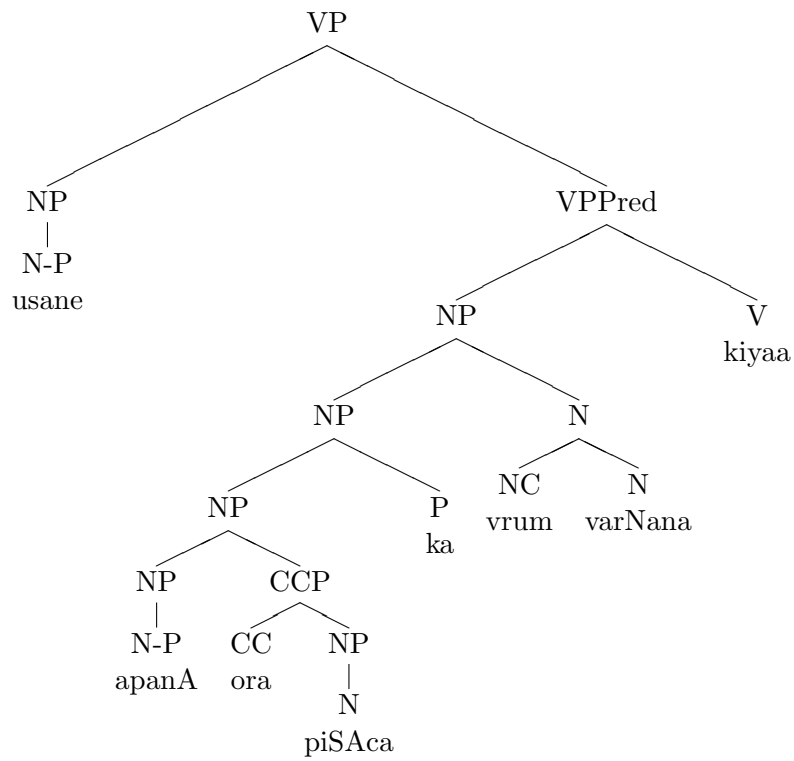
‘Rambacana Yadav of BKP (Maoist) has been arrested.’

The treatment of compounds involves adjunction of the linearly left words in the compound to the NP projected by the rightmost word in the compound, as shown schematically below.

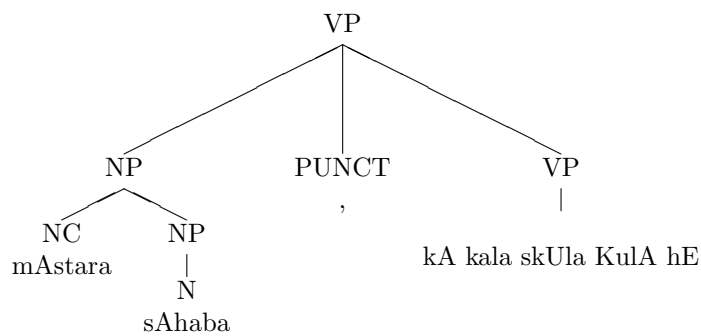
(128) *Compound analysis* (Generalized-Structure-of-Compounds)

The X in the label XC is a variable that is replaced in the actual annotation by the relevant part of speech tag. As such, the label for non-final words in non-proper compound nouns is NC , the label for non-final parts of compound proper name parts is $NPropC$.

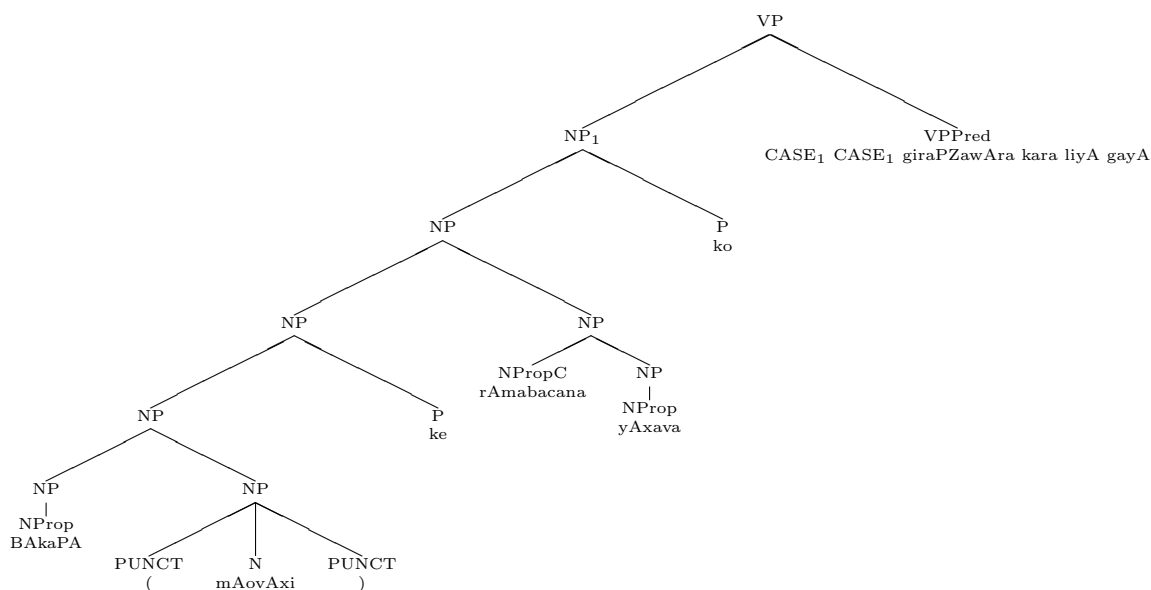
(129) a. Tree for (127a) (Conjunct-verb-DS-1) Compound noun:



b. Tree for (127b) (Relation-DS-rad-2) Compound noun:



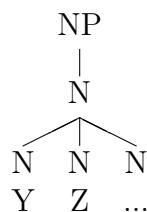
c. Tree for Relation-DS-fragof-1 (Relation-DS-fragof-1) Compound name:



25.3.1 Additional considerations and alternative analyses

While we are compelled by conversion considerations to provide a syntactic parse for compounds we do not consider the internal structure of compound nouns to be syntactically interesting; our assumption is that the syntax manipulates only the compound as a whole. Given this view, a conceivable PS analysis of compound nouns would involve multiple branching at the N level, with every word of the compound a daughter of the N node, as shown below.

- (130) *Alternative, non-binary branching compound analysis*
(not used because of binary branching constraint)



This alternative analysis correctly captures the lack of syntactic independence of the compound's internal morphemes; however, it results in nonbinary branching because the sister of NP, N, has multiple daughters.⁴ We choose to avoid such non-binary branching structures in favor of the structures above, which comply with the uniform non-binary branching desideratum.

⁴In order to facilitate conversion, we avoid this type of nonbinary branching structure—where an X daughter of XP has two or more daughters—to the extent possible throughout these guidelines.

25.4 Nominal Appositives

In appositive structures, two NPs occur consecutively, and the second NP functions as an appositive of the first, which it specifies further.⁵

(131) a. (Nominal-Appositives-PS-1)

vahAM para wEnAwa surakRAkarmiyom ne
 vahAM para wEnAwa surakRAkarmiyom ne
 there on stationed security.workers Erg
bOYIIvuda aBinewrI bipASA basu ke sAWa xuyarvahAra kiyA
bOYIIvuda aBinewrI bipASA basu ke sAWa xuyarvahAra kiyA
Bollywood beauty Bipasha Basu Gen with misbehavior do.Pfv.MSg
 ‘Security guards stationed there misbehaved with Bollywood actress Bipasha Basu.’

b. (Nominal-Appositives-PS-2)

prakASa JA kI **naI Pilma apaharaNa** kA BI prImiyara honA
 prakASa JA kI **naI Pilma apaharaNa** kA BI prImiyara honA
 Prakash Jha Gen **new film *Abduction*** Gen also premiere be-Inf
 hE
 hE
 be.Prs.Sg
 ‘The premiere of Prakash Jha’s new film *Abduction* also has to take place.’

c. (Nominal-Appositives-PS-3)

xevAnaMxa ne **apanI naI Pilma mistara prAima ministara** kA
 xevAnaMxa ne **apanI naI Pilma mistara prAima ministara** kA
 Dev.Anand Erg **his new film Mr. Prime Minister** Gen
 praxarSana nahIM karane kA PEsaLA kiyA hE
 praxarSana nahIM karane kA PEsaLA kiyA hE
 display Neg do-Inf Gen decision do.Pfv.Msg be.Prs.Sg
 ‘Dev Anand has decided not to show his new film *Mr. Prime Minister*.’

Such appositive structures raise many interesting theoretical questions, which we do not grapple with here. Our analysis involves NP adjunction of the appositive NP to the modified NP. The appositive NP is labelled *NP-App*. The following trees, which give the phrase structures of the first two examples above, illustrate the analysis.

Tree for (??) (Nominal appositives-PS-1):

Tree for (??) (Nominal appositives-PS-2):

⁵The example sentences in (131) are abbreviated versions of sentences from the corpus: (131a) correspond to file 1, Sentence id=2, Sentence id=7, and Sentence id=14, respectively.

Chapter 26

Adjectival Phrases and Comparatives

What constituents are associated with the AP? We find other APs (*halkaa* ‘light’ in *halkaa niilaa* ‘light blue’), degree heads (*kam* ‘less’, *zyadaa* ‘more’), degree phrases (*Ram se* ‘Ram than’), degree modifiers (*bahut* ‘very’), and differential/measure phrases (*do kilo* ‘two kilos’).

We make a distinction between degree phrases and the other modifiers. The other modifiers all attach directly or indirectly via adjunction to the AP that they modify.

(132) Attachment to AP

- a. AP modifiers like *halkaa* ‘light’ in *halkaa niilaa* ‘light blue’
- b. degree heads (*kam* ‘less’, *zyadaa* ‘more’)
- c. degree modifiers (*bahut* ‘very’)
- d. measure phrases (*do kilo* ‘two kilos’)
- e. differential phrases (*do kilo* ‘two kilos’) in the absence of an overt degree head
- f. degree modifiers (*bahut* ‘very’) in the absence of an overt degree head

Note that degree phrases can attach to a wide range of categories, not just to APs.

Hindi-Urdu comparatives do not always involve an overt comparative degree head. It is possible to treat such cases as involving a silent comparative degree head. However, for reasons of simplicity, we will not represent silent comparative degree heads. If an overt degree head is present, differential phrases and degree modifiers are attached to the Degree Phrase, which is in turn adjoined to the AP. When there is no overt comparative degree head, differential phrases and degree modifiers are directly attached to AP.

(133) Attachment to DegP

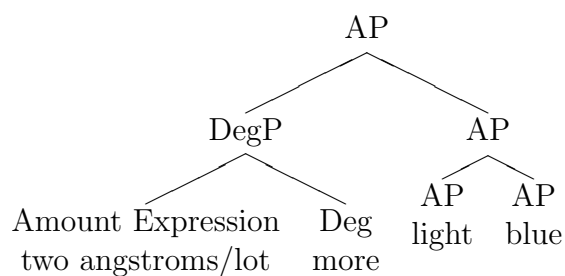
- a. differential phrases (*do kilo* ‘two kilos’) in the presence of an overt degree head
- b. degree modifiers (*bahut* ‘very’) in the presence of an overt degree head

Even though degree phrases such as *Ram se* ‘Ram than’ are licensed by the presence of a degree head, we do not generate degree phrases together with the degree head. Instead these are treated as clausal adjuncts. Degree phrases bear an index **compref**, which indicates

which NP they are compared with; thus in *John is taller than Bill*, *John* and *Bill* will have the same **compref**. In the trees here, we indicate this via a superscript preceded by **c**.

Like PPs and NPs, APs and DegPs do not involve an A' or Deg'-level. The following is a schematic structure.

- (134) (Label:Generalized-Structure-of-AdjP)
This is two angstroms/a lot more light blue.

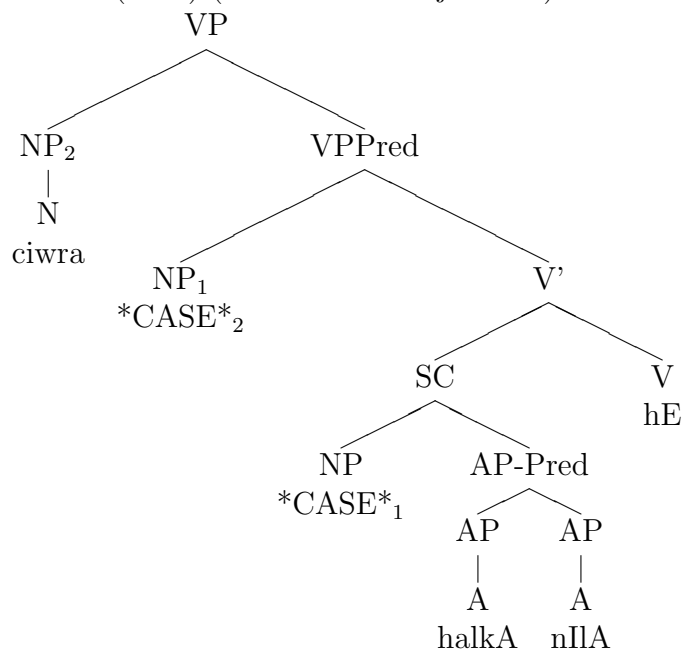


Apart from the head A, any of the elements in the above tree could be missing. The adjective *halkaa* 'light' is adjoined to the main AP.

- (135) a. (Label:Structure-of-AdjP-PS-0) AP-modifier of A
 chitra [halkaa niila:] hai
 ciwra halkA nIIA hE
 picture.m light.MSg blue.MSg be.Prs.Sg

'The picture is light blue.'

Tree for (135a) (Structure-of-AdjP-PS-0):

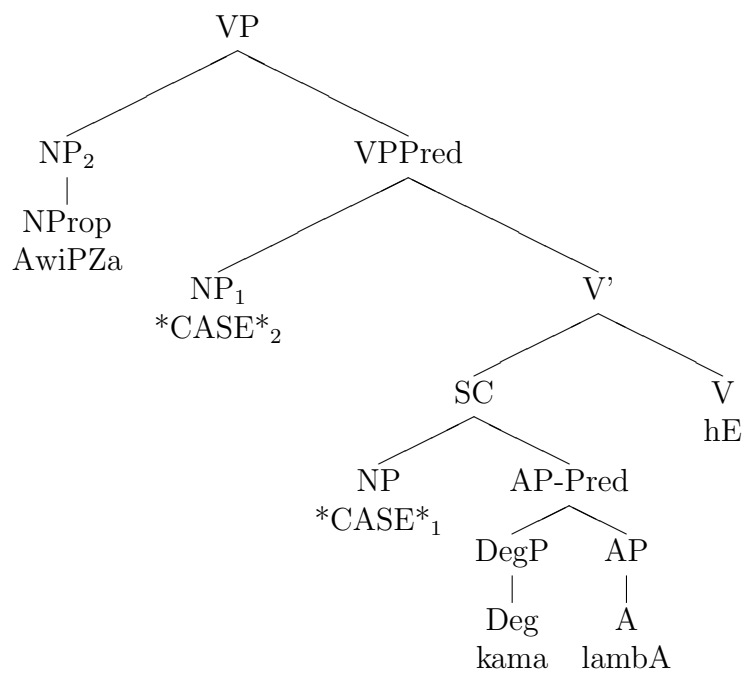


b. (Label:Structure-of-AdjP-PS-1) Degree Heads (more/less)

Atif [zyAxaaa/kam lambaa] hai
 AwiPZa jZyAxA/kama lambA hE
 Atif.m more/less tall.MSg is

‘Atif is taller/less tall.’

Tree for (135b) (Structure-of-AdjP-PS-1):

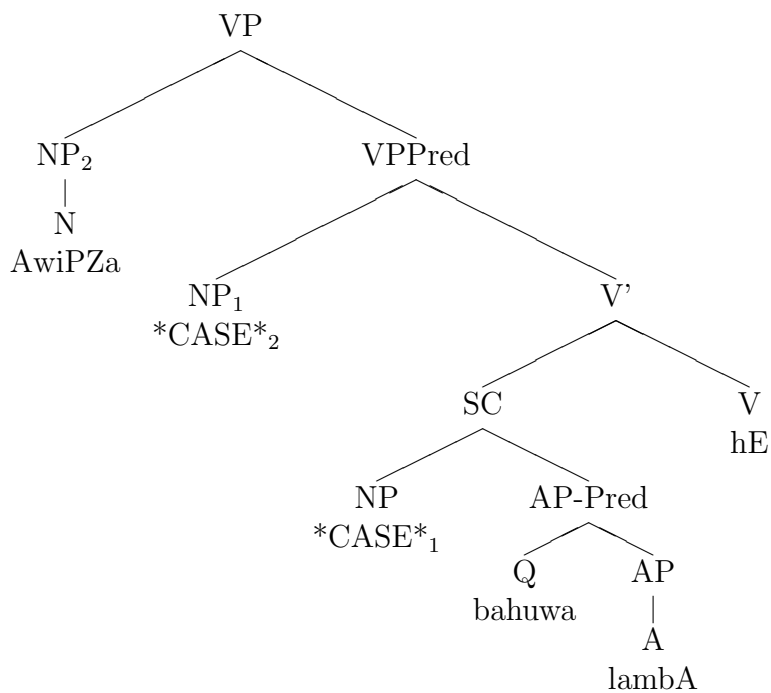


c. (Label:Structure-of-AdjP-PS-2) Degree Modifiers (extremely)

Atif [bahut lambaa] hai
 AwiPZa bahuwa lambA hE
 Atif.m very tall.MSg is

‘Atif is very tall.’

Tree for (135c) (Structure-of-AdjP-PS-2):

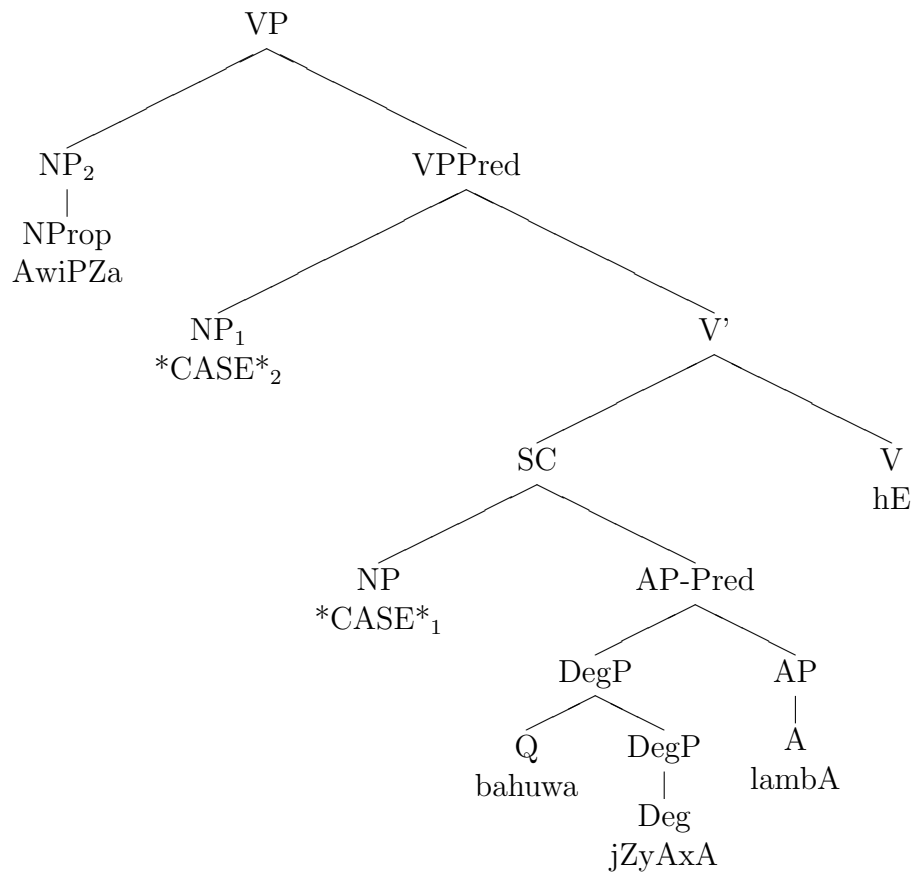


- d. (Label:Structure-of-AdjP-PS-21) Degree Modifier + Degree Head *zyaadaa* ‘more’

Atif [bahut zyaadaa lambaa] hai
 AwiPZa bahuwa jZyAxA lambA hE
 Atif.m very more tall.MSg is

‘Atif is very tall.’

Tree for (135d) (Structure-of-AdjP-PS-21):

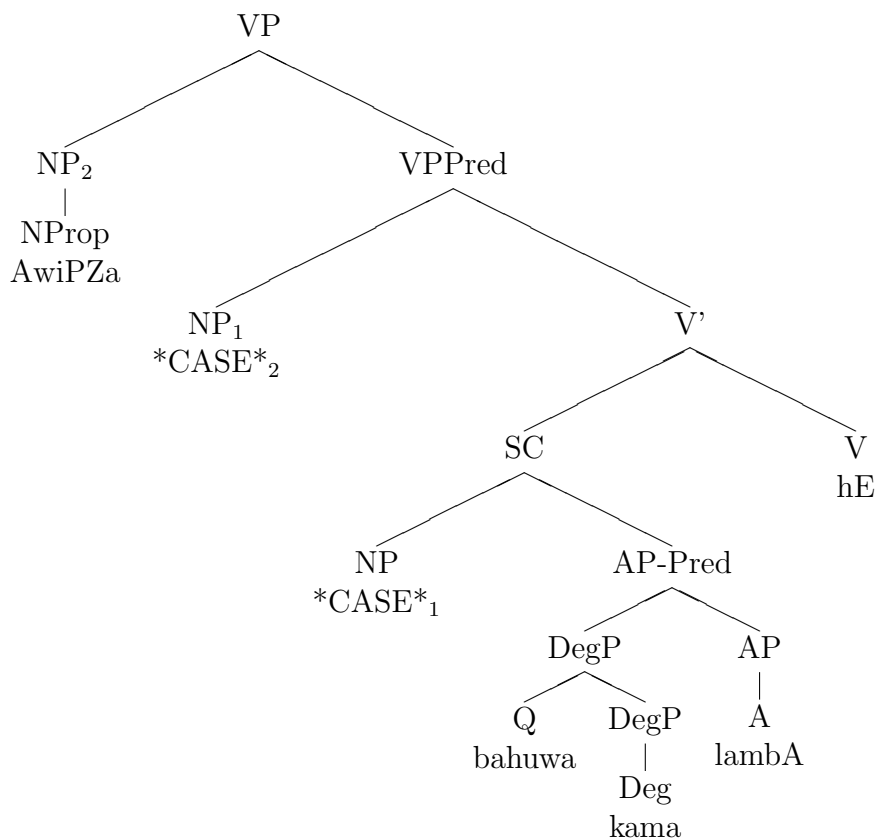


- e. (Label:Structure-of-AdjP-PS-22) Degree Modifier + Degree Head *kam* 'less'

Atif [bahut kam lambaa] hai
 AwiPZa bahuwa kama lambA hE
 Atif.m very less tall.MSg is

'Atif is very short.'

Tree for (135e) (Structure-of-AdjP-PS-22):

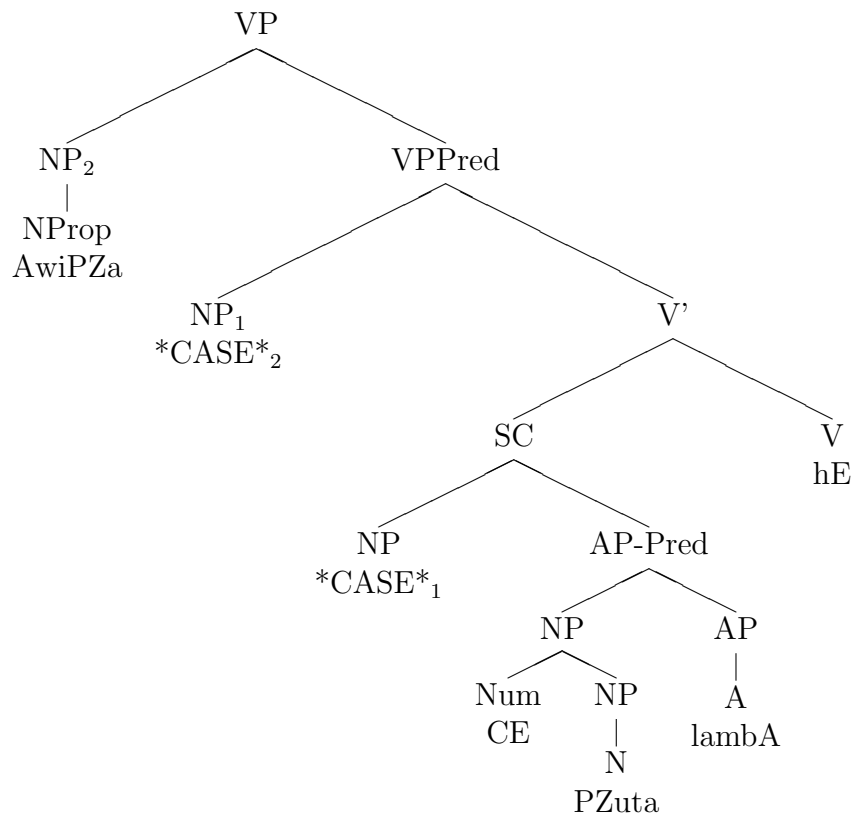


- f. (Label:Structure-of-AdjP-PS-3) Measure Phrases (6ft.)

Atif [6 foot lambaa] hai
 AwiPZa CE PZuta lambA hE
 Atif.m 6 foot tall.MSg is

‘Atif is 6 foot tall.’

Tree for (135f) (Structure-of-AdjP-PS-3):

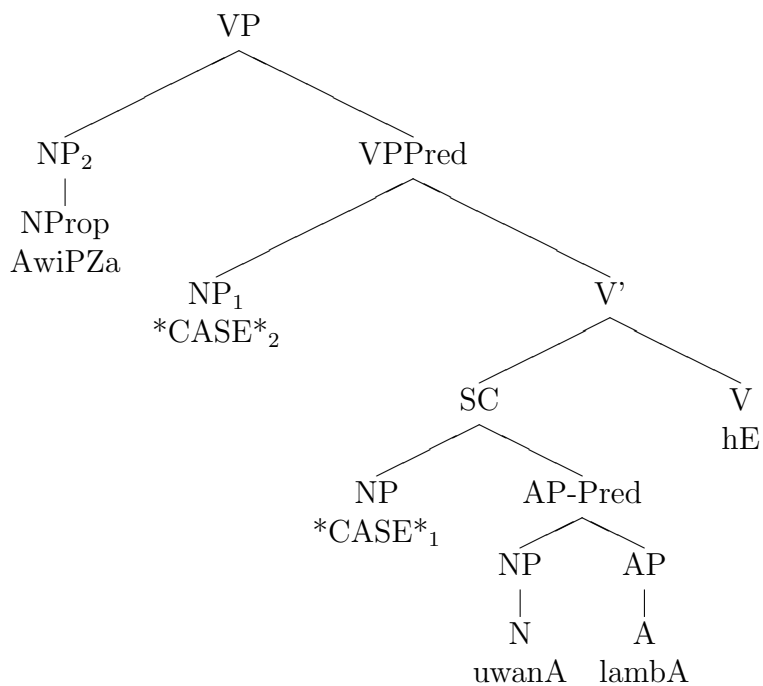


g. (Label:Structure-of-AdjP-PS-31) Measure Phrases (*utnaa* 'that much')

Atif [utnaa lambaa] hai
 AwiPZa uwanA lambA hE
 Atif.m that.much tall.MSg is

'Atif is that tall.'

Tree for (135g) (Structure-of-AdjP-PS-31):



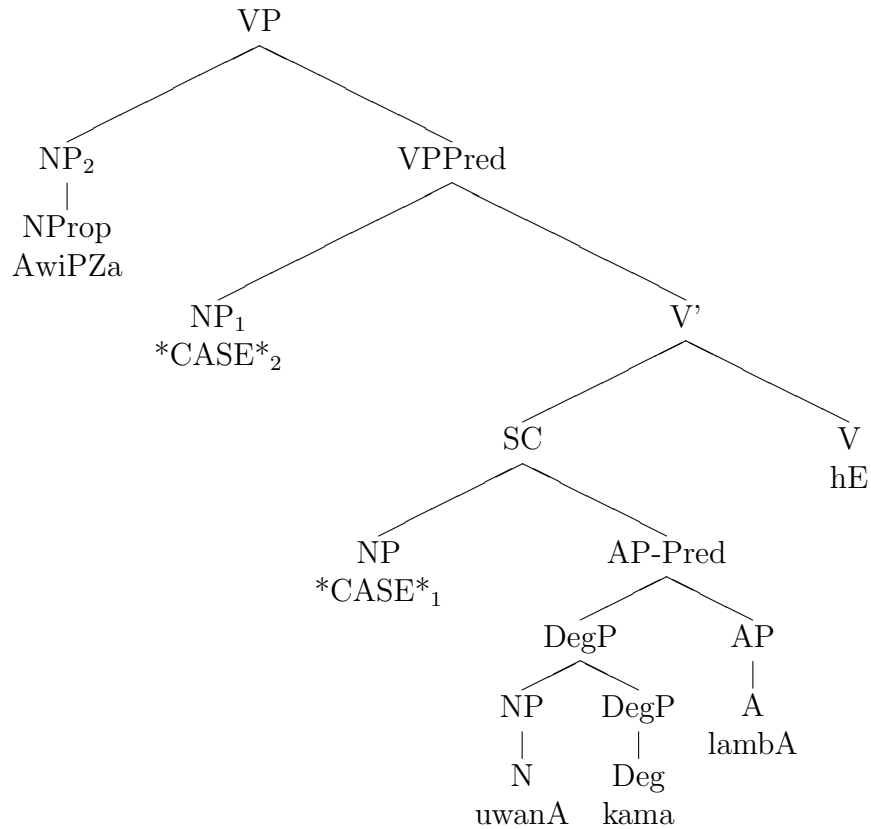
- h. (Label:Structure-of-AdjP-PS-31) Measure Phrases (*utnaa* ‘that much’)

Atif [utnaa kam lambaa] hai
 AwiPZa uwanA kama lambA hE
 Atif.m that.much less tall.MSg is

‘Atif is that much less tall.’

ocr Check category of *utnaa*

Tree for (135h) (Structure-of-AdjP-PS-31):

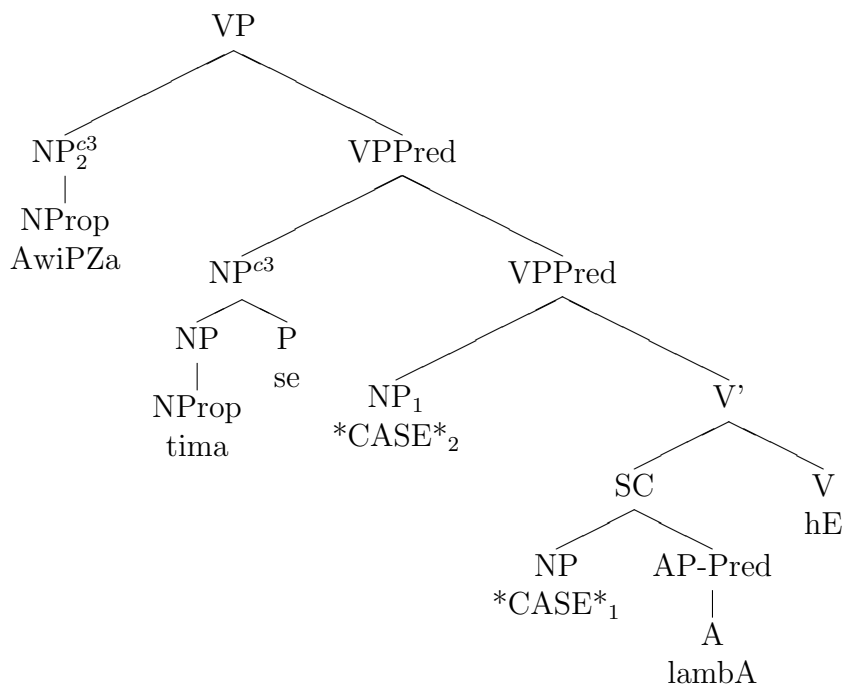


- i. (Label:Structure-of-AdjP-PS-4) Degree Phrases 1 *more*: (than John): with most adjectives, the degree head *zyaadaa* is silent; we do not represent silent degree heads.

Atif Tim se lambaa hai
 AwiPZa tima se lambA hE
 Atif.m Tim than tall.MSg is

‘Atif is taller than Tim.’

Tree for (135i) (Structure-of-AdjP-PS-4):

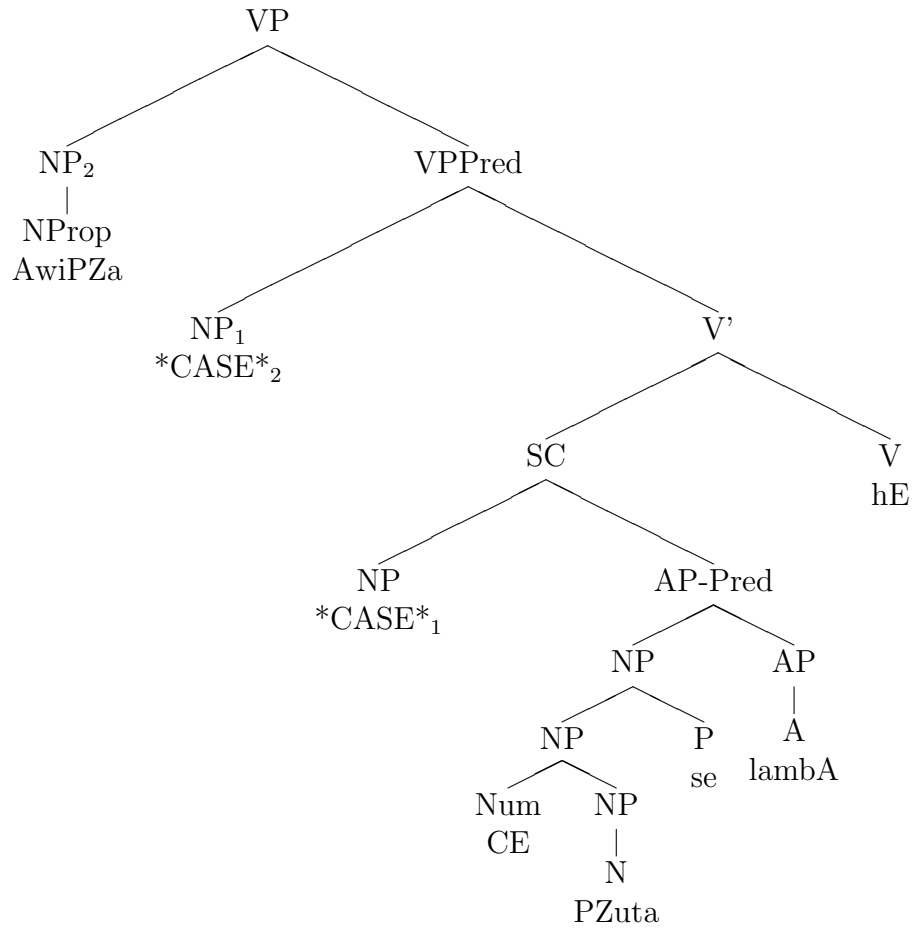


- j. (Label:Structure-of-AdjP-PS-41) Degree Phrases 1 *more*: (than 6 foot): with most adjectives, the degree head *zyAxaana* is silent. This is not a direct comparison between individuals and therefore the *than 6 foot* is attached to the AP, and not to the VP. There is also no **compref**.

Atif [chE foot se lambaa] hai
 AwiPZa CE PZuta se lambA hE
 Atif.m 6 foot than tall.MSg is

‘Atif is taller than 6 feet.’

Tree for (135j) (Structure-of-AdjP-PS-41):

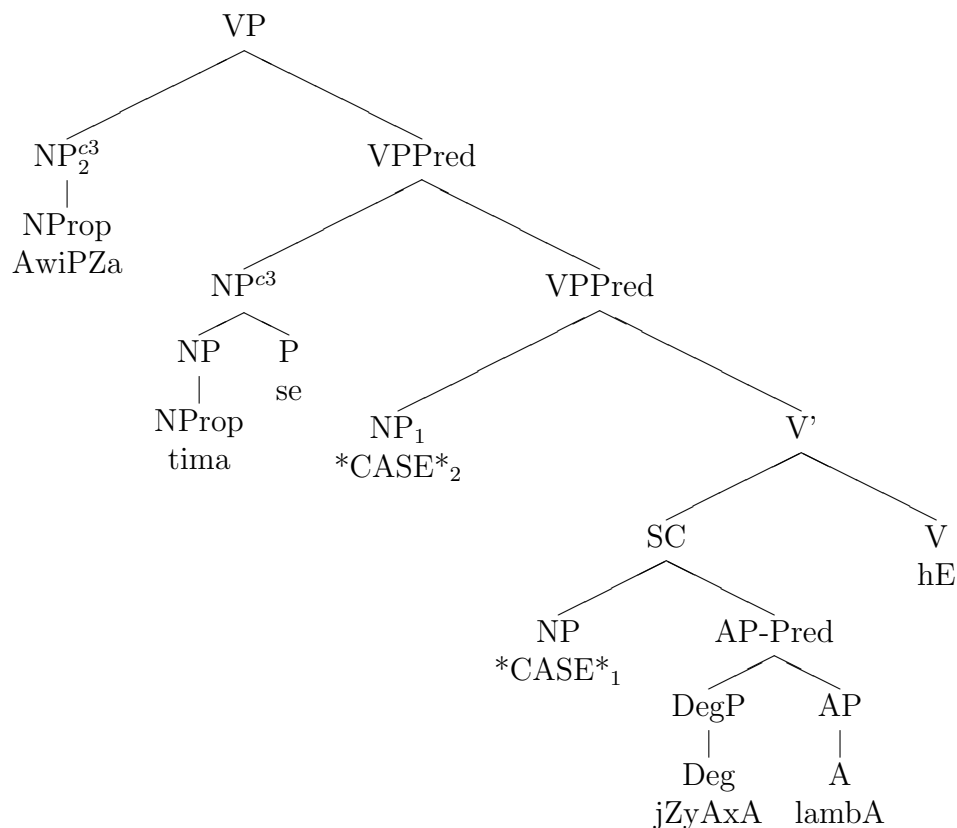


- k. (Label:Structure-of-AdjP-PS-42) Degree Phrases 1 *more*: (than John):

Atif Tim se zyaadaa lambaa hai
 AwiPZa tima se jZyAxA lambA hE
 Atif.m Tim than more tall.MSg is

‘Atif is taller than Tim.’

Tree for (135k) (Structure-of-AdjP-PS-42):

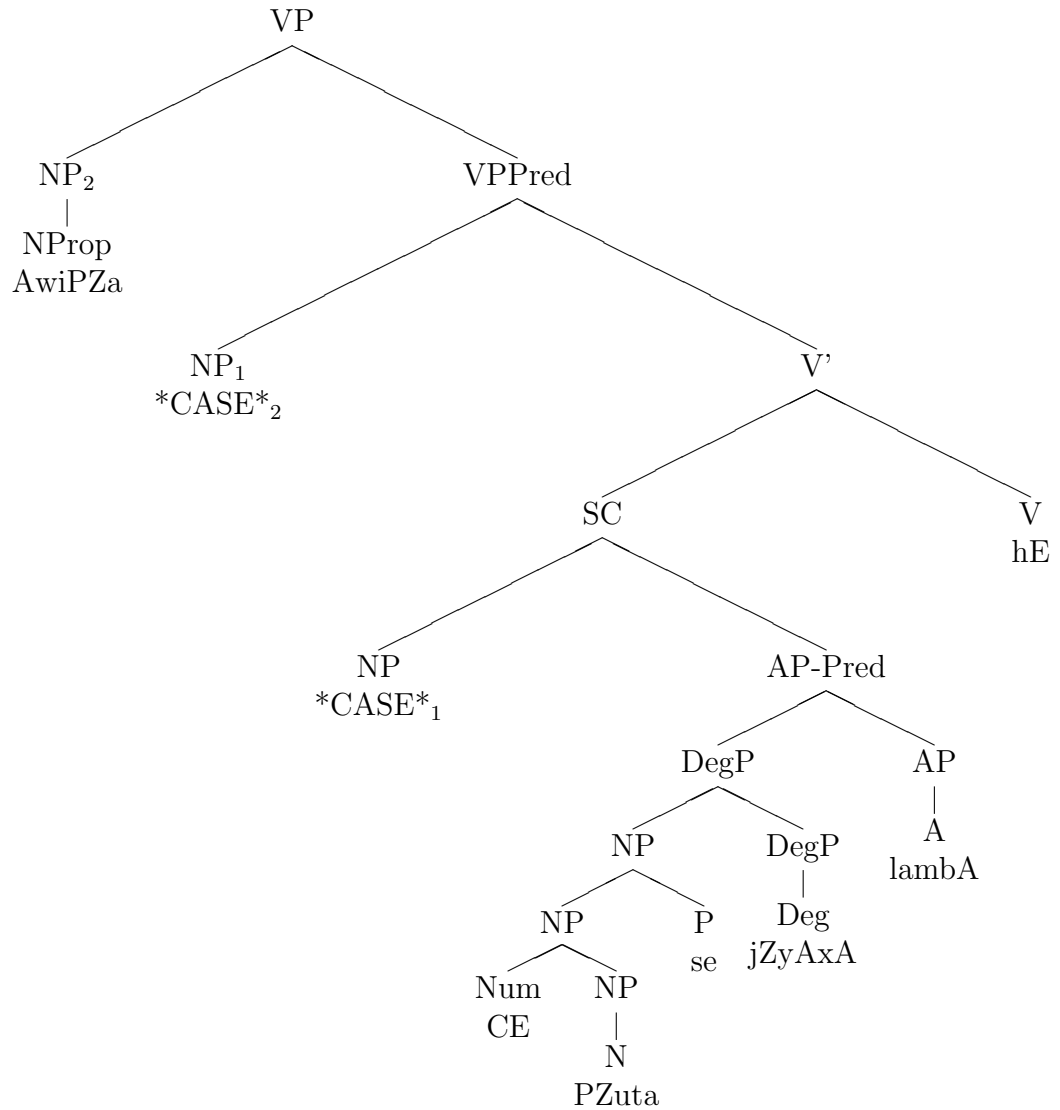


1. (Label:Structure-of-AdjP-PS-43) Degree Phrases 1 *more*: (than 6 foot): This is not a direct comparison between individuals and therefore the *than 6 foot* is attached to the DegP, and not to the VP. There is also no **compref**.

Atif [chE foot se zyaadaa lambaa] hai
 AwiPZa CE PZuta se jZyAxA lambA hE
 Atif.m 6 foot than tall.MSg is

‘Atif is taller than 6 feet.’

Tree for (135l) (Structure-of-AdjP-PS-43):

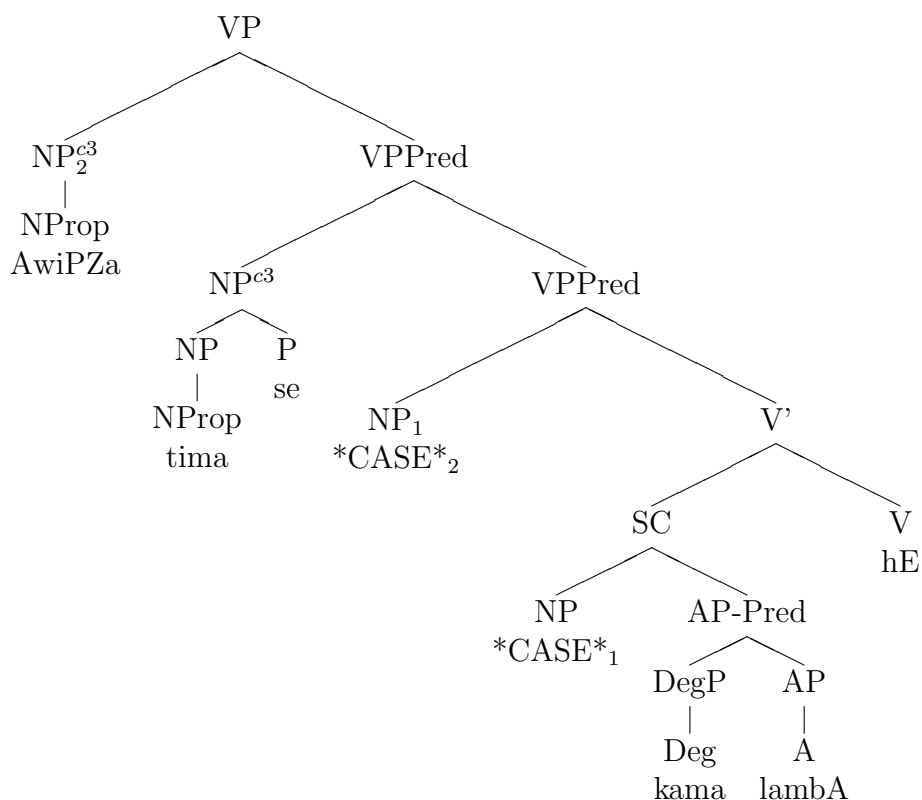


m. (Label:Structure-of-AdjP-PS-5) Degree Phrases 2 *less*: (than John)

Atif Tim se kam lambaa hai
 AwiPZa tima se kama lambA hE
 Atif.m Tim than less tall.MSg is

‘Atif is less tall than Tim.’

Tree for (135m) (Structure-of-AdjP-PS-5):

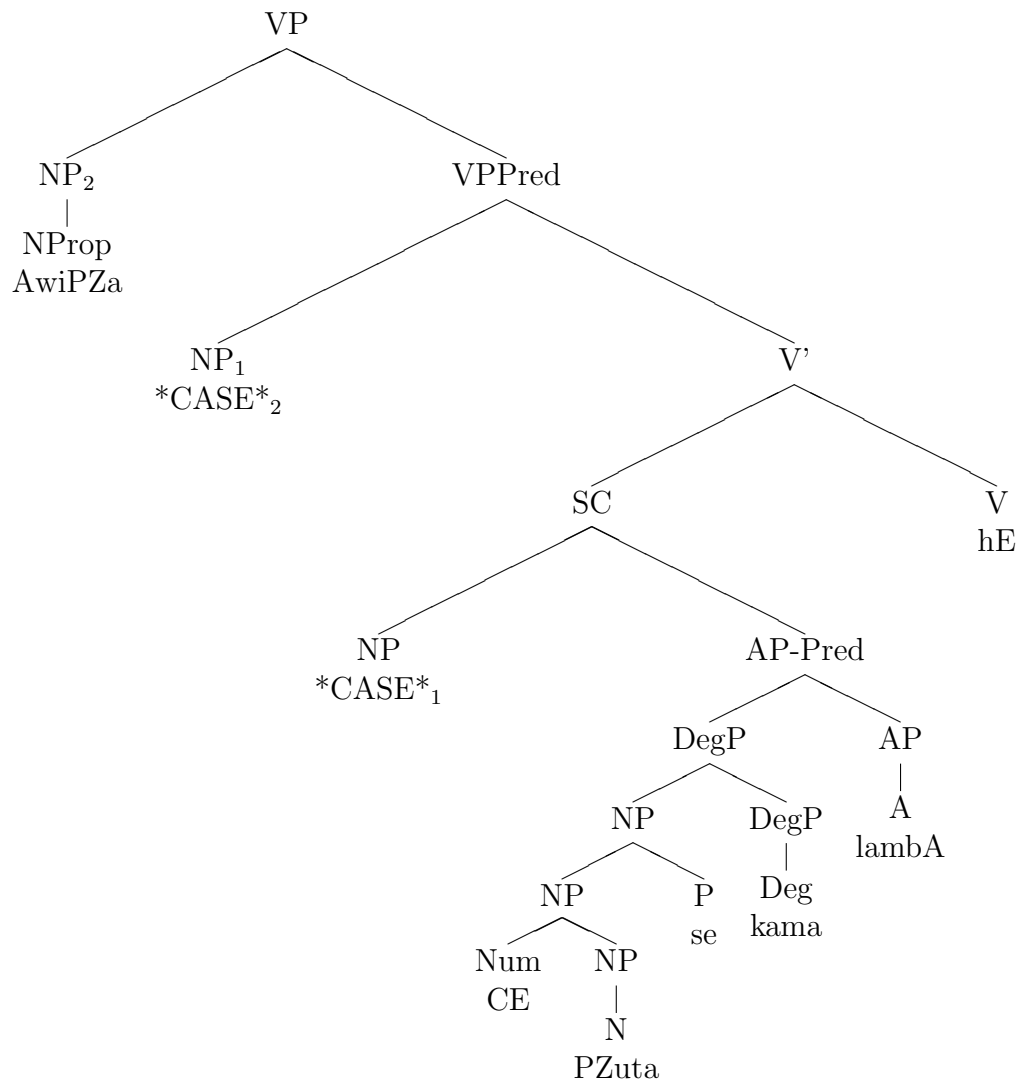


- n. (Label:Structure-of-AdjP-PS-51) Degree Phrases 2 *less*: (than 6 foot): This is not a direct comparison between individuals and therefore the *than 6 foot* is attached to the DegP, and not to the VP. There is also no **compref**.

Atif [chE foot se kam lambaa] hai
 AwiPZa CE PZuta se kama lambA hE
 Atif.m 6 foot than less tall.MSg is

‘Atif is less tall than 6 feet.’

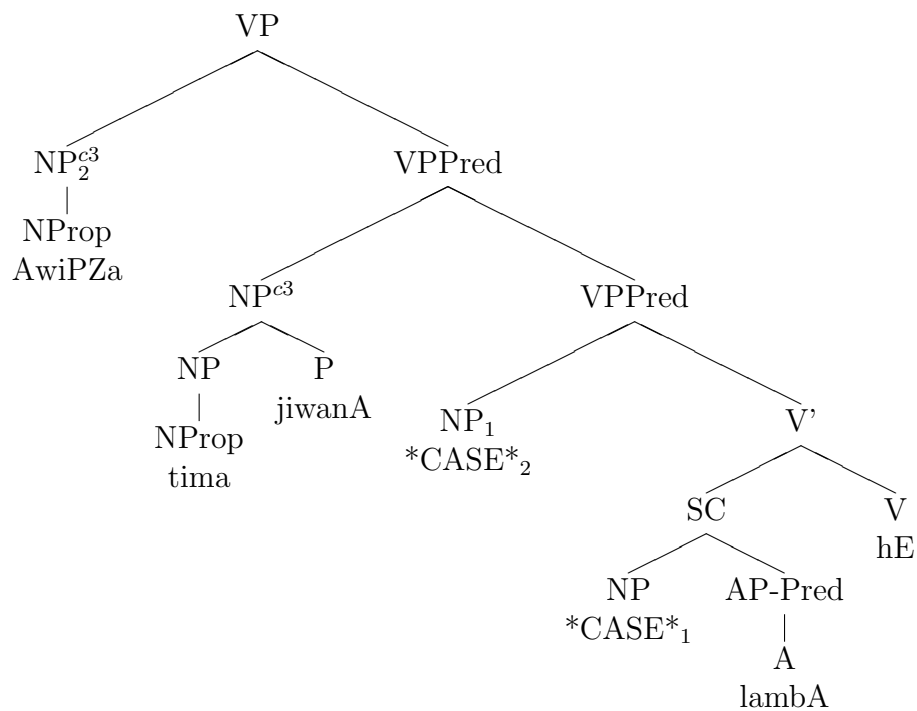
Tree for (135n) (Structure-of-AdjP-PS-51):



- o. (AdjP-PS-6) Degree Phrases 3 *as much*:

Atif Tim jitnaa lambaa hai
 AwiPZa tima jiwana lambA hE
 Atif.m Tim as.much less tall.MSg

‘Atif is as tall as Tim.’
 Tree for (135o) (Label: Structure-of-AdjP-PS-6)

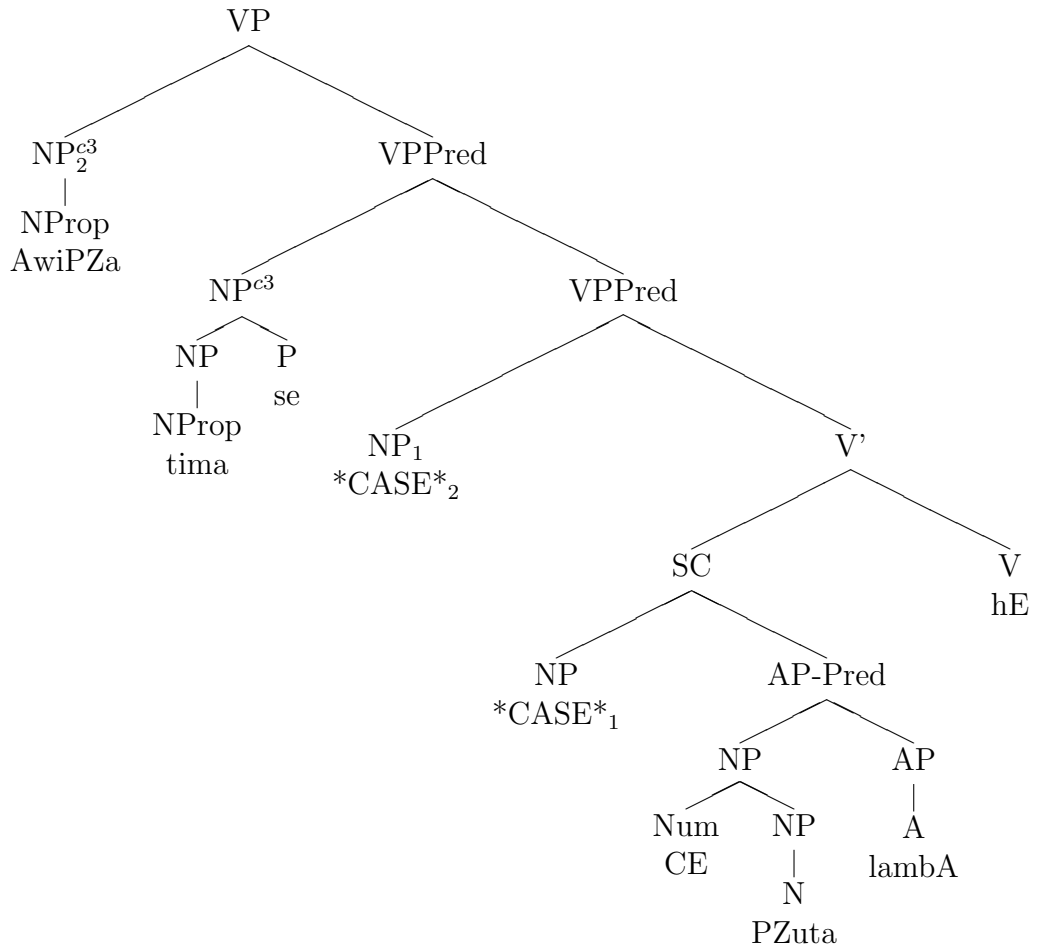


p. (Label: sent-AdjP-PS-81) Degree Phrases + Differential Amount Expression

Atif Tim se chE foot lambaa hai
 AwiPZa tima se CE PZuta lambA hE
 Atif.m Tim than tall.MSg is

‘Atif is taller than Tim.’

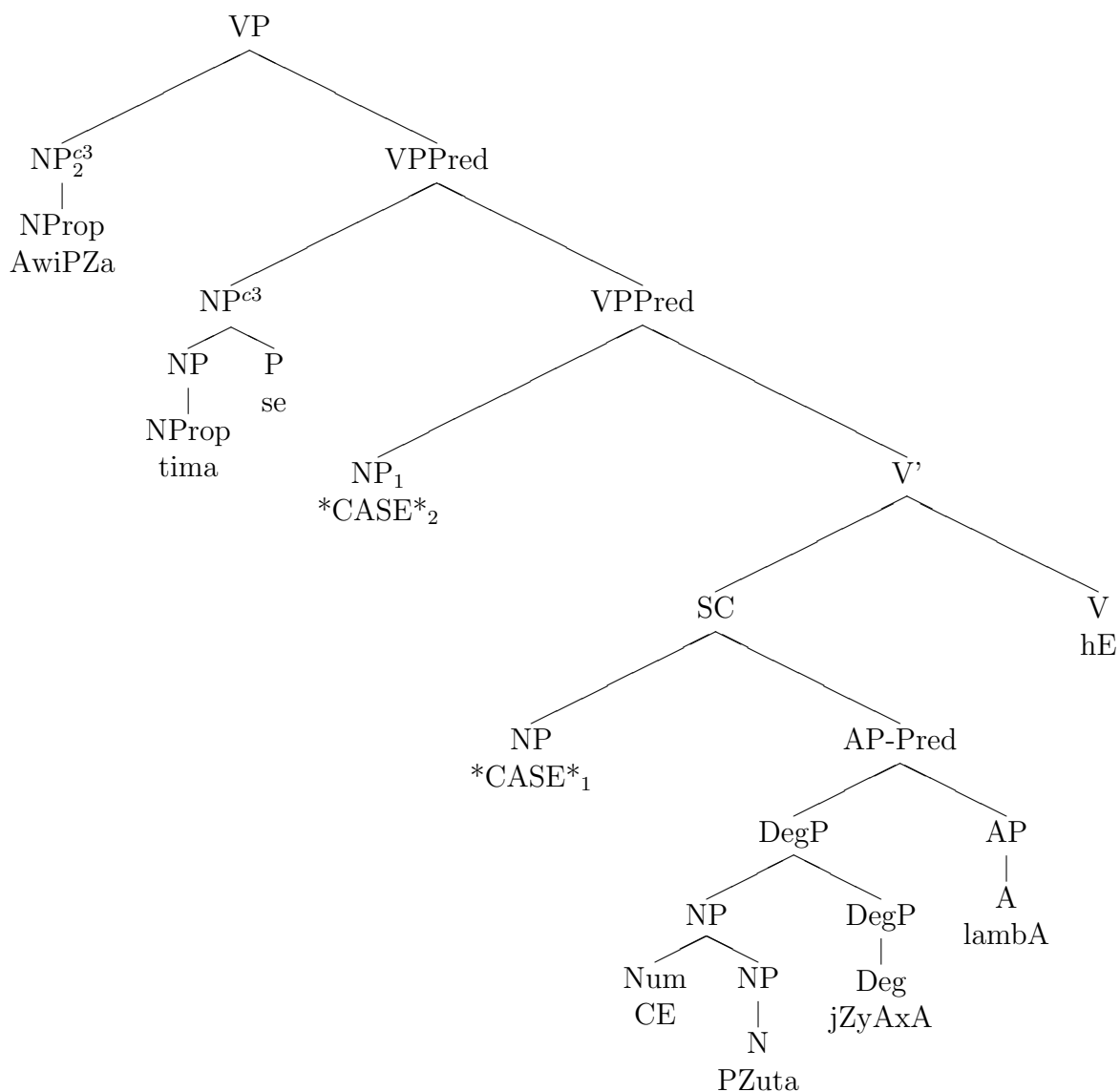
Tree for (135p) (Label: sent-AdjP-PS-81):



- q. (Label: AdjP-PS-82) Degree Phrases + Differential Amount Expression + Degree Head

Atif Tim se chE foot zyAxaaa lambaa hai
 AwiPZa tima se CE PZuta jZyAxA lambA hE
 Atif.m Tim than tall.MSg is

‘Atif is taller than Tim.’ Tree for (135q) (Label: Structure-of-sent-AdjP-PS-82)

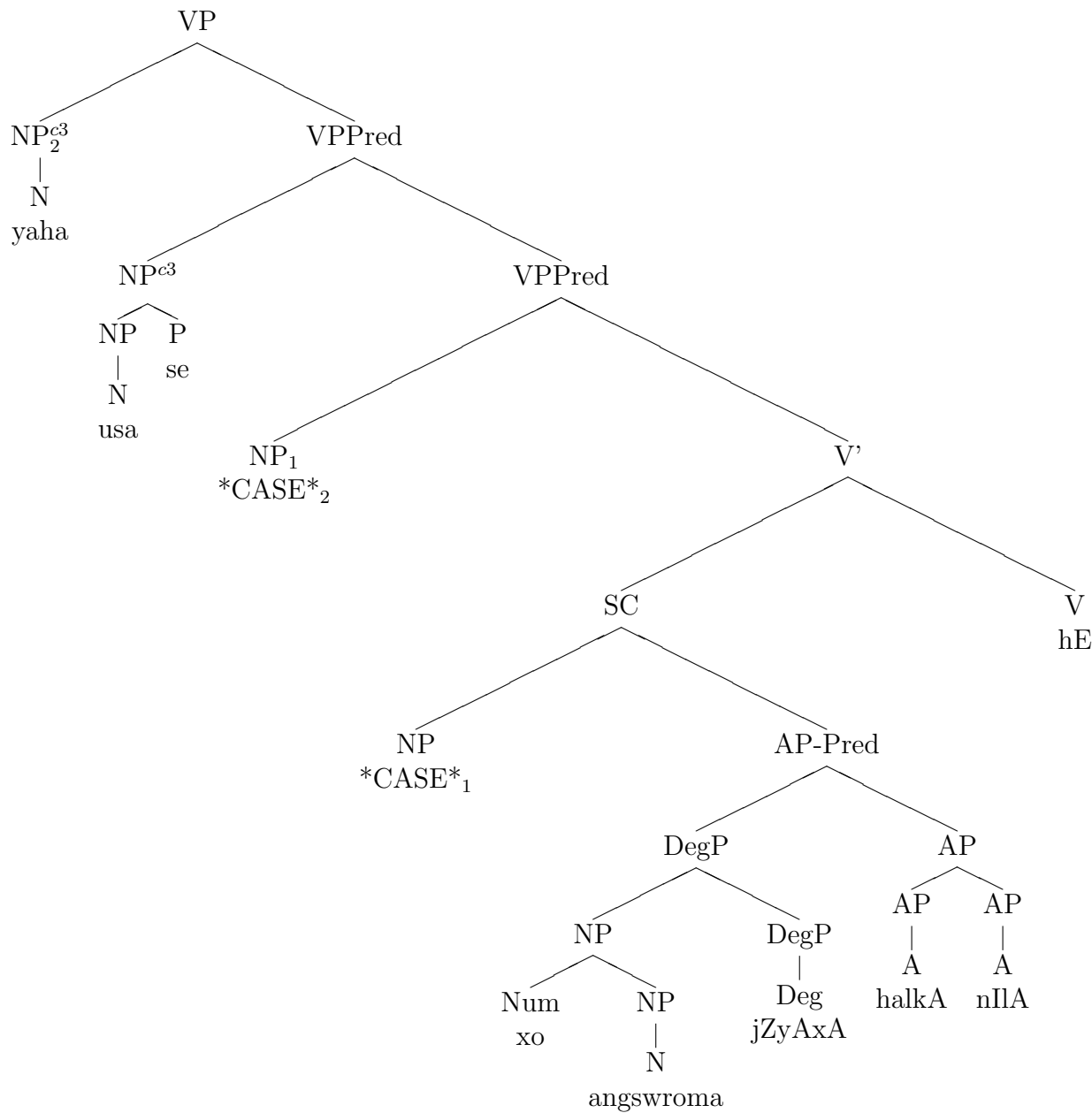


- r. (Label:Structure-of-AdjP-PS-83) Degree Phrases + Differential Amount Expression + Degree Head + AP modifier

yeh us se do angstrom zyAxaaa halkaa niilaa hai
 yaha usa se xo angswroma jZyAxA halkA nIIA hE
 this that than two angstrom more light blue be.Prs.Sg

‘This is two angstroms more light blue than that.’

Tree for (135r) (Structure-of-AdjP-PS-83):



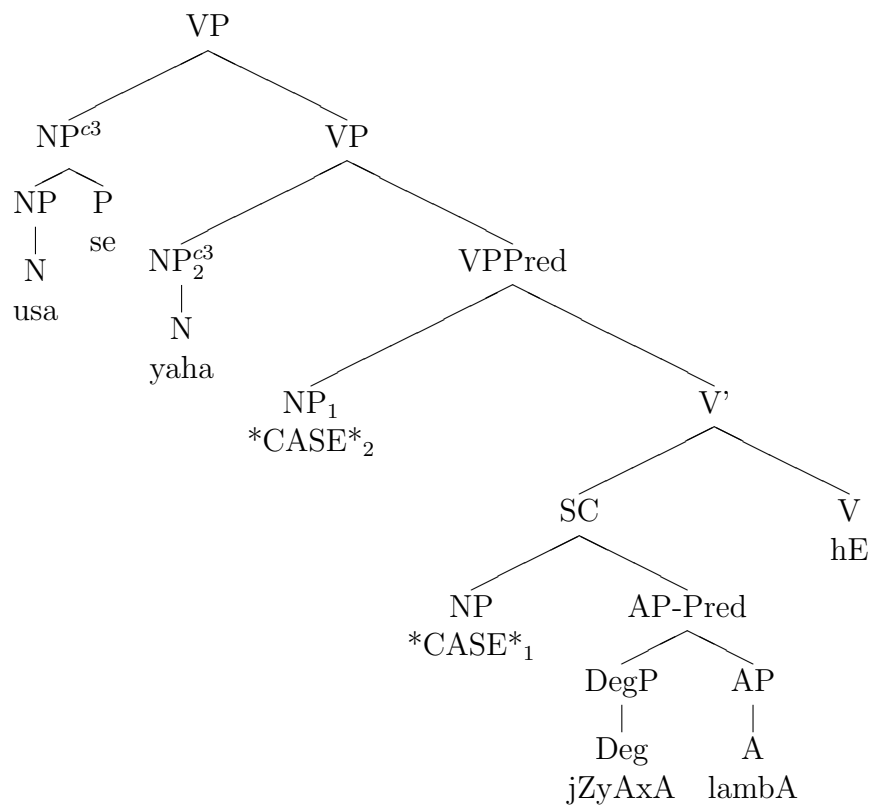
Since we treat most *than*-phrases as VP adjuncts, their local scrambling is not represented via traces. They are just adjoined where they appear.

(136) a. (Lanel:AP-Scrambling-PS-1)

us se yaha [zyAxaaa niilaa] hai
 usa se yaha jZyAxA nIIA hE
 that than this more blue be.Prs.Sg

‘This is more blue than that.’

Tree for (136a) (AP-Scrambling-PS-1):

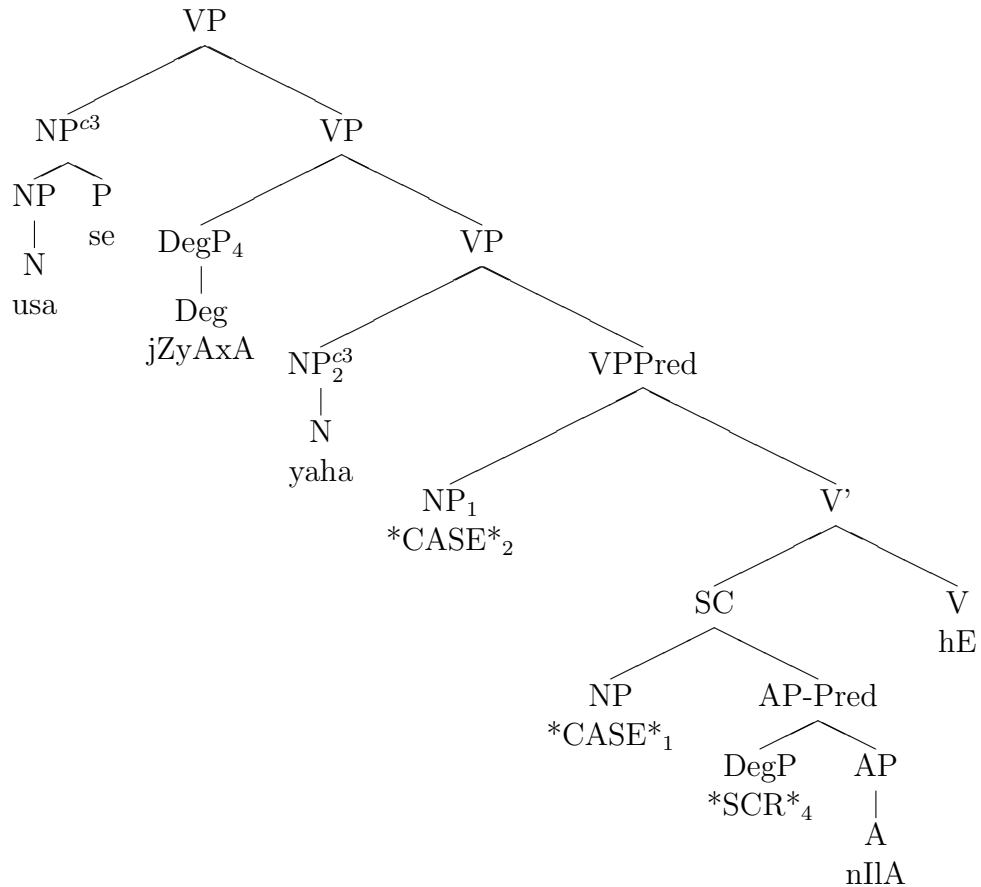


b. (Label:AP-Scrambling-PS-2)

us se zyAxaaa yeh [niilaa] hai
 usa se jZyAxA yaha nIIA hE
 that than more this blue be.Prs.Sg

‘Compared to that, this is more blue.’

Tree for (136b) (AP-Scrambling-PS-2):



(I am uncertain about the analysis of this example. I will do a mechanical treatment for now. My intuition is that here the DegP clearly forms a unit with the *-se* phrase.)

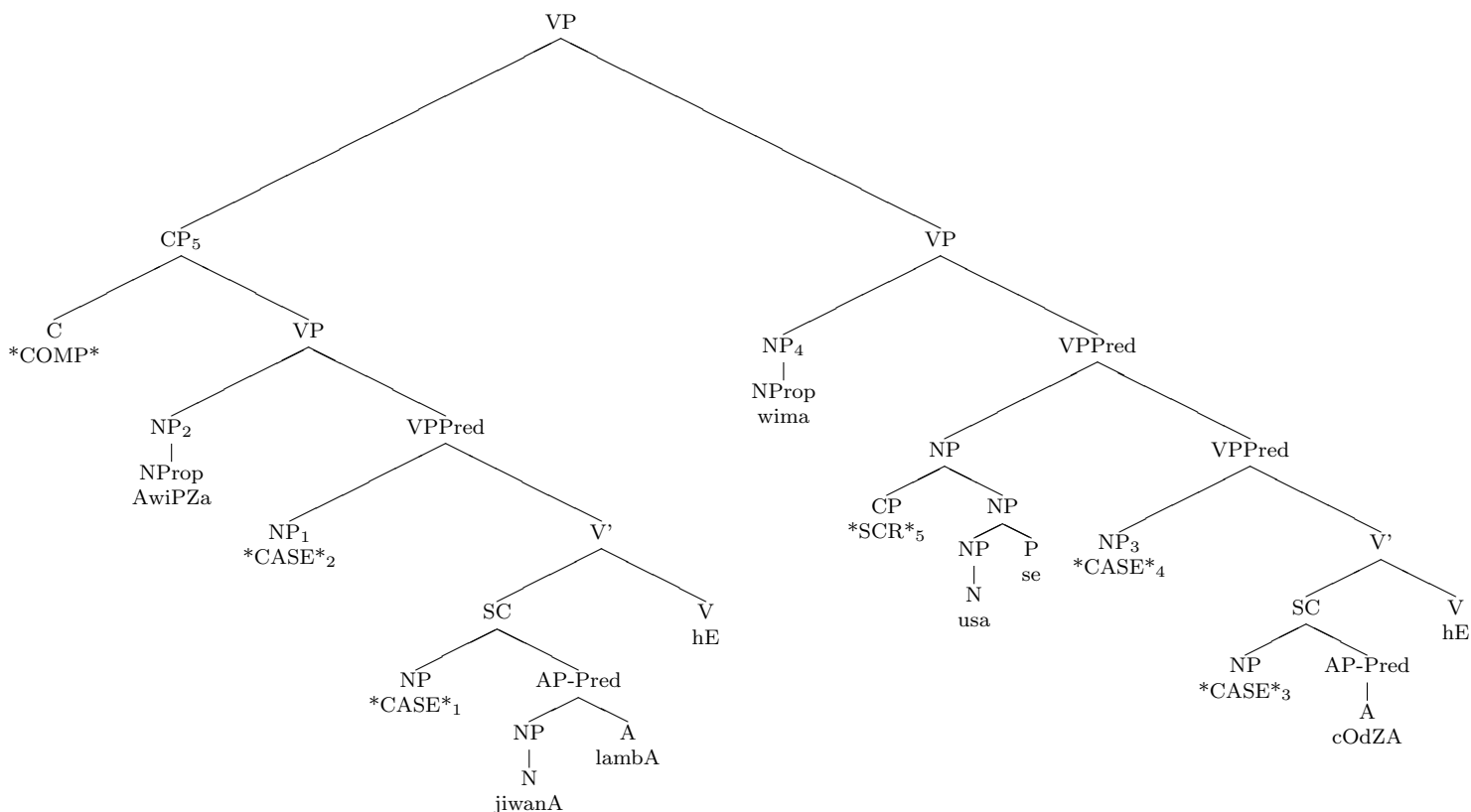
Cases like (135h) with *utnaa* ‘that.much’ are used together with a correlative clause to form clausal comparatives/equatives.

(137) (Label:AP-Comparative-PS-9)

| | | | | | | | | |
|--------|----------|--------|-----------|------|------|------|---------|-----------|
| [Atif | jitnaa | lambaa | hai], | [Tim | us | se | chauraa | hai] |
| AwiPZa | jiwanA | lambA | hE | wima | usa | se | cOdZA | hE |
| Atif | how.much | tall | be.Prs.Sg | Tim | that | than | wide | be.Prs.Sg |

‘Tim is wider than Atif is tall.’

Tree for (137) (AP-Comparative-PS-9):



The *than*-phrase here picks out an amount so it would be better to treat it as directly attached to the AP and not assign it a **compref**. This is also the case in the following example.

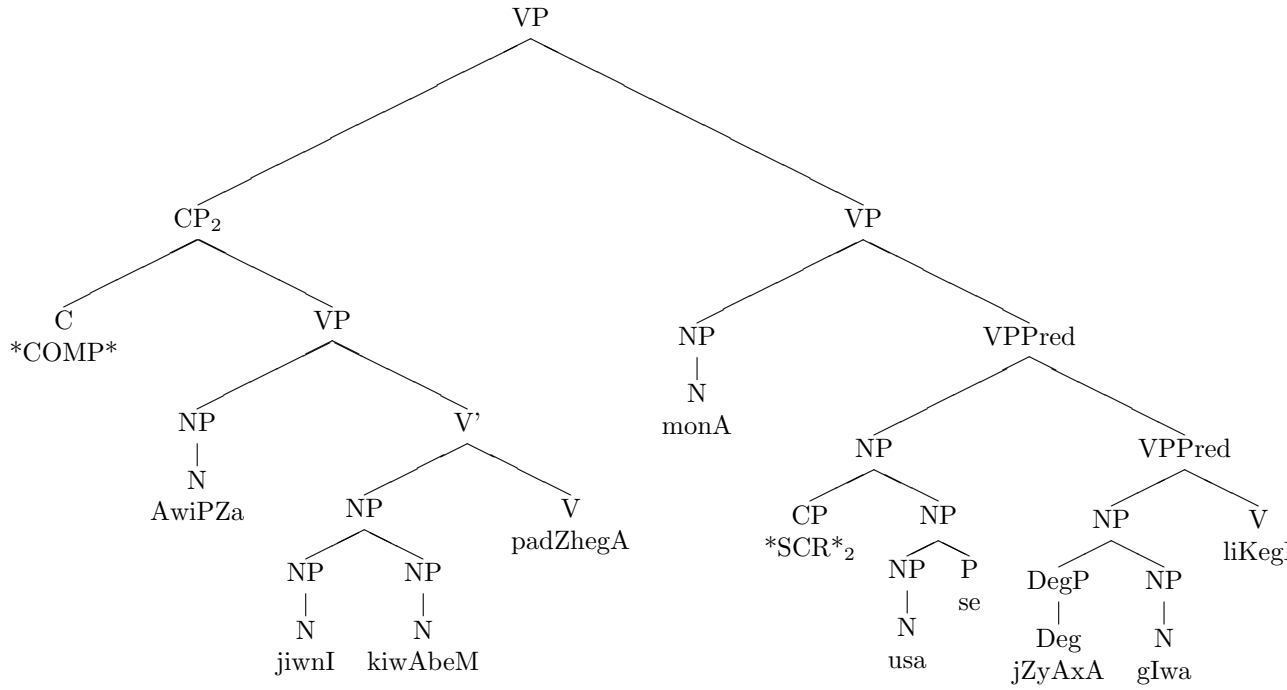
DegPs can exist independently of APs, for example in amount subcomparatives based on NPs.

(138) (Label: AdjP-PS-7)

[Atif jitnii kitaabē parh-egaa] [Mona us se zyAxaaa giit
 AwiPZa jiwnI kiwAbeM padZhegA monA usa se jZyAxA gIwa
 Atif how.many.f books read-Fut.3MSg Mona that than more song
 likh-egii]
 liKegI
 write-Fut.3MSg

‘Mona will write more songs than Atif will read books.’

Tree for (138) (Label:Structure of AdjP-PS-7)



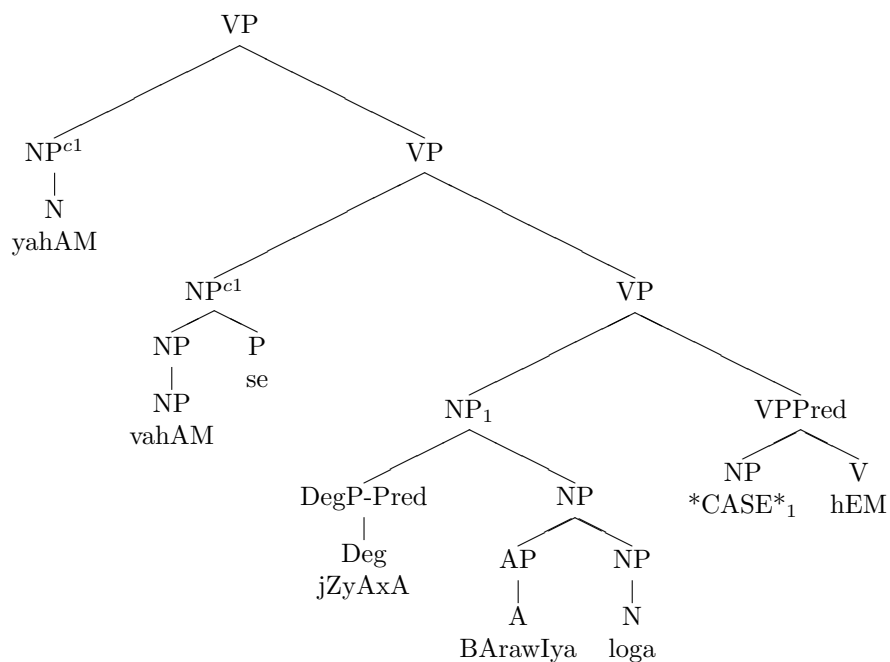
DegPs can also appear separate from an NP/AP. We will handle these cases as involving fronting of the NP that the DegP modifies.

- (139) a. (Label:Structure-of-AdjP-PS-71)
 DegP Predicate 1: DegP is on NP

yahã: vahã: se zyAxaaa bhaartiya log hË
 yahAM vahAM se jZyAxA BArawIya loga hEM
 here there than more Indian people be.Prs.Pl

‘There are more Indian people here than there.’

Tree for (139a) (Structure-of-AdjP-PS-71):

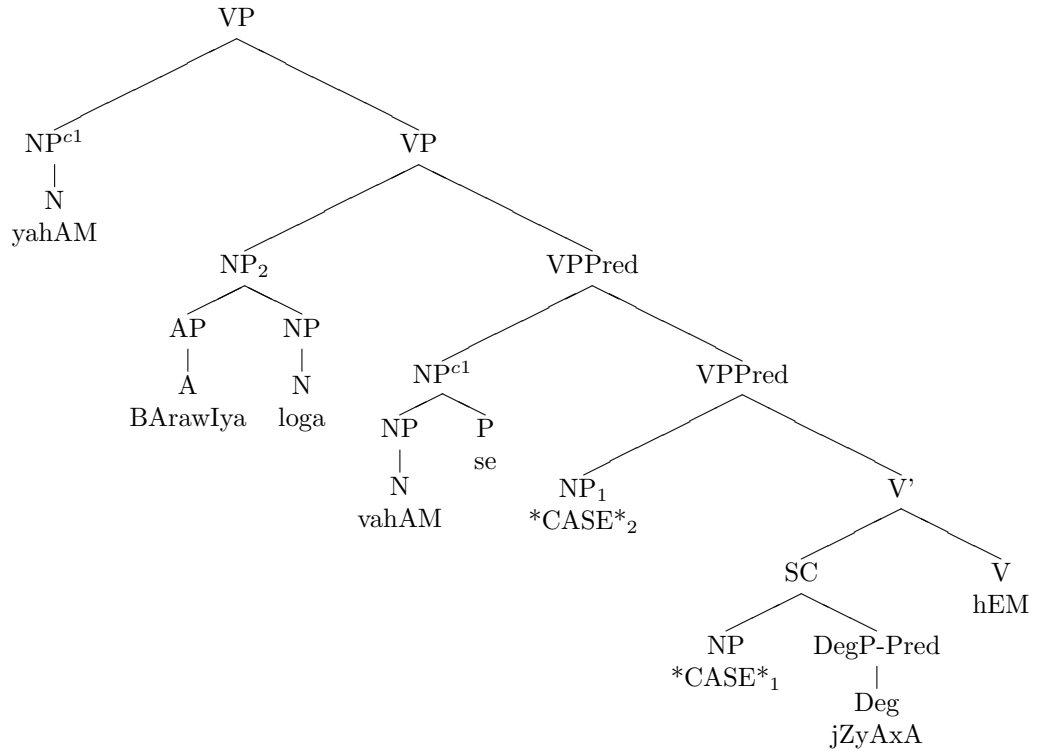


- b. (Label:Structure-of-AdjP-PS-72)
 DegP Predicate 2: DegP is discontinuous from NP

yahã: bhaartiya log vahã: se zyAxaaa hÊ
 yahAM BArawIya loga vahAM se jZyAxAA hEM
 here Indian people there than more be.Prs.Pl

‘There are more Indian people here than there..’

Tree for (139b) (Structure-of-AdjP-PS-72):



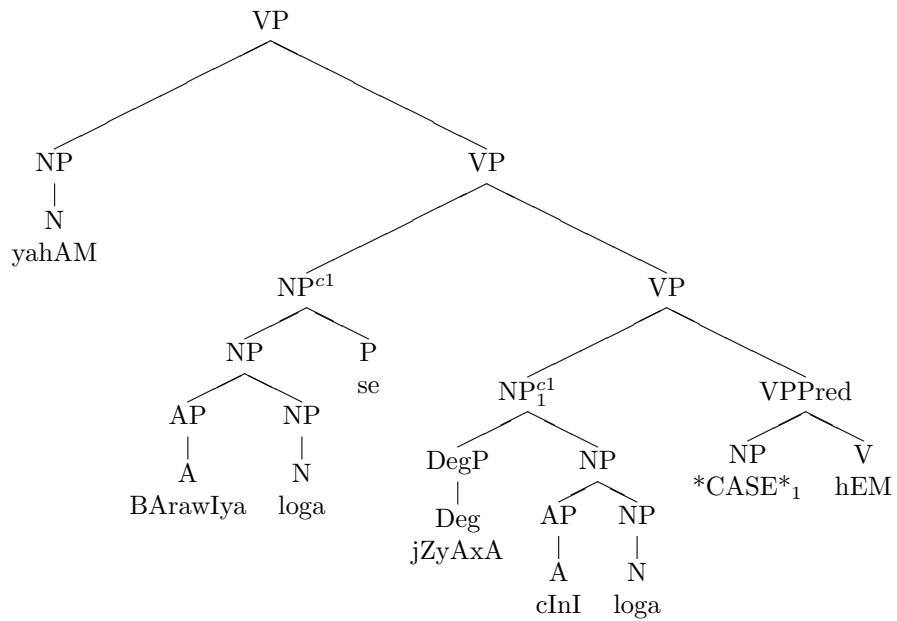
c. (Label:Structure-of-AdjP-PS-73)

DegP Predicate 3: DegP is on NP

yahã: bhaartiya log se zyAxaaa chiinii log hË
 yahAM BARawIya loga se jZyAxA cInI loga hEM
 here Indian people than more Chinese people be.Prs.Pl

‘There are more Chinese people here than Indian people.’

Tree for (139c) (Structure-of-AdjP-PS-73):



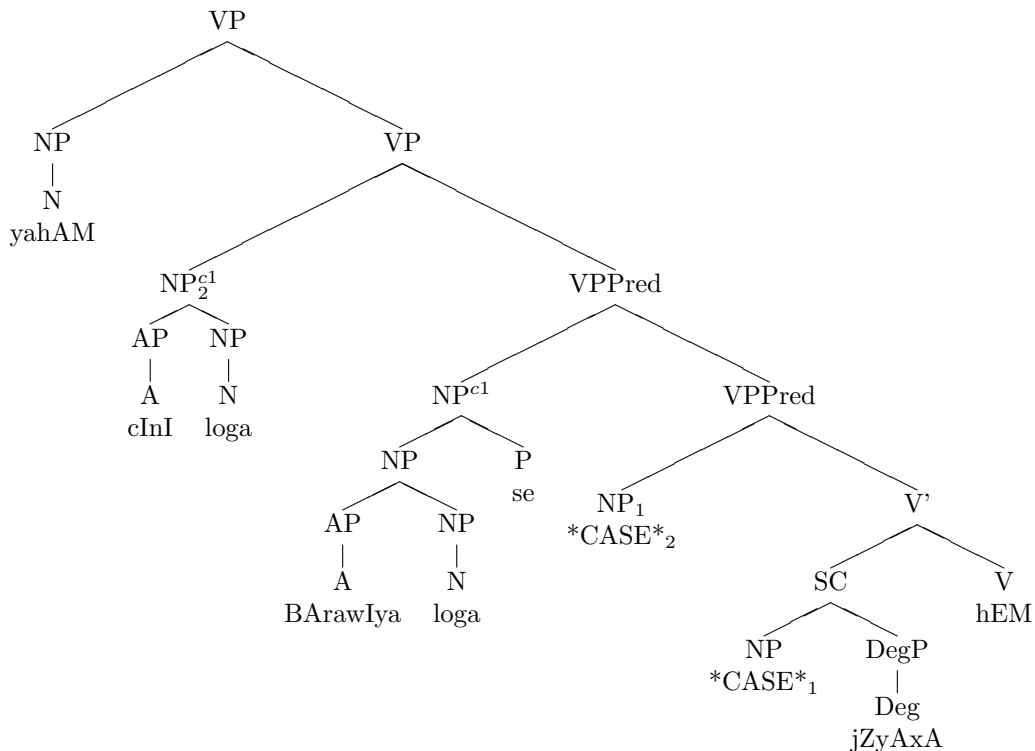
d. (Label:Structure-of-AdjP-PS-74)

DegP Predicate 4: DegP is discontinuous from NP

yahã: chiinii log bhaartiya log se zyAxaaa hË
 yahAM cInI loga BArawIya loga se jZyAxA hEM
 here Chinese people Indian people than more be.Prs.Pl

‘There are more Chinese people here than Indian people.’

Tree for (139d) (Structure-of-AdjP-PS-74):



Like in English, DegP can also attach to categories other than AP and NP such as VP.¹

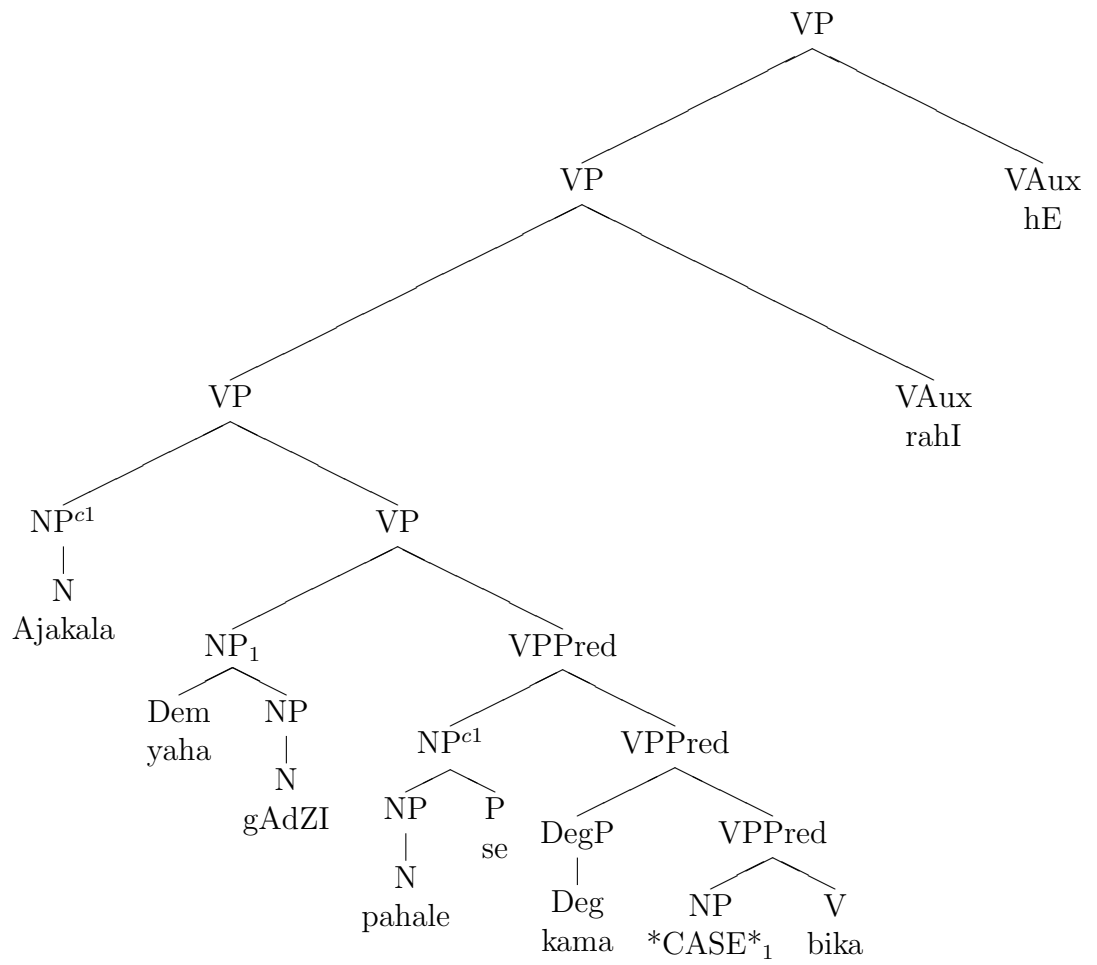
(140) a. (Label:DegP-on-VP-PS-1)

aaikal yeh gaariii pehle se kam bik rahii hai
 Ajakala yaha gAdZI pahale se kama bika rahI hE
 these.days this car before than less sell Prog.F be.Prs.Sg

‘These days this car is selling less than before.’

Tree for (140a) (DegP-on-VP-PS-1):

¹What is the category of elements like *aaikal* ‘these days’ (lit. today-yesterday) and *pehle* ‘before’? Maybe the fact that *se* can combine with it indicates that it is an NP.



Part VII

Ellipsis and Coordination

Chapter 27

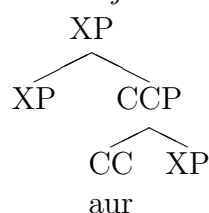
Coordination

****ocr Need to say that we still have a CONJ empty head!****

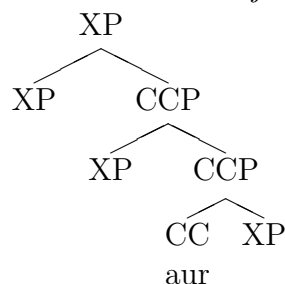
Phrases of most categories, with the major exception of CPs, can be coordinated. Orthographically, in Hindi, lists of three or more conjuncts do not require commas.

The basic syntax for coordinate phrases is illustrated below.

(141) a. *Two conjuncts:*



b. *More than two conjuncts:*

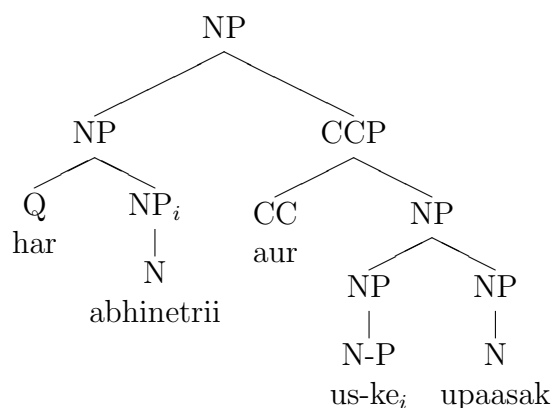


As illustrated by the diagrams above, our treatment of coordinate phrases utilizes a CCP (projected from the conjunction, CC). CCP indicates a constituent that can be adjoined to an XP, resulting in a structure that is interpreted as coordination.

Among the evidence for a hierarchical structure is the fact that the first phrase c-commands the second, as evidenced by the ability of the quantifier phrase in the first conjunct in (142) to bind the pronoun in the second conjunct.

(142) har abhinetrii_i aur us-ke_i upaasak
every actress and her-Gen.MPl devotee

‘Every actress and her devotees’



We note that a hierarchical structure for coordinations also accords with our general binary branching desideratum.

The coordinate phrase is always rightmost within the conjunction. One piece of evidence for this requirement is that the second phrase in a VP coordination, for example can stand on its own:

- (143) aur aap in prograamõ ko record bhii kar sakte hẽ
 and you.Ho these programs Acc record also do be.Prs.3Pl
 ‘And you can also record these programs.’

Further evidence that the coordinate phrase must always be rightmost comes from the behavior of coordinations of more than two conjuncts, to which we turn below.

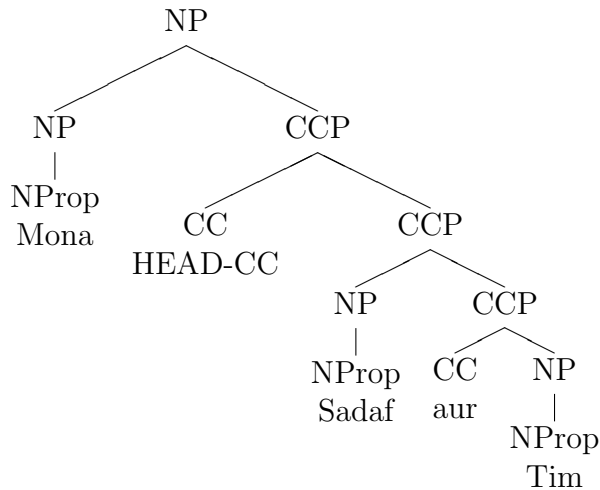
Coordinations in Hindi can consist of two XPs of the same category, or they can involve lists of three or more XPs.

- (144) a. Tim and Mona
 b. Tim and Mona and Atif

When more than two XPs are coordinated, an overt conjunct can occur following every (non-final) conjunct, or, much more likely, an overt conjunct occurs preceding the final conjunct in the list only:

- (145) a. Mona aur Sadaf aur Tim
 Mona and Sadaf and Tim
 b. Mona Sadaf aur Tim

These facts suggest that we invoke a covert coordinator, as shown in the possible (but not ultimately adopted) tree for (145b).



However, it is not clear how to constrain the distribution of such a covert coordinator in a principled fashion. Its distribution would have to be highly constrained in order to account for the generalization noted above—when only one overt conjunction occurs in a coordinate phrase, it can only occur preceding the final conjunct. An account which assumes a silent version of the overt coordinator is thus too permissive; it fails to rule out examples like (146).

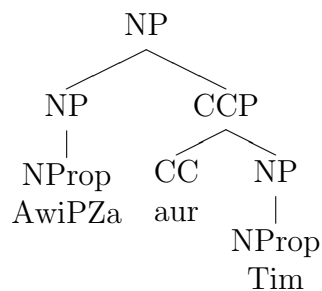
- (146) *Mona aur Sadaf Tim
 Mona and Sadaf Tim
 ‘Mona, Sadaf, and Tim’

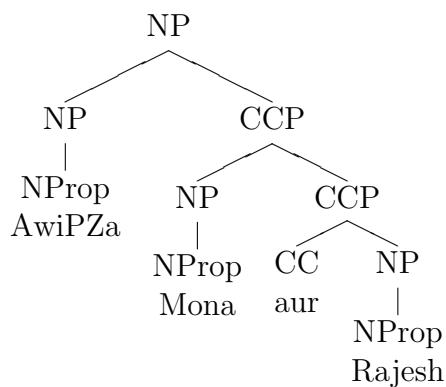
Instead of introducing a covert coordinator, we introduce a CCP projected from conjunction.

An XP and an overt coordinator (CC) combine to form a CCP. An XP and a CCP can combine to form either a CCP or an XP. If an XP is created, all conjuncts have been included, while if a CCP is created it must be adjoined to a further conjunct.

We give two examples, one with two conjuncts and one with three.

- (147) When two phrases are coordinated:





The only exception to this convention involves ****ocr What?****.

Note that if a comma is present in a conjunction, we treat it as a CC and as the head of a CCP. Commas that occur between conjuncts in coordinate phrases are one exception to our assumption that punctuation is not syntactically meaningful.

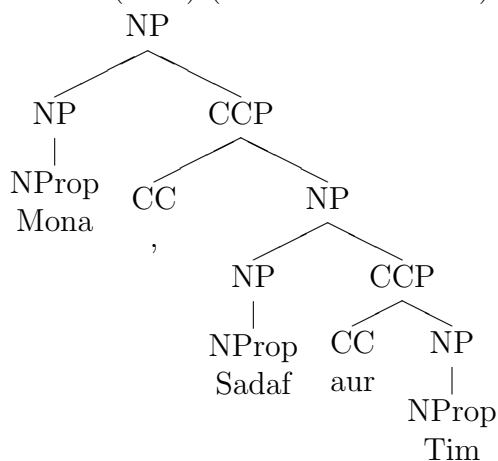
(148) a. (sent-coordinations-PS-6)

Atif , Mona aur Tim

Atif , Mona and Tim

‘Atif, Mona, and Tim’

b. Tree for (148a) (coordinations-PS-6)



Chapter 28

Shared arguments

28.1 Description and data

In this section we describe and present our analysis of a null argument phenomenon specific to coordinations, which we call Shared Argument constructions.

In VP coordinations, phrasal arguments of any type (subjects, direct objects, dative objects) can be non-overt in the second conjunct when they are referential with a corresponding argument in the preceding conjunct. The examples in (149) illustrate: (??), (??), and (??) show examples of shared subjects, and (??) shows a shared direct object. Finally, the sentence in (??) illustrates the possibility of multiple shared arguments within a single coordination: in this example, both the subject and one of the objects are null in the second conjunct.

- (149) a. (Elided-arg-DS-1) Shared subjects; case marking on first conjunct and second conjunct subjects differs:
- Ram ne ek kitaab paRhii aur so gayaa
rAma ne eka kiwAba padZi Ora so gayA
Ram Erg a book read and sleep go-Past
'Ram read a book and slept.'
- b. (Shared-arg-PS-1) Shared subjects; identical case marking on first and second conjunct subjects:
- Ram ne ek kitaab paRhii aur ek seb khaayaa
rAma ne eka kiwAba padZi Ora eka seba KayA
Ram Erg a book read and a apple ate
'Ram read a book and ate an apple.'
- c. (Elided-arg-DS-2) Shared subjects; identical case marking on first and second conjunct subjects:
- mere brother Ram ne aur usake Shyam ne baahar khaanaa BI
mere BAI rAma ne Ora usake xoswa SyAma ne bAhara KAnA BI
my brother Ram Erg and his friend Shyam Erg outside food also

khaayaa aur BI dekh-aa
 KayA Ora pikcara BI xeKI
 ate and movie ALSO saw-Pfv

‘My brother Ram and his friend Shyam had a meal outside and also watched a movie.’

- d. (Shared-arg-PS-2) Shared direct object:

yeh kitaab Ram ne chunii aur Sita ne khariidii
 yaha kiwAba rAma ne Ora sIwA ne KArlxI
 this book Ram Erg chose and Sita Erg bought

‘Ram chose this book and Sita bought (it).’

- e. (Shared-arg-PS-3) Shared subject and shared dative object:

Ram ne Ramesh ko paise diye aur bazaar bhej diyaa
 rAma ne rameSa ko pEse xiye ora bazAra Beja xiyA
 Ram Erg Ravi Dat money gave and market sent give-Pfv

‘Ram gave money to Ramesh and sent him to the market.’

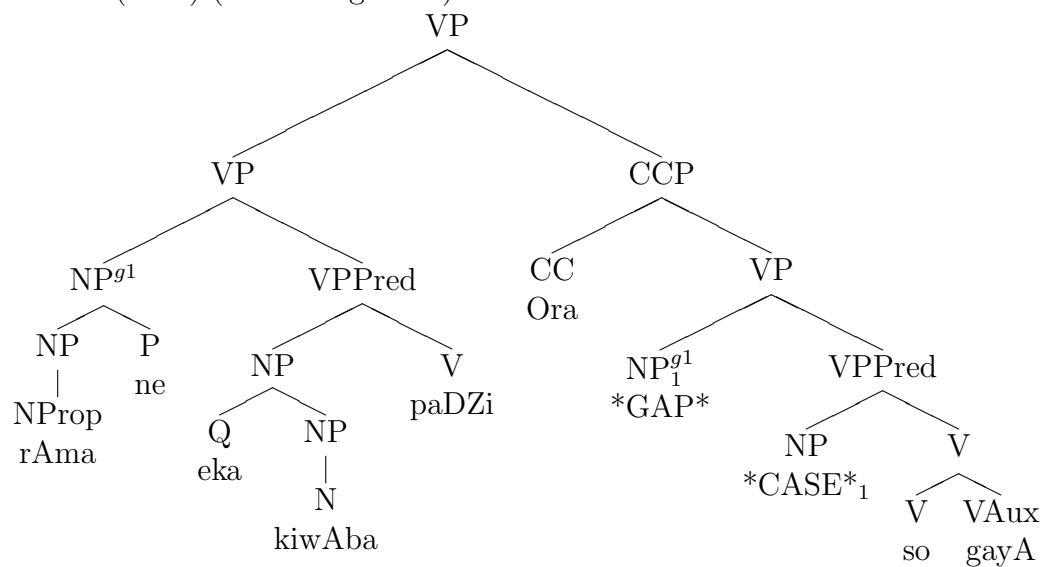
28.2 Analysis

The PS analysis of these constructions follows PB’s treatment, which introduces a GAP—a null element that is coreferential with the corresponding argument in the preceding conjunct—in the position of the null argument in the non-initial conjunct.¹ Accordingly, the PS analysis involves an ordinary VP coordination structure with empty categories (ECs) in the relevant argument positions. These ECs have the ec-type “GAP.” Coreference between the null argument and its correspondent in the preceding conjunct is indicated with a gap reference (“gapref”) feature (indicated in the trees below with a superscript).

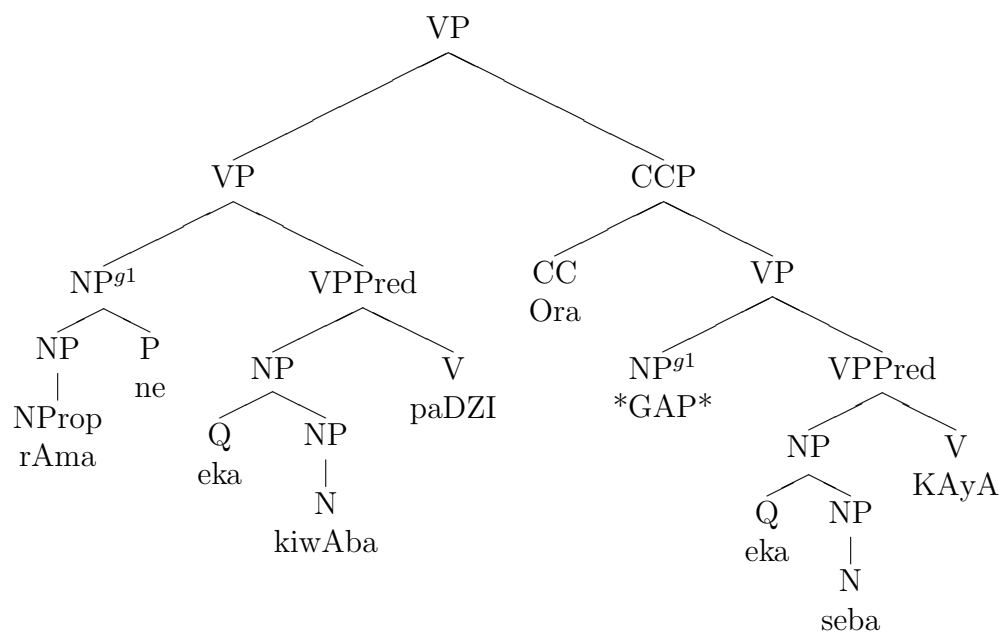
¹While some cases of shared arguments can be handled by Across-the-Board (ATB)-movement, not all can. Cases such as (??), for example, should not be handled by ATB movement, since the case marking on the two subjects differs—ergative in the first conjunct, but unaccusative in the second conjunct VP. On the other hand, given the independent existence of *pro* subject and object arguments in Hindi, the use of the null element GAP, as opposed to *pro*, in sentences like those in (149) might be questioned.

While this approach is intuitive particularly in cases where the subjects do not match in case marking and as such may not have the same syntax within their own conjunct, cases where the subjects are formally identical suggest treatment in terms of ATB-movement, or gapping, assuming that “gapped” subjects must be formally identical, as noted above. Accordingly, an alternative approach would distinguish between shared subject coordination constructions in which the subjects are identical up to case marking, and those in which they are not, by using GAP in the former cases, and *pro* in the latter. Alternatively, if *pro* is used for cases where the null argument and its correspondence are not identical, an ATB-analysis could be used for the remaining cases. While we consider this to be an equally plausible approach, as noted above, from the PS perspective, it is not as desirable for conversion, since DS and PB do not make a distinction between null arguments in coordinations based on (lack of) case matching, and because a true ATB analysis would cause additional complications for conversion. We therefore follow PB in distinguishing all null subjects of non-initial conjuncts in coordinations from all other null subjects. Null subjects inside coordinations are treated as GAP; null subjects elsewhere can be various types of ECs, including *pro* (in finite contexts) and PRO (in nonfinite contexts).

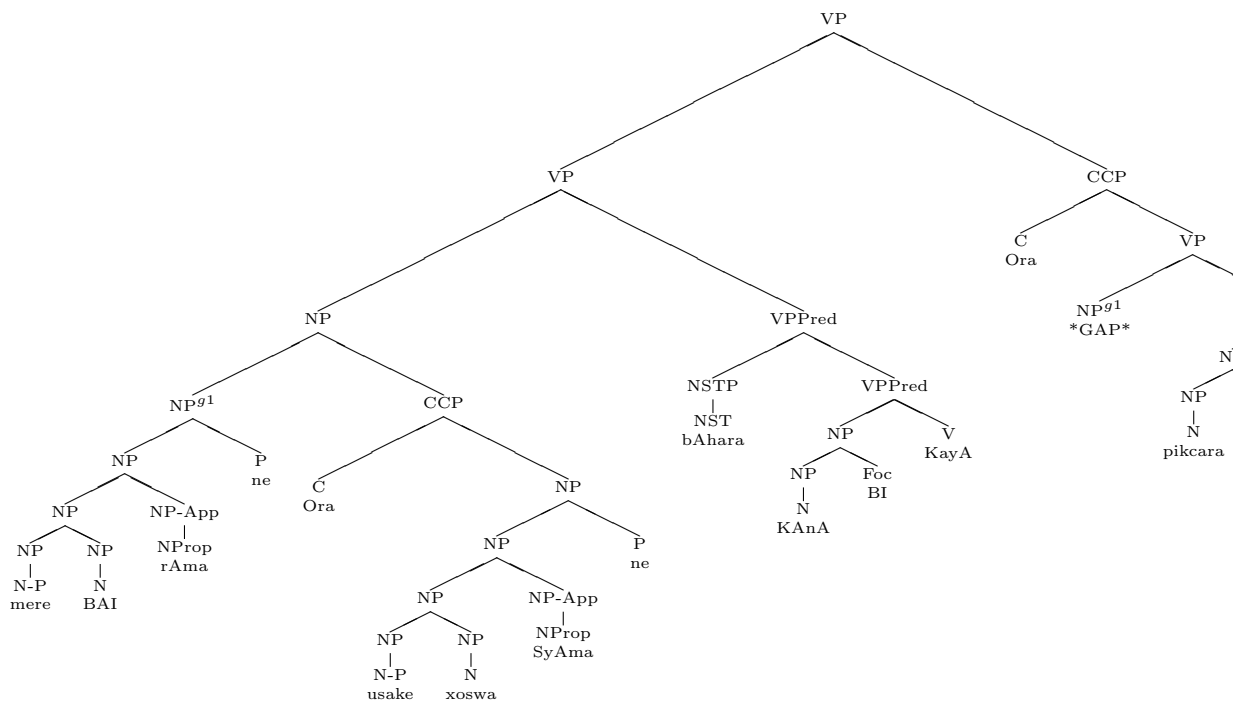
(150) a. Tree for (149a) (Elided-arg-DS-1)



b. Tree for (149b) (Shared-arg-PS-1)

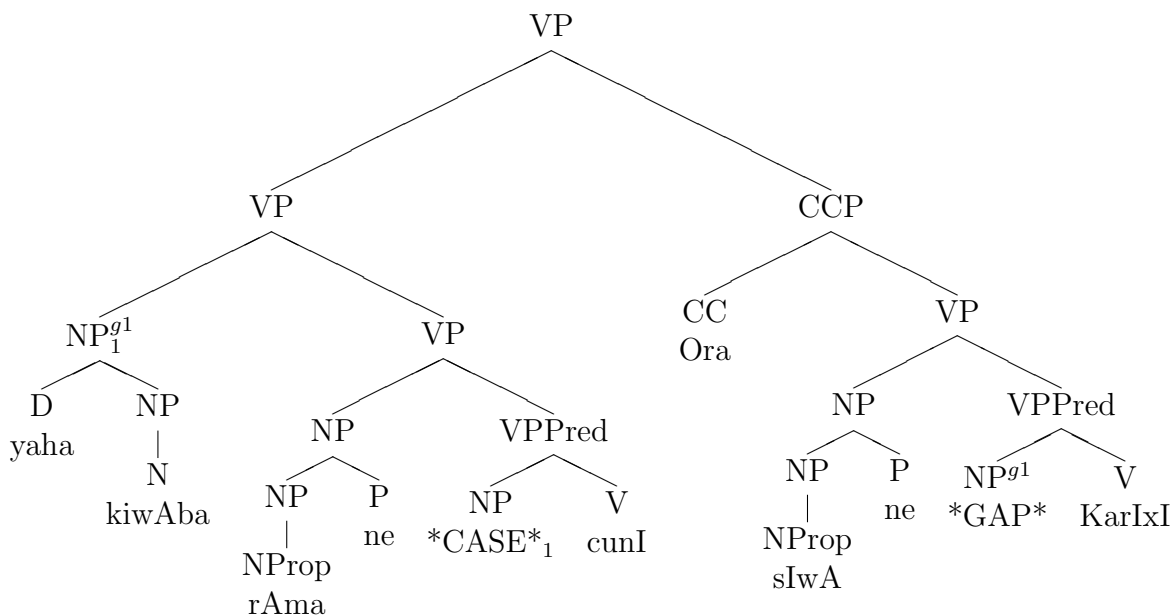


c. Tree for (149c) (Elided-arg-DS-2)

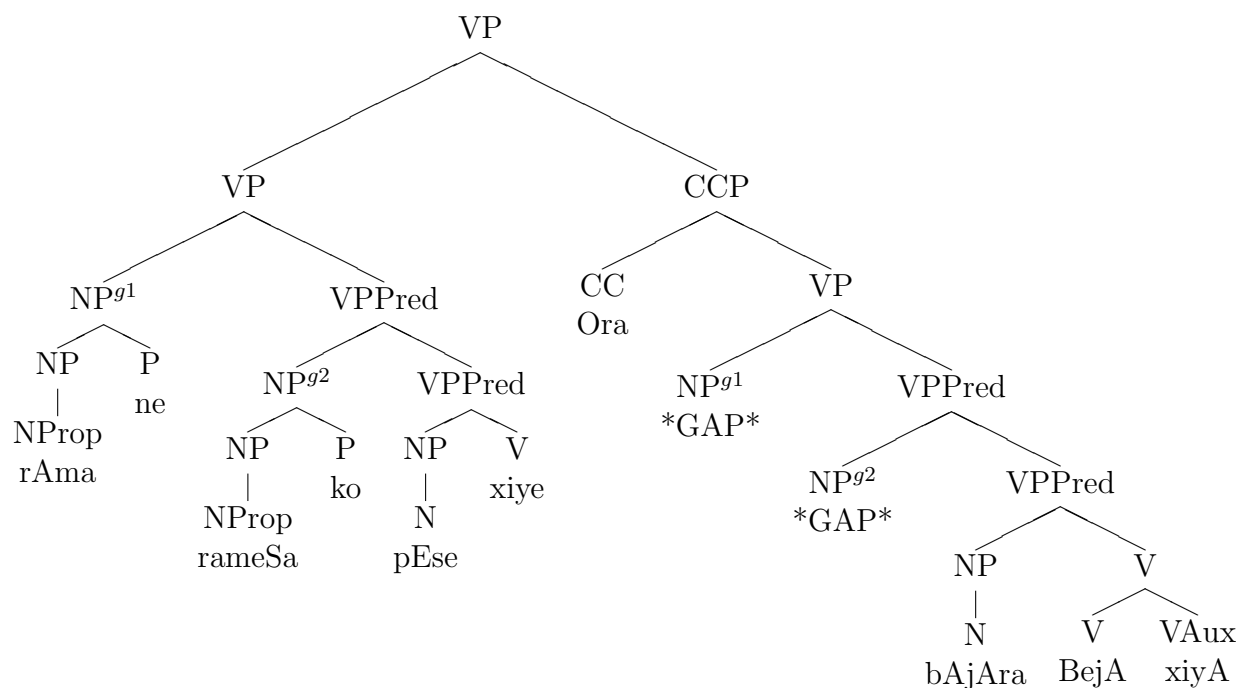


(151) Shared objects in coordinations

a. Tree for (149d) (Shared-arg-PS-2)



b. Tree for (149e) (Shared-arg-PS-3)



Note that we assume the minimum amount of movement necessary to achieve the correct surface order in the second conjunct. Thus we do not assume full parallelism between the two conjuncts. This is illustrated in (149d): while the object scrambles to the left edge of the VP in the first conjunct, as evidenced by its surface order with respect to the subject, the GAP corresponding to it in the second conjunct does not undergo a parallel movement.

Chapter 29

Ellipsis

29.1 NP-Ellipsis

NPs can rather freely go missing in Hindi-Urdu - the argument drop property of Hindi-Urdu - and this extends to parts of NPs.

- (152) a. (Ellipsis-Empty-Pronouns-PS-1) Genitive possessor remains:
Atif-ne Mona-kii kitaab paṛh-ii aur mĒ-ne **Tina-kii**.
AwiPZa ne monA kĪ kiwAba padZI Ora mone ne
Atif-Erg Mona-Gen.f book.f read-Pfv.f and I-Erg Tina-Gen.f

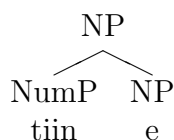
wĪna kĪ

'Atif read Mona's book and Mona read Tina's.'
- b. (Ellipsis-Empty-Pronouns-PS-2) Numeral Remains:
Atif-ko do kitaabē caahiye aur mujhe **tiin**
AwiPZa ko xo kiwAba cAhie Ora muJe wĪna
Atif-Dat two books wants and I.Dat three

'Atif wants two books and I want three.'
- c. (Ellipsis-Empty-Pronouns-PS-3) Adjective Remains:
Atif-ko kaalaa pen caahiye aur mujhe **haraa**
AwiPZa ko kalaA pena cAhie Ora muJe harA
Atif-Dat black pen wants and I.Dat green

'Atif wants (a/the) black pen and I want (a/the) green.'

Note that in these cases, the numeral/adjective is not morphologically nominalized - there is no additional morphology on these elements. A plausible way of representing these cases is as involving an AP/NumP etc. adjoined to an empty NP.



It remains to be decided whether the antecedent of the ellipsis will be disambiguated in the annotation. Given that this relationship is like pronominal anaphora resolution in many ways, it would seem reasonable to not attempt this in the annotation. ****ocr***af: 9-Feb-2011* I think we have now finalized this decision. We will distinguish two classes of ellipses: one, including NP-Ellipsis, Null Complement Anaphora, and Sluicing as well as argument drop, which we consider to be contextually recoverable and as such do not coindex, and another, including Stripping and Gapping, which we consider to be structurally conditioned, and do use (a particular type of) coindexation.*****

Cases of NP ellipsis may be flanked by a postposition on the right.

(153) (Ellipsis-Empty-Pronouns-PS-4)

Atif-ne lambe laṛke-ko ḍā:ṭ-aa aur mĒ-ne **na:ṭe-ko**
 Atif-Erg tall.Obl boy-Acc scold-Pfv and I-Erg short.Obl-Acc

‘Atif scolded the tall boy and I the short (one).’

29.2 Gapping

29.3 Gapping

Gapping is one of the major ellipsis processes operative in Hindi-Urdu. It elides the verb together with any tense/aspect morphology (obligatorily) and possibly associated XPs (optionally) in a non-initial conjunct of a coordination, under identity with structurally parallel corresponding elements in the preceding conjunct.

As such, gapping superficially resembles the Shared Argument construction discussed in Chapter , with the defining difference that the main, finite verb must be among the null elements. However, arguments and adjuncts can be, and often are, also missing along with the verb. (??) shows a basic example where just the verb is missing in the second conjunct, while (??)-(??) show examples of other argument and adjunct phrases in the verb phrase being gapped along with the verb. Finally, the example in (??), which involves a gapped infinitival complement, shows that gaps can be quite large, and can include more than one verb. (??) shows that a finite clause embedding verb can also be gapped.

(154) a. (Gapping-PS-1)

Verb only gap:

Ram ne seb khaayaa aur Sita ne santraa
 rAma ne seba KayA Ora sIwA ne saMwarA
 Ram Erg apple ate and Sita Erg orange

‘Ram ate an apple and Sita an orange.’

b. (Gapping-PS-2)

Verb plus direct object gap:

Ram ne Sita ko ek kitaab dii aur Mina ne Ravi ko
 rAma ne sIwA ko eka kiwAba xI Ora mInA ne Ravi ko
 Ram Erg Sita Dat a book gave and Mina Erg Ravi Dat

‘Ram gave a book to Sita and Mina to Ravi.’

c. (Gapping-PS-3)

Verb plus subject gap:

Ram ne Sita ko ek kitaab dii aur Mina ko akhbaar
 rAma ne sIwA ko eka kiwAba xI Ora Mina ko aKabAra
 Ram Erg Sita Dat a book gave and Mina Dat newspaper

‘Ram gave a book to Sita and a newspaper to Mina.’

d. (Gapping-PS-4)

Verb, subject, and adverb gap:

Ram ne Sita ko kal ek kitaab dii aur aaj ek akhbaar
 rAma ne sIwA ko kala ek kiwAba xI Ora Aja eka aKabAra
 Ram Erg Sita Dat yesterday a book gave and today a newspaper

‘Ram gave Sita a book yesterday and a newspaper today.’

e. (Gapping-PS-5)

Verb plus infinitival verb complement gap

Ram amruud khaanaa chaahtaa hai aur Sita seb
 rAma AmrUd khAnA chAtA Ora sIwA seba
 Ram guava eat.Inf want.Hab is and Sita apple

‘Ram wants to eat guavas and Sita apples.’

f. (Gapping-PS-6)

Atif ne Mona se kah-aa [ki vo der se aa egaa] aur
 AwiPZa ne monA se kahA kI vo ha xera se AegA Ora
 Atif Erg Mona with say-Pfv that he delay with come-Fut and Ravi
 Ravi ne [ki vo time pe aa egaa]
 ravi ne kI vo ha tAima pe AegA
 Erg that he time on come-Fut

‘Atif told Mona [that he would come late] and Ravi [that he would come on time].’

We adopt the following terminology: the unpronounced elements in the second conjunct, which crucially include the verb, comprise the **gap**, while the remaining, pronounced, elements, are the **remnants**.

The examples above illustrate the following additional defining characteristics of gapping. First, the gapped sequence does not have to be contiguous in the antecedent clause (see 154f). Second, as (154f) shows, the gapped sequence does not have to be right-peripheral. Finally, agreement mismatches do not lead to deviance in Gapping (see ?? where the missing verb would have had masculine agreement). Tense and aspectual mismatches, in contrast, are not permitted; accordingly, tense and aspectual auxiliaries are obligatorily included in the gap, as in (155a). (Other types of auxiliary verbs, as well as the light verbs of complex predicates (see other chapters) are also obligatorily included in the gap; they are not possible remnants, even if contrastive.) Negation is also always shared and obligatorily takes scope within each conjunct (155b).

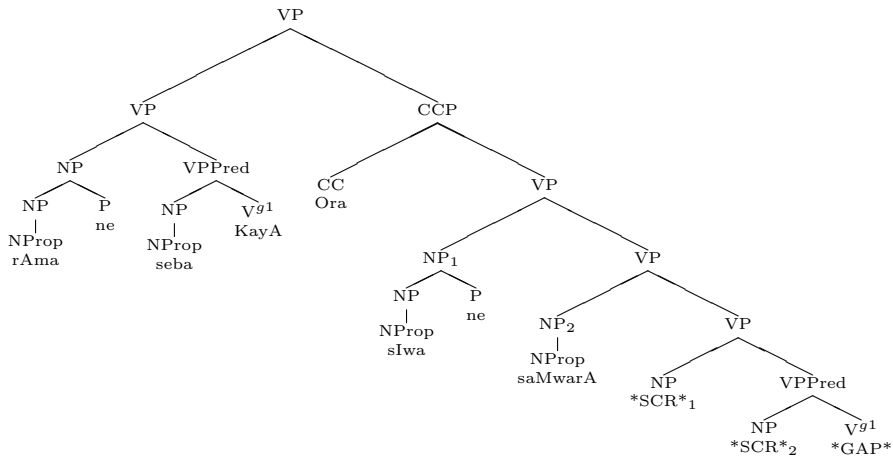
- (155) a. *Auxiliary obligatorily gapped with main verb*
 Ram kek khaa legaa aur Mina miThaai
 rAma keka KA legA Ora mInA miThAI
 Ram cake eat take.Fut and Mina sweets
 ‘Ram will eat the cake and Mina the sweets.’
- b. Ram ne kek nahiiN khaayaa aur Sita ne miThaai
 rAma ne keka nahIM KAyA Ora sIwA ne miThAI
 Ram Erg cake Neg eat.Pfv and Sita Erg sweets
 ‘Ram didn’t eat the cake and Sita the sweets.’ (cannot just mean: ‘Ram didn’t eat the cake and Sita eat the sweets.’ i.e. the following reading is not available
 $\neg(\text{eat}(R, c) \wedge \text{see}(S, s))$)

Our analysis of verbal gapping starts from the assumption that the verb phrase is the target of the ellipsis; however, for perspicuity and also in order to facilitate conversion, we introduce a separate null element for each phrase within the gapped VP.¹ The missing verb and any gapped arguments or adjuncts are present in the structures as null elements of the type GAP (the same type of null elements used in shared argument constructions.) As in the analysis of shared arguments, coreference between each gapped element is indicated by a “gapref” feature whose value is identical to a corresponding “gapref” feature on the gapped element’s correlate in the preceding conjunct; these are indicated with superscripts preceded by “g” (to distinguish them from movement indices) in the trees below.

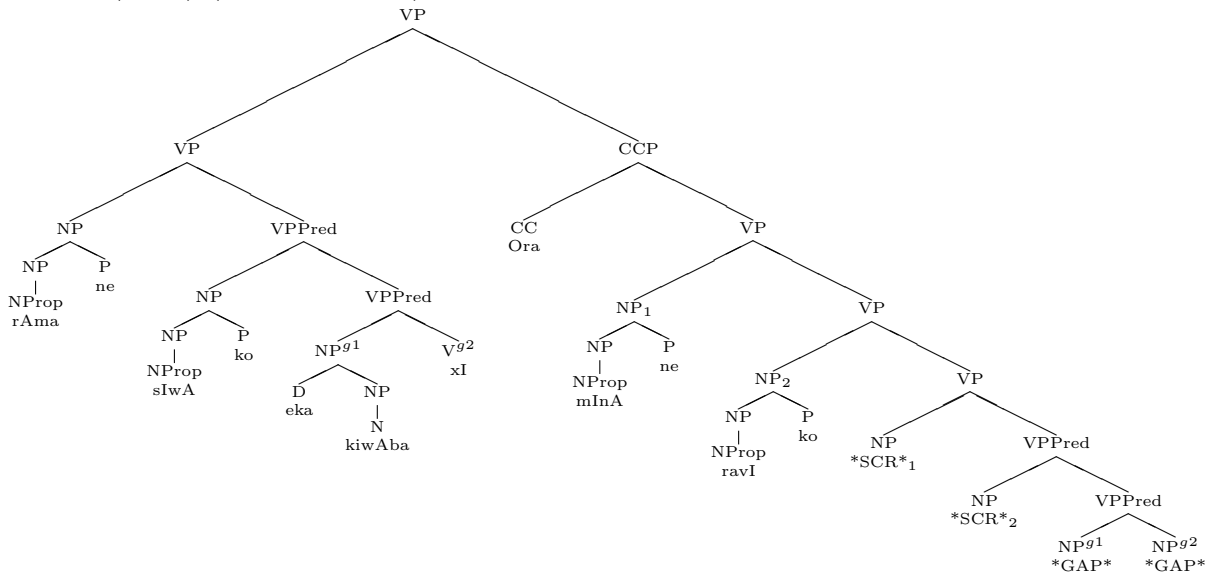
The remnants, in contrast, are analyzed as having moved out of the targeted VP, in accordance with the assumption that the VP itself is the constituent targeted by gapping. Thus, they appear as ordinary traces of scrambling (ec type=SCR) within the VP, and are indexed, via the indices used elsewhere in movement, with their overt counterparts adjoined to the VP.

- (156) a. Tree for (154a) (Gapping-PS-1)

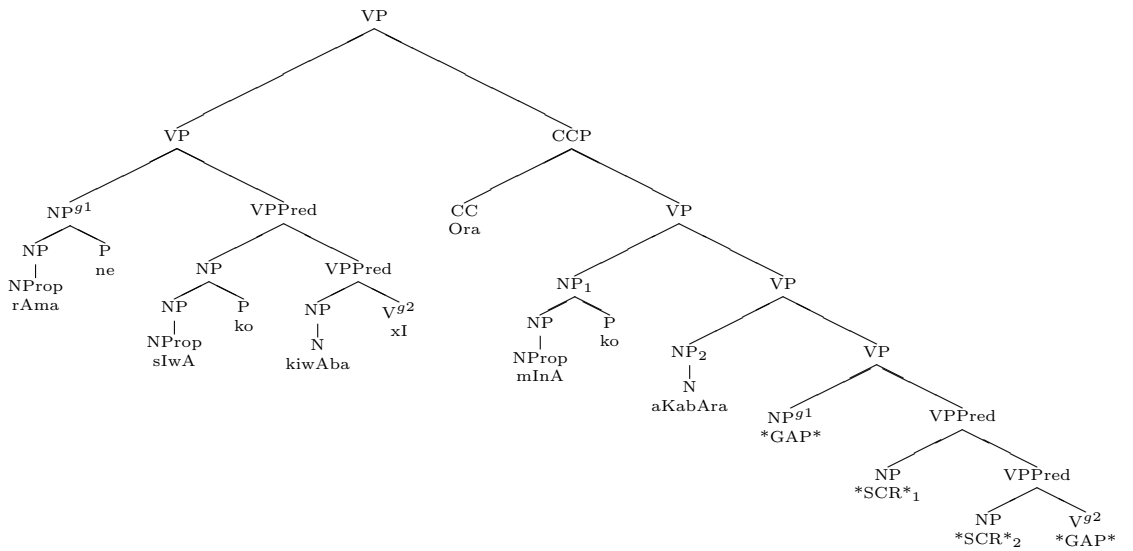
¹An alternative would be to introduce a null VP, thus eliminating separate V and NP GAPS within the VP. While this approach is consistent with the generative assumption that the verb phrase is the target of ellipsis, it does not allow us to represent the internal structure of the null VP. Also, it is not as desirable from the perspective of conversion, since the DS and PB levels represent each null item within the gapped VP with individual GAPS.



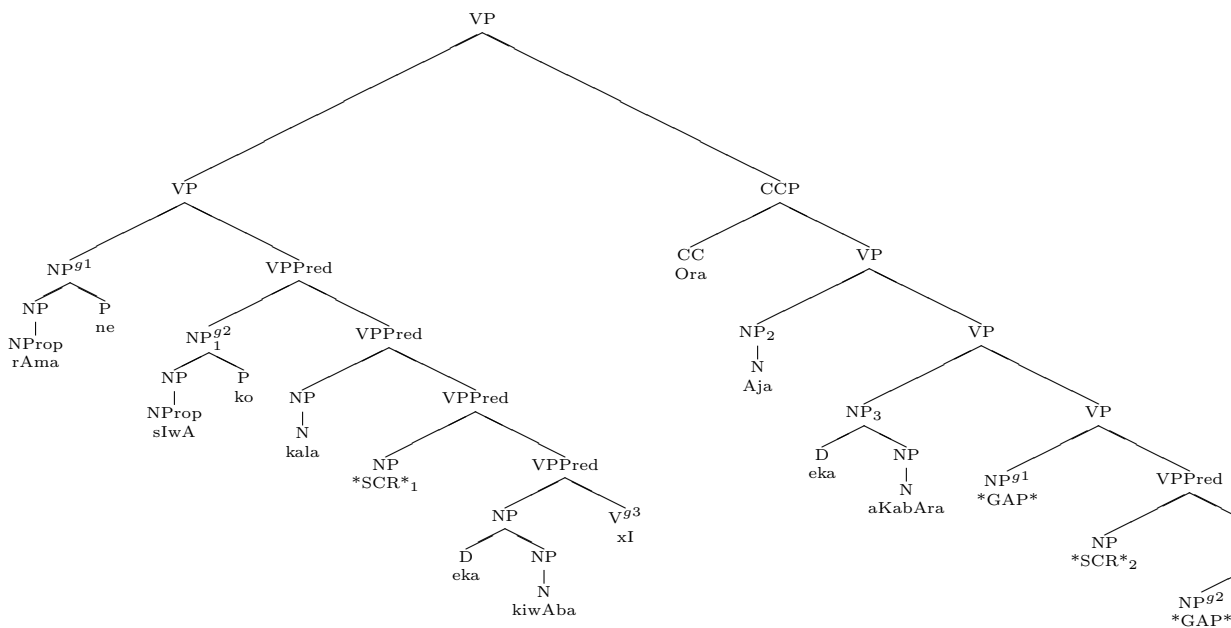
b. Tree for (154b) (Gapping-PS-2)



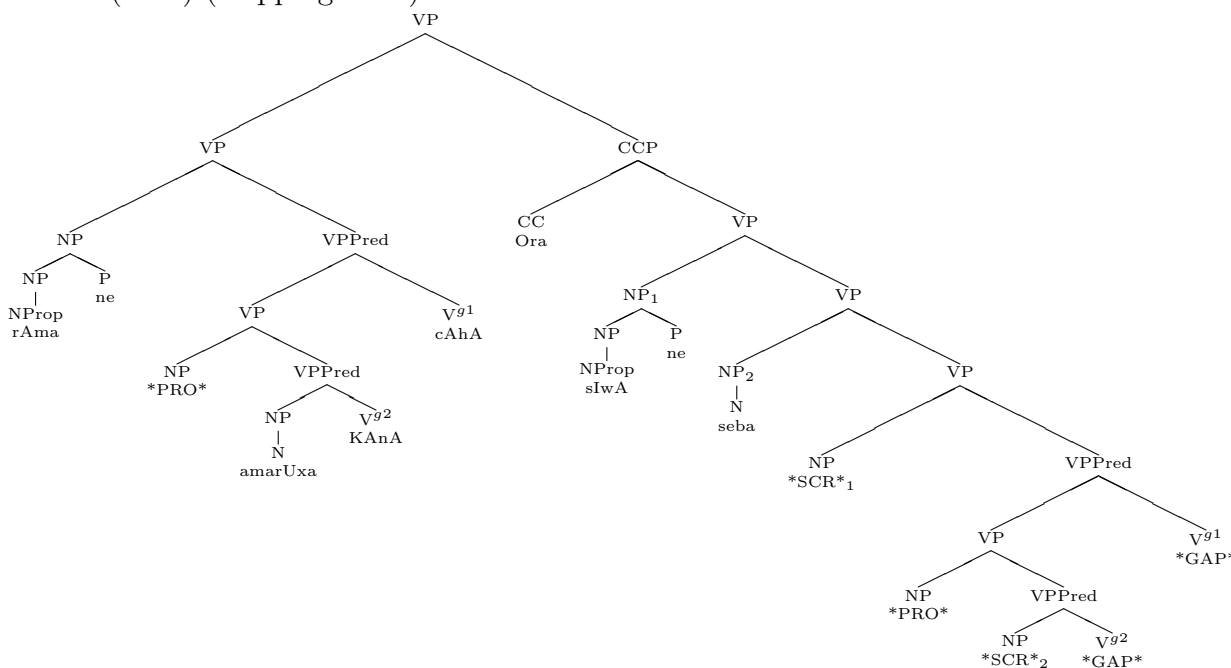
c. Tree for (154c) (Gapping-PS-3)



d. Tree for (154d) (Gapping-PS-4)



e. Tree for (154e) (Gapping-PS-5)

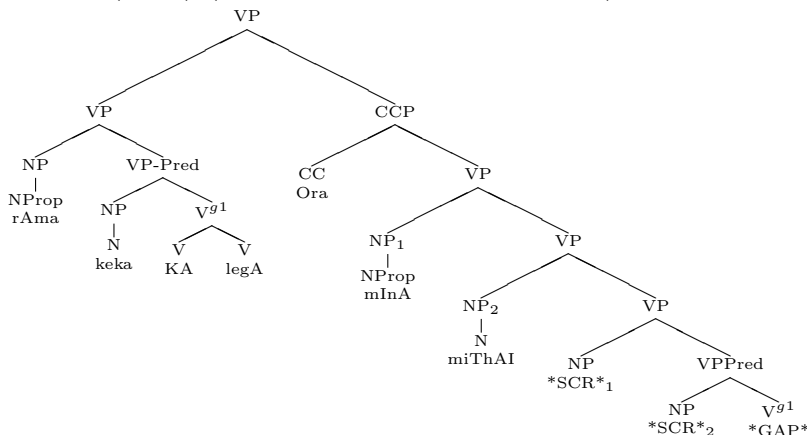


Note that we assume movement of all remnants, whether they are arguments of the verb or not. Thus, movement of the remnant extends to cases involving adjunct remnants as well, such as (??).²

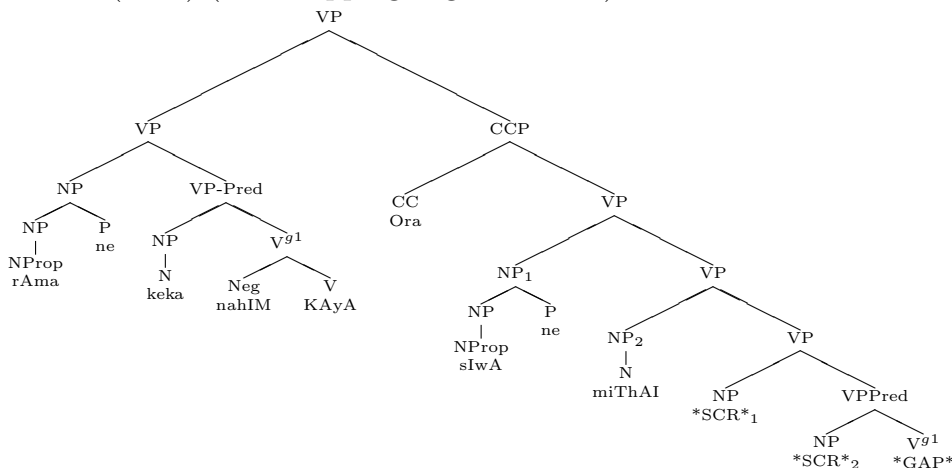
²It can reasonably be objected that it is more difficult to be certain in the case of non-argument remnants whether they have necessarily moved from within the verb phrase. This suggests an alternative possible analysis, according to which adjunct remnants do not undergo movement but are simply base adjoined in their surface positions. While this would make the overall analysis of gapping non-uniform, we consider it to be equally feasible and could also adopt this treatment if it proves to be more conducive to conversion. A relevant observations is that while all arguments that do not have a contrastive counterpart in the second

The PS structures for examples such as (155a) and (155b), which involve adjunction to V, involves coreference between the complex V (V plus V and V + Neg, respectively), represented as a single null V, and its overt counterpart in the first conjunct.³

(157) a. Tree for (155a) (sent-Gapping-auxiliary-PS-7)



b. Tree for (155b) (sent-Gapping-negation-PS-8)



Gapping is one of the major ellipsis processes operative in Hindi-Urdu. It elides the verb together with any tense/aspect morphology (obligatorily) and possibly associated XPs (optionally) in the second conjunct. Again, I hesitated here; ‘ellipsis processes’ and ‘elides’ seem very generative specific.

(158) a. (Ellipsis-in-Coordination-PS-1) verb missing:

Atif-ne Mona-ko pencil dii aur Ravi-ne Tina-ko pen
 AwiPZa ne monA ko peMsila xI Ora ravi ne
 Atif-Erg Mona-Dat pencil.f give.Pfv.f and Ravi-Erg Tina-Dat pen.m

clause are shared from the first clause, adjuncts don't have to be. The default reading involves sharing of the adjunct but another reading where the adjunct is not shared also exists.

³We note that this introduces a possible asymmetry with the DS analysis, which may involve sharing of the main verb alone.

tInA ko pena

‘Atif gave a pencil to Mona and Ravi a pen to Tina.’

- b. (Ellipsis-in-Coordination-PS-2) verb and direct object missing:

Atif-ne Mona-ko pencil dii aur Ravi-ne Tina-ko
 AwiPZa ne monA ko peMsila xI Ora ravi ne tInA
 Atif-Erg Mona-Dat pencil.f give.Pfv.f and Ravi-Erg Tina-Dat

ko

‘Atif gave a pencil to Mona and Ravi to Tina.’

- c. (Ellipsis-in-Coordination-PS-3) verb and indirect object missing:

Atif-ne Mona-ko pencil dii aur Ravi-ne pen
 AwiPZa ne monA ko peMsila xI Ora ravi ne pena
 Atif-Erg Mona-Dat pencil.f give.Pfv.f and Ravi-Erg pen.m

‘Atif gave a pencil to Mona and Ravi a pen.’

- d. (Ellipsis-in-Coordination-PS-4) verb and inner argument missing, arguments left on both sides:

Atif-ne Mona-se kah-aa [ki vo der-se aa-egaa] aur Ravi-ne
 AwiPZa ne monA se kahA kI voha xera se
 Atif-Erg Mona-with say-Pfv that he delay-with come-Fut and Ravi-Erg
 [ki vo time-pe aa-egaa]
 AegA Ora ravi ne kI voha tAima pe AegA
 that he time-on come-Fut

‘Atif told Mona [that he would come late] and Ravi [that he would come on time].’

- agreement mismatches do not lead to deviance in Gapping (see 158a where the missing verb would have had masculine agreement).

- the gapped sequence does not have to be contiguous in the antecedent clause (see 158c).

- the gapped sequence does not have to be right-peripheral. (see 158d).

The coordination in Gapping can also be ‘lower’:

- (159) a. (Ellipsis-in-Coordination-PS-5) verb *dii* is shared on the right and the subject is shared on the left:

Atif-ne [Mona-ko ek kitaab] dii aur [Ravi-ko ek tasviir]
 AwiPZa ne monA ko eka kiwAba xI Ora ravi ko
 Atif-Erg Mona-Dat a book.f give.Pfv.f and Ravi-Dat a picture.f

eka wasvIra

‘Atif gave a book to Mona and Tina a picture to Ravi.’

- b. (Ellipsis-in-Coordination-PS-6) direct object + verb *ek tasviir dii* is shared on the right and the subject on the left:

Atif-ne [Mona-ko aaj] ek tasviir dii aur [Ravi-ko kal]
 AwiPZa ne Aja monA ko eka wasvIra xI Ora
 Atif-Erg Mona-Dat today a picture give.Pfv.f and Ravi-Dat yesterday

ravi ko kala

‘Atif gave a picture to Mona today and to Ravi yesterday.’

- possible that the shared material on the left gets there by ATB-movement.

While all arguments that do not have a contrastive counterpart in the second clause are shared from the first clause, adjuncts don’t have to be. The default reading involves sharing of the adjunct but another reading where the adjunct is not shared also exists. Negation, however, is always shared and obligatorily takes scope within each conjunct.

- (160) a. (Ellipsis-in-Coordination-PS-7) adverbs:

Atif-ne ek ghanṭe-me ret-kaa kilaa banaa-yaa thaa aur Shahjahan-ne
 AwiPZa ne eka GaMte meM rewa kA kilA banAyA
 Atif-Erg one hour-in sand-Gen castle make-Pfv be.Pst and Shahjahan-Erg
 Taj Mahal
 WA Ora SAhajahAz ne wAja mahala
 Taj Mahal

‘Atif had made a sandcastle in an hour and Shahjahan the Taj Mahal.’

Reading 1: ‘S. made the Taj Mahal.’

Reading 2: ‘S. made the Taj Mahal in an hour.’

- b. (Ellipsis-in-Coordination-PS-8) negation:

Atif-ne JJWS nahī: dekh-ii aur Mona-ne KANK
 AwiPZa ne JJWS nahIM xeKi Ora monA ne KANK
 Atif-Erg JJWS neg see-Pfv.f and Mona-Erg KANK

‘Atif didn’t see JJWS and Mona didn’t see KANK.’

(cannot just mean: ‘Atif didn’t see JJWS and Mona see KANK.’ i.e. the following reading is not available $\neg(\text{see}(a, j) \wedge \text{see}(m, k))$)

- in the behavior of negation, Hindi-Urdu is different from English. German seems to allow both the Hindi-Urdu and the English patterns.

The plan is that ellipsis created by gapping will be fully disambiguated in the annotation. Much of this will be fully automatic. The only issue will be with respect to whether adjuncts should be copied or not. This will be left to the judgement of the annotator.

Constraints on Gapping:

(161) a gap cannot cross a finite clause boundary:

a. (Ellipsis-in-Coordination-PS-9)

Atif-ne Sita-ko kitaab de-nii caah-ii aur Tim-ne Mona-ko
 AwiPZa ne sIwA ko kiwAba xenI cAhI Ora tIma ne
 Atif-Erg Sita-Dat book.f give-Inf.f want-Pfv.f and Tim-Erg Mona-Dat

monA ko

‘Atif wanted to give a book to Sita and Tim to Mona.

b. (Ellipsis-in-Coordination-PS-10)

Atif caah-taa hai ki vo res jiiit-e aur Tom
 AwiPZa cAhawA hE kI voha resa jIwe Ora tOYma
 Atif want-Hab be.Prs.Sg that he race win-Sbjv.MSg and Tom

haar-e

hAre

lose-Sbjv.MSg

Available Reading: Atif wants that he win the race and Tim lose the race.

Unavailable reading: Atif wants that he win the race and Tom *wants that he* lose *the race*.

c. (Ellipsis-in-Coordination-PS-11)

*Atif-ne caah-aa hai ki vo res jiiit-e aur Tom-ne
 AwiPZa ne cAhA kI voha resa jIwe Ora tOYma
 Atif want-Pfv be.Prs that he race win-Sbjv.MSg and Tom

haar-e

hAre

lose-Sbjv.MSg

Available Reading: Atif wants that he win the race and Tim lose the race.

Unavailable reading: Atif wants that he win the race and Tom *wants that he* lose *the race*.

d. (Ellipsis-in-Coordination-PS-12) Complex NP

*Atif-ko yeh baat [ki Sita lambii hai] pataa hai aur Tina-ko
 Atif-Dat this thing that Sita tall.f be.Prs.Sg and Tina-Dat this thing
 yeh baat [ki Mona].

that Mona

Intended: Atif knows the claim that Sita is tall and Tina *knows* the claim that Mona *is tall*.

(Boundedness)

(162) Gapping cannot skip the finite part of the verb being gapped. Also auxiliaries cannot be gapped by themselves.

a. (Ellipsis-in-Coordination-PS-13) not possible to gap verb leaving auxiliary behind:

*Atif kapṛe **dho** rahaa hai aur Ravi farsh rahaa hai
 AwiPZa kapade Xo rahA hE Ora ravi ParSa rahA hE
 Atif.m clothes wash Prog.MSg be.Prs.Sg and Ravi floor Prog.MSg be.Prs.Sg

‘Atif is washing clothes and Ravi is wiping the floor.’

b. (Ellipsis-in-Coordination-PS-14) progressive auxiliary cannot be gapped:

*Atif kapṛe dho **rahaa hai** aur Ravi farsh pōch
 AwiPZa kapade Go rahA hE Ora ravi ParSa poCa rahA hE
 Atif.m clothes wash Prog.MSg be.Prs.Sg and Ravi floor wipe

‘Atif is washing clothes and Ravi is wiping the floor.’

c. (Ellipsis-in-Coordination-PS-15) tensed auxiliary cannot be gapped:

*Atif kapṛe dho-**taa hai** aur Ravi farsh pōch-**taa**
 Atif.m clothes wash-Hab.MSg be.Prs.Sg and Ravi floor wipe-Hab.MSg

‘Atif washes clothes and Ravi wipes the floor.’

d. (Ellipsis-in-Coordination-PS-16) the embedding verb can be gapped (along with their auxiliaries)

(?)Atif Mona-ko kitaab de-naa **caah-*taa* hai** aur Ravi
 Atif.M Mona-Acc book.f give-Inf want-Hab.MSg be.Prs.Sg and Ravi.M
 Mona-se kitaab le-naa
 Mona-from book.f take-Inf

‘Atif wants to give Mona a book and Ravi wants to take a book from Mona.’

This means that Hindi-Urdu does not allow Subgapping. This also seems to be true of embedded clauses in German (Pgs. 109-11 of Lechner).

(163) a. (Ellipsis-in-Coordination-PS-18) Gapping is marginal out of embedded finite clause 1:

?mujhe lag-**taa hai** [ki Sita Mumbai jaa-egii] aur tumhē lag-**taa**
 me.Dat feel-Hab be.Prs that Sita.f Mumbai go-Fut.f and you.Dat seem-Hab
 hai [ki Mona]
 be.Prs that Mona.f

‘I think that Sita will go to Mumbai and you think that Mona.’

b. (Ellipsis-in-Coordination-PS-19) Gapping is marginal out of embedded finite clause 2:

?Atif-ne kah-aa ki Sitaa Mumbai jaa-egii aur Rohit-ne kah-aa ki Tina
 Atif-Erg say-Pfv that Sita Mumbai go-Fut.f and Rohit-Erg say-Pfv that Tina
 Dilli
 Delhi

‘Atif said that Sita will go to Mumbai and Rohit said that Tina Delhi.’

- c. (Ellipsis-in-Coordination-PS-20) Gapping is marginal out of embedded finite clause 3:

??mujhe nahī: lag-taa ki [ek-bhii laṛkii 601 le-gii] aur tumhē nahī:
 me.Dat Neg seem-Hab that one-ever girl.f 601 tage-Fut.f and you.Dat Neg
 lag-taa ki [ek-bhii laṛkaa 602]
 seem-Hab that one-ever boy 602

‘*I don’t think that any girl will take 601 and you don’t think that any boy 602.’

- d. (Ellipsis-in-Coordination-PS-21) Gapping is not possible out of relative clauses:

*Atif-ne paṛh-ii thii [vo kitaab [jo Ravi-ne likh-ii thii]] aur
 Atif-Erg read-Pfv.f be.Pst.f that book.f Rel Ravi-Erg write-Pfv.f be.Pst.f and
 Mona-ne paṛh-ii thii [vo kahaanii [jo Tina-ne]]
 Mona-Erg read-Pfv.f be.Pst.f that story.f Rel Tina-Erg

‘*Atif read the book which Ravi wrote and Mona read the story which Tina.’

Cases like (163a, b) are strongly out in English. What is going on in Hindi-Urdu? Are these cases genuinely gapping? Could they be sluicing?

- (164) More than one XP must be left behind by gapping:

- a. (Ellipsis-in-Coordination-PS-22) single remnant:

???Atif-ne Sita-ko dekh-aa aur Mona-ne
 Atif-Erg Sita-Acc see-Pfv and Mona-Erg

intended: ‘Atif saw Sita and Mona (saw Sita).’

(Even with a *-bhii* ‘also’ on the remnant, the example is degraded. In fact, the two clause counterpart where the *Mona-ne-bhii* is a separate utterance is much better.)

- b. (Ellipsis-in-Coordination-PS-23) multiple remnants:

Atif-ne Sita-ko aaj dekh-aa aur Mona-ne kal
 Atif-Erg Sita-Acc today see-Pfv and Mona-Erg yesterday

intended: ‘Atif saw Sita today and Mona (saw Sita) yesterday.’

29.4 Right Node Raising

Right Node Raising can lead to sharing of verbal sequences but actually the shared material need not be a verbal sequence at all - right peripherality is what seems to matter.

- (165) a. (Ellipsis-in-Coordination-PS-24) verb *dii* is shared:

- Atif-ne Mona-ko ek kitaab aur Tina-ne Ravi-ko ek tasviir dii
 Atif-Erg Mona-Dat a book.f and Tina-Erg Ravi-Dat a picture.f give.Pfv.f
 ‘Atif gave a book to Mona and Tina a picture to Ravi.’
- b. (Ellipsis-in-Coordination-PS-25) direct object + verb *ek tasviir dii* is shared:
 Atif-ne Mona-ko aur Tina-ne Ravi-ko ek tasviir dii
 Atif-Erg Mona-Dat and Tina-Erg Ravi-Dat a picture.f give.Pfv.f
 ‘Atif gave a picture to Mona and Tina (did) to Ravi.’
- c. (Ellipsis-in-Coordination-PS-26) CP is shared:
 Atif-ko lag-taa hai aur Mona-ko pataa hai [ki dharti gol
 Atif-Dat seem-Hab be.Prs and Mona-Dat known be.Prs that earth.f round
 hai]
 be.Prs
 ‘Atif thinks and Mona knows that the earth is round.’
- d. (Ellipsis-in-Coordination-PS-27) even auxiliary verbs (*rahaa hai*) can be (marginally) shared:
 (?)Atif kapṛe dho aur Ravi farsh pōch rahaa hai
 Atif.m clothes wash and Ravi floor wipe Prog.MSg be.Prs.Sg
 ‘Atif is washing clothes and Ravi is wiping the floor.’

The coordination in RNR can also be ‘lower’:

- (166) a. (Ellipsis-in-Coordination-PS-28) *dii* is shared on the right and the subject is shared on the left:
 Atif-ne [Mona-ko ek kitaab] aur [Ravi-ko ek tasviir] dii
 Atif-Erg Mona-Dat a book.f and Ravi-Dat a picture.f give.Pfv.f
 ‘Atif gave a book to Mona and Tina a picture to Ravi.’
- b. (Ellipsis-in-Coordination-PS-29) —Sdirect object + verb *ek tasviir dii* is shared:
 Atif-ne [Mona-ko aa]j aur [Ravi-ko kal] ek tasviir dii
 Atif-Erg Mona-Dat today and Ravi-Dat yesterday a picture.f give.Pfv.f
 ‘Atif gave a picture to Mona today and to Ravi yesterday.’
 - possible that the shared material on the left gets there by ATB-movement.
- (167) Negation
- a. (Ellipsis-in-Coordination-PS-30) Negation can be shared: seems to take scope under the conjunction
 Atif-ne ek-bhii kitaab aur Mona-ne ek-bhii ciṭṭhii nahī: paṛh-ii
 Atif-Erg one-even book.f and Mona-Erg one-even letter.f Neg read-Pfv.f
 ‘Atif didn’t read even one book and Mona didn’t read even one book.’
- b. (Ellipsis-in-Coordination-PS-31) Negation must be shared: must take scope under the conjunction inside each clause

Tina kuch kitaabẽ aur Mona ek-bhii ciṭṭhii nahī: paṛh-egii
 Tina.f some.Pl book.FPl and Mona.f one-even letter.f Neg read-Pfv.f

‘Tina won’t read some books and Mona won’t read any books.’

- c. (Ellipsis-in-Coordination-PS-32) Negation can’t be left in the first conjunct normally:

#Tina kitaab nahī: aur Mona ciṭṭhii paṛh-egii
 Tina.f book.f Neg and Mona.f letter.f read-Fut.f

intended: ‘Tina won’t read a book and Mona will read a letter.’

- d. (Ellipsis-in-Coordination-PS-33) Negation can marginally be left in the first conjunct if there is a negation in the second conjunct also:

(?)Tina kitaab nahī: aur Mona ciṭṭhii nahī: paṛh-egii
 Tina.f book.f Neg and Mona.f letter.f Neg read-Fut.f

intended: ‘Tina won’t read a book and Mona won’t read a letter.’

- (168) More than one XP must be left behind by RNR:

- a. (Ellipsis-in-Coordination-PS-34) RNR parse unavailable, as indicate by agreement

Atif aur Ravi Sita-se mil-ẽge/*mil-egaa
 Atif.m and Ravi.m Sita-with meet-Fut.MPl/*meet-Fut.MSg

‘Atif and Ravi will meet with Sita.’

- b. (Ellipsis-in-Coordination-PS-35)

Atif aaj aur Ravi kal Sita-se mil-egaa/*mil-ẽge
 Atif.m today and Ravi.m tomorrow Sita-with meet-Fut.MSg/*meet-Fut.MPl

‘Atif will meet with Sita today and Ravi tomorrow.’

- i.e. when a coordination parse is available, an RNR parse becomes unavailable. Annotators should only explore the RNR option when simple coordination of XPs is not possible.

Agreement: when the elided verb and the overt verb in the second conjunct have distinct agreement features, we get marginality. Agreement with an argument of the first conjunct is strictly ungrammatical but local agreement is not perfect either. Sometimes speakers will use a non-syntactically licensed plural agreement.

- (169) Gender Mismatch:

- a. (Ellipsis-in-Coordination-PS-36) M + F: ?

Ram Dilli aur Sita Agra ??jaa-egii/*jaa-egaa/???jaa-ẽge
 Ram.m Delhi and Sita.f Agra go-Fut.f/go-Fut.MSg/go-Fut.MPl

‘Ram will go to Delhi and Sita to Agra.’

- b. (Ellipsis-in-Coordination-PS-37)

Sita Dilli aur Ram Agra ?jaa-egaa/*jaa-egii/???jaa-ẽge
 Ram.m Delhi and Sita.f Agra go-Fut.f/go-Fut.MSg/go-Fut.MPl

Sita will go to Delhi and Ram to Agra.’

(170) Person Mismatch:

- a. (Ellipsis-in-Coordination-PS-38) 1 + 3:
 mĒ Dilli aur vo Agra ?jaa-egaa/*jaa-ũgaa/???jaa-ẽge
 I.m Delhi and he Agra go-Fut.3MSg/go-Fut.1MSg/go-Fut.MPl
 ‘I will go to Delhi and he to Agra.’
- b. (Ellipsis-in-Coordination-PS-39) 3 + 1:
 vo Dilli aur mĒ Agra *jaa-egaa/?jaa-ũgaa/???jaa-ẽge
 he.m Delhi and I Agra go-Fut.3MSg/go-Fut.1MSg/go-Fut.MPl
 ‘He will go to Delhi and I to Agra.’

When local agreement in the second conjunct yields ‘default’ features: 3/M/Sg, the deviance is less severe.

Less Restrictions: we saw that Gapping could not cross finite clause boundaries. This seems to be possible (if mildly degraded) with RNR:

- (171) a. (Ellipsis-in-Coordination-PS-40) RNR possible out of embedded finite clause:
 mujhe lag-taa hai [ki Sita] aur tumhẽ lag-taa hai [ki Mona
 me.Dat feel-Hab be.Prs that Sita.f and you.Dat seem-Hab be.Prs that Mona.f
Mumbai jaa-egii]
 Mumbai go-Fut.f
 ‘I think that Sita, and you think that Mona, will go to Mumbai.’
- b. (Ellipsis-in-Coordination-PS-41) RNR is possible out of relative clauses:
 Atif-ne paṛh-ii thii [vo kitaab [jo Ravi-ne]] aur Mona-ne paṛh-ii
 Atif-Erg read-Pfv.f be.Pst.f that book.f Rel Ravi-Erg and Mona-Erg read-Pfv.f
 thii [vo kahaanii [jo Tina-ne likh-ii thii]]
 be.Pst.f that story.f Rel Tina-Erg write-Pfv.f be.Pst.f
 ‘Atif read the book that Ravi and Mona read the story that Tina [wrote].’
 Gapping out of embedded finite clauses is degraded but perhaps not completely out.

Gapping out of relative clauses is totally out.

Right Peripherality

- (172) a. (Ellipsis-in-Coordination-PS-42) SO CP and SO [V Aux] CP:
 *Atif-ne Sita-se [ki dharti gol hai] aur Tina-ne Mona-se kah-aa
 Atif-Erg Sita-with that earth round is and Tina-Erg Mona-with say-Pfv
 thaa [ki chaand chaukor hai]
 be.Pst that moon square is

‘Atif told Sita that the earth was round and Tina Mona that the moon was square.’

- b. (Ellipsis-in-Coordination-PS-43) Right Peripherality restored: S [NP[CP]] and S [NP[CP]] V Aux:

Atif-ne [yeh baat [ki dhartii gol hai]] aur Tina-ne [yeh baat [ki Atif-Erg this thing.f that earth round is and Tina-Erg this thing.f that chaand chaukor hai]] kah-ii thii moon square is say-Pfv be.Pst

‘Atif said that the earth was round and Tina that the moon is square.’

RNR fed by rightward movement:

- (173) a. (Ellipsis-in-Coordination-PS-44) (the English translation involves both RNR and gapping)

Atif-ne kitaab aur Mona-ne tasviir dii thii Tina-ko
Atif-Erg book.f and Mona-Erg picture.f give.Pfv.f be.Pst.f Tina-Dat

‘Atif had given a book and Mona a picture to Tina.’

- b. (Ellipsis-in-Coordination-PS-45) rightward movement in second conjunct only: not totally out!

(??)Atif-ne Ravi-ko kitaab aur Mona-ne tasviir dii thii
Atif-Erg Ravi-Dat book.f and Mona-Erg picture.f give.Pfv.f be.Pst.f

Tina-ko

Tina-Dat

‘Atif had given a book to Ravi and Mona a picture to Tina.’

→ maybe the right peripherality constraint needs to be fully satisfied only in the first conjunct.

- (174) (Ellipsis-in-Coordination-PS-46) a structure may have both RNR and Gapping:

Atif-ne Mona-se kah-aa aur Tina-ne Ravi-se [ki dhartii gold hai]
Atif-Erg Mona-with say-Pfv and Tina-Erg Ravi-with that earth round is

‘Atif told Mona and Tina to Ravi that the earth is round.’

- (175) Differences between Gapping and RNR in Hindi-Urdu:

- a. RNR is subject to agreement restrictions, Gapping is not
- b. RNR obeys a right peripherality condition, Gapping does not.
- c. Gapping cannot cross finite clause boundaries, RNR can. [check]
- d. Gapping must take tense/aspect auxiliaries along with it, RNR doesn't have to.

29.5 Sluicing

Tbd.

Part VIII

Miscellaneous Syntax

Chapter 30

Questions

Nothing here.

Chapter 31

Imperatives

Nothing here.

Chapter 32

Vocatives

Nothing here.

Chapter 33

Parentheticals

Nothing here.

Chapter 34

Fragments

Nothing here.

Chapter 35

Idiomatic Constructions

35.1 Introduction

Several types of lexical items and syntactic constructions which occur with relative frequency in Hindi-Urdu are outside the range of phenomena accounted for by the principles of Hindi grammar laid out in the rest of these guidelines, either because they are semantically idiosyncratic or idiomatic, or because we do not consider their internal structures to be of syntactic interest. This section and following section describe the treatment of the following such types of lexical items and syntactic constructions.

1. Names (Section 36, page 281), numbers, dates
2. Foreign words (Section 37, page 285)
3. *X-ke calte hue* Constructions
4. *X-ke rang-me bhang paR gayaa* Constructions

The latter two constructions are discussed in this section. They are characterized by idiosyncratic mappings between their forms and their meanings.

35.2 *X-ke calte hue* Constructions

The construction we will refer to as the *X-ke calte hue* construction is schematized below.

- (176) X *ke* calte hue
X Gen walk Part
'While X was happening' *X is replaced by any NP*

In this construction, which always occurs as an adjunct, the participial form of the verb *calte* 'walk' occurs with a genitive phrase and conveys a meaning 'happen' or 'going on', resulting in an adjunct phrase with the meaning *While X was happening, ...*

The following sentences illustrate this construction with naturally occurring examples.¹

(177) a. (sent-X-ke-calte-hue-PS-1)

surakRAkarmiyoM xvArA kI baxawamIjI ke calawe bipASA kA
 surakRAkarmiyoM xvArA kI baxawamIjI ke calawe bipASA kA
 security.workers by done.f misbehavior Gen moving.Part Bipasha Gen
 gussA sAwaveM AsamAna para jA pahuMcA
 gussA sAwaveM AsamAna para jA pahuMcA
 anger 7th sky go arrived

‘Due to the misbehavior of the security staff, Bipasha’s anger reached the seventh sky.’

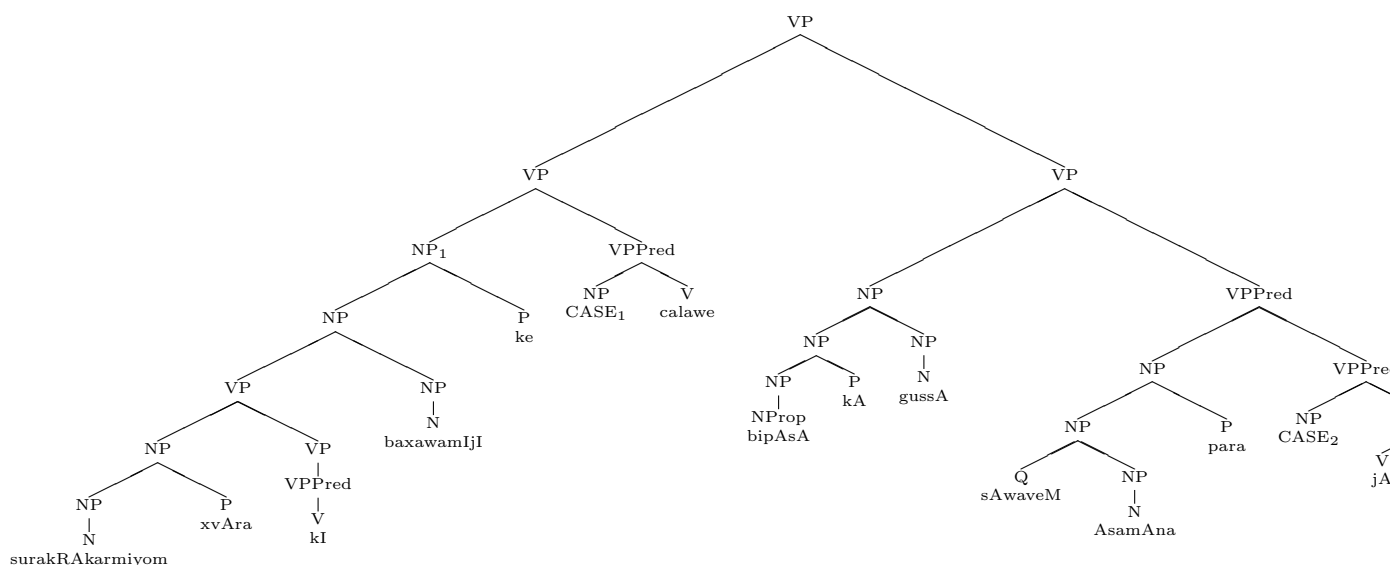
b. (sent-X-ke-calte-hue-PS-2)

baMgAlI bAlA ke sAWa baxawamIjI ke calawe gArdsa ke KilAPa koI
 baMgAlI bAlA ke sAWa baxawamIjI ke calawe gArdsa ke KilAPa koI
 Bengali lady Gen with misbehavior Gen moving guards Gen against any
 kArravAI nahIM kI gal
 kArravAI nahIM kI gal
 action Neg do GO.Pfv

‘Still considering the issue of misbehavior with the Bengali Beauty, no action has yet been taken against the guards.’

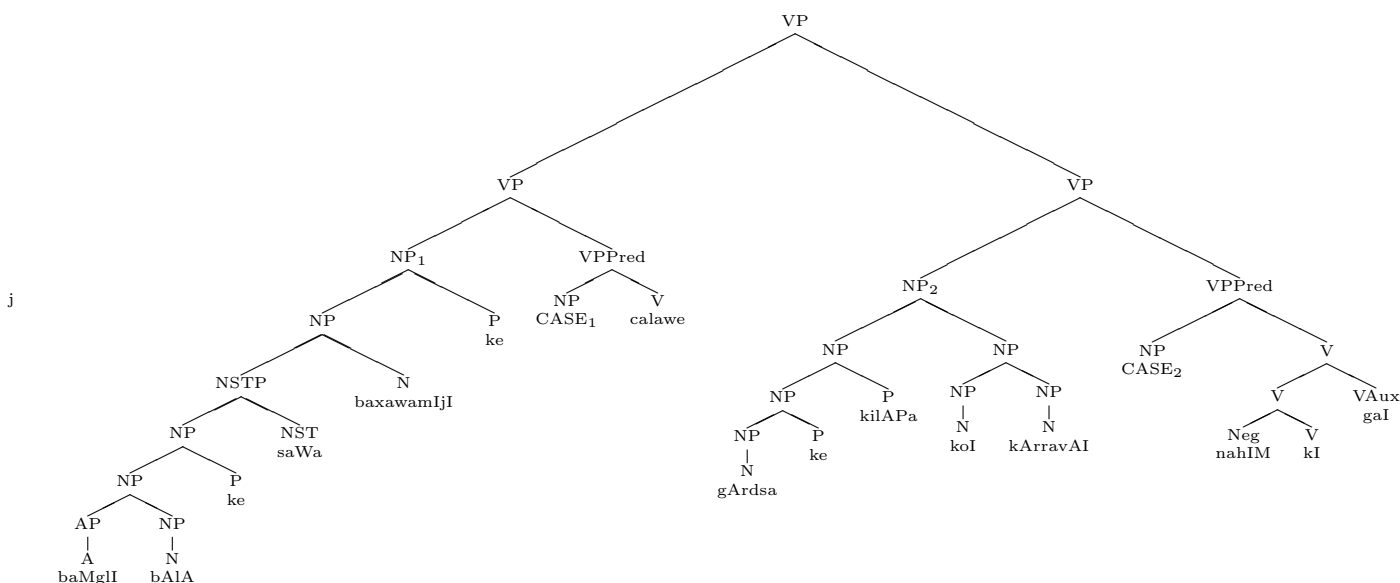
We treat these constructions as unaccusative verb phrases, with the genitive as the unaccusative subject of *calawe*. The whole unaccusative verb phrase is adjoined to the relevant main clause VP.

Tree for (177a) (sent-X-ke-calte-hue-PS-1)



¹The sentence in (177a) is modified slightly from 500-random, file 2, Sentence id=3; that in (177b) is modified from 500-random, file 2, Sentence id=6.

Tree for (177b) (sent-X-ke-calte-hue-PS-2)



35.3 X-ke rang-me bhang paR gayaa Construction

The second idiomatic construction has the form shown in (178).

- (178) (X-ke) rang me bhang paR gayaa
 X-Gen color in contaminant fall go
 ‘Things (at/in X) got messed up..’

In this construction, the phrase *rang me bhang paR gayaa* (lit. ‘contaminant fall into color’) has the idiomatic meaning ‘things got messed up.’ An optional, but usually present, genitive phrase on *rang* ‘color’ provides further information about X.

The following examples illustrate this construction with sentences from the corpus. Note that, in the second example, the *ke*-phrase is not present.

- (179) a. (sent-X-ke-rang-me-bhang-paR-gayaa-PS-1)
 bqhaspawivAra ko jZI meM SurU hue 36veM aMwarrARtrIya Pilma
 bqhaspawivAra ko jZI meM SurU hue 36veM aMwarrARtrIya Pilma
 Wednesday on Zee in start be.Pfv 36th international film
 mahowsava ke raMga meM BaMga padZA jaba vahAM para wEnAwa
 mahowsava ke raMga meM BaMga padZA jaba vahAM para wEnAwa
 festival Gen color in contaminant fall.Pfv when there on stationed
 surakRAkarmiyOM ne bOYIIVuda aBinewrI bipASA basu ke sAWa
 surakRAkarmiyOM ne bOYIIVuda aBinewrI bipASA basu ke sAWa
 security.workers Erg Bollywood Gen.f actress Bipasha Basu Gen
 xuyyarvahAra kiyA
 xuyyarvahAra kiyA
 with misbehavior

‘Things got messed up at the 36th International Film Festival, having started on Thursday on the Zee channel, when the security guards stationed there misbehaved with Bollywood actress Bipasha Basu.

b. (sent-X-ke-rang-me-bhang-paR-gayaa-PS-2)

lekina aBinewrI ke isa kaxama se vahAM raMga meM BaMga
 lekina aBinewrI ke isa kaxama se vahAM raMga meM BaMga
 but actress Gen this step from there color in contaminant
 padZa gayA
 padZa gayA
 fell Go.Pfv

‘But due this step by the actor, things got messed up there.’

In the PS, these phrases are treated as normal unaccusative verb phrases: *bhang* ‘contaminant’ is the unaccusative subject of *paR* ‘fly’ and the postposition headed by *rang* is adjoined to the unaccusative VP headed by *paR* ‘fly.’

Tree for (??)

Tree for (??)

The treatment of this idiomatic expression as syntactically normal is supported by the fact that its component phrases can be modified and interact with other syntactic processes, such as scrambling. In the example below, an adverbial expression intervenes between *bhang* and the postposition headed by *rang* ‘color.’ The tree is equivalent to the tree for (179a), except that the time expression is adjoined to VPPred, to the Spec, VP position above which the unaccusative subject must move.²

(180) (sent-X-ke-rang-me-bhang-paR-gayaa-PS-3)

bqhaspawivAra ko jZI meM SurU hue 36veM aMwarrARtrIya Pilma
 bqhaspawivAra ko jZI meM SurU hue 36veM aMwarrARtrIya Pilma
 Wednesday on Zee in start be.Pfv 36th international film
 mahowsava ke raMga meM BaMga usa samaya padZA jaba vahAM para
 mahowsava ke raMga meM BaMga usa samaya padZA jaba vahAM para
 festival Gen color in contaminant that time fall.Pfv when there on
 wEnAwa surakRAkarmiyOM ne bOYIIvuda aBinewrI bipASA basu ke sAWa
 wEnAwa surakRAkarmiyOM ne bOYIIvuda aBinewrI bipASA basu ke sAWa
 stationed security.workers Erg Bollywood beauty Bipasha Basu Gen with
 xuyarvahAra kiyA
 xuyarvahAra kiyA
 misbehavior do.Pfv

‘Things got messed up at the 36th International Film Festival, having started on Thursday on the Zee channel, when the security guards stationed there misbehaved with Bollywood actress Bipasha Basu.’

Tree for (180) (sent-X-ke-rang-me-bhang-paR-gayaa-PS-3)

²(180) is Sentence id=2, file 1)

Chapter 36

Names

Names and certain other nouns are distinguished from all other nouns through the use of the POS tag *NProp*. Items which receive this label include—in addition to all proper names (names of persons, places, companies, etc.)—dates and days of the week.

(181) Compound noun:

a. (Relation-DS-jk1-2) Names:

rAma ne **mohana** xvAra **rAjA** ko tikata KArixayAye
rAma ne **mohana** xvAra **rAjA** ko tikata KArixayAye
Ram Erg Mohan by Raja Dat ticket buy.CAUSE.Pfv.M

‘Ram made Mohan buy tickets for Raja.’

b. (Relation-DS-rsp-1) Date; country:

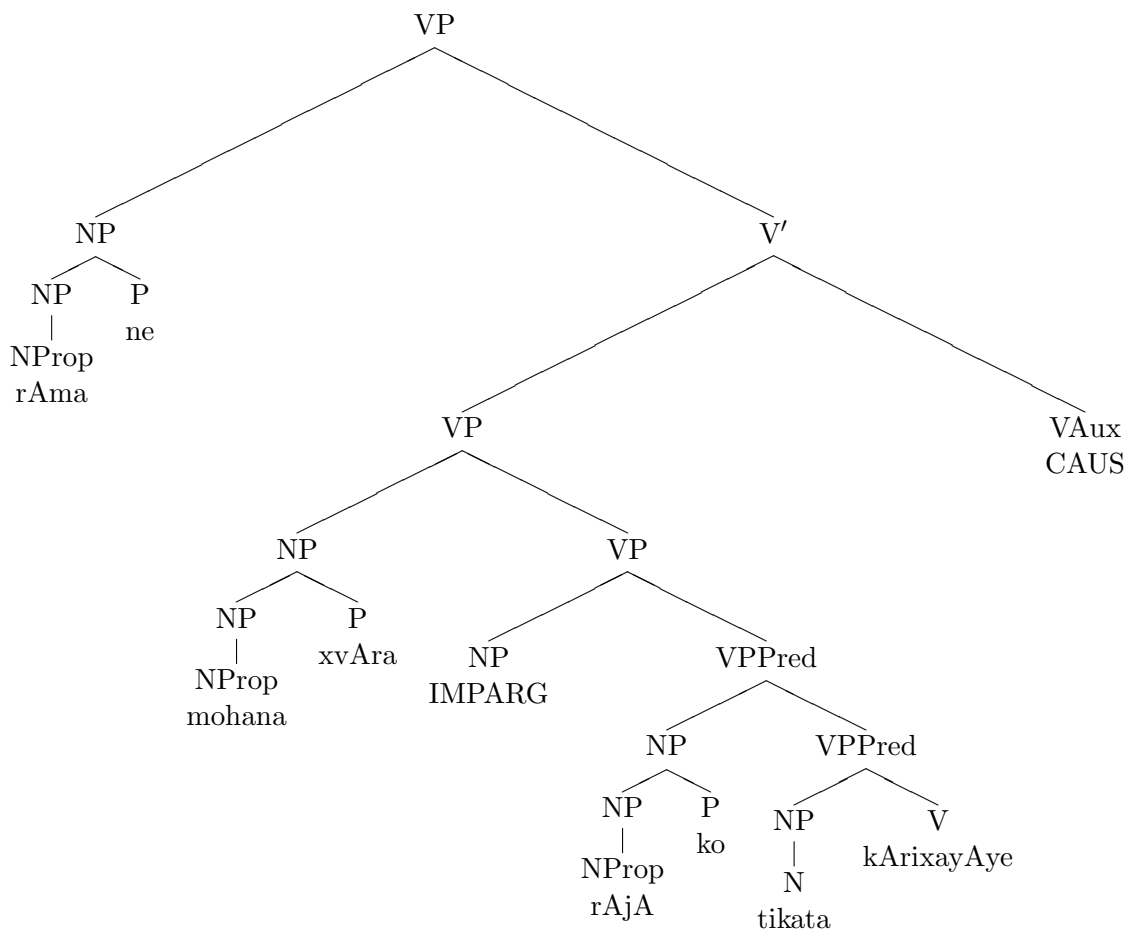
1990 se lekara **2000** waka **BArawa** kI pragawi wejZa rahI
1990 se lekara **2000** waka **BArawa** kI pragawi wejZa rahI
1990 from taking 2000 till India of development fast Prog.F

‘India was fast developing from 1990 till 2000.’

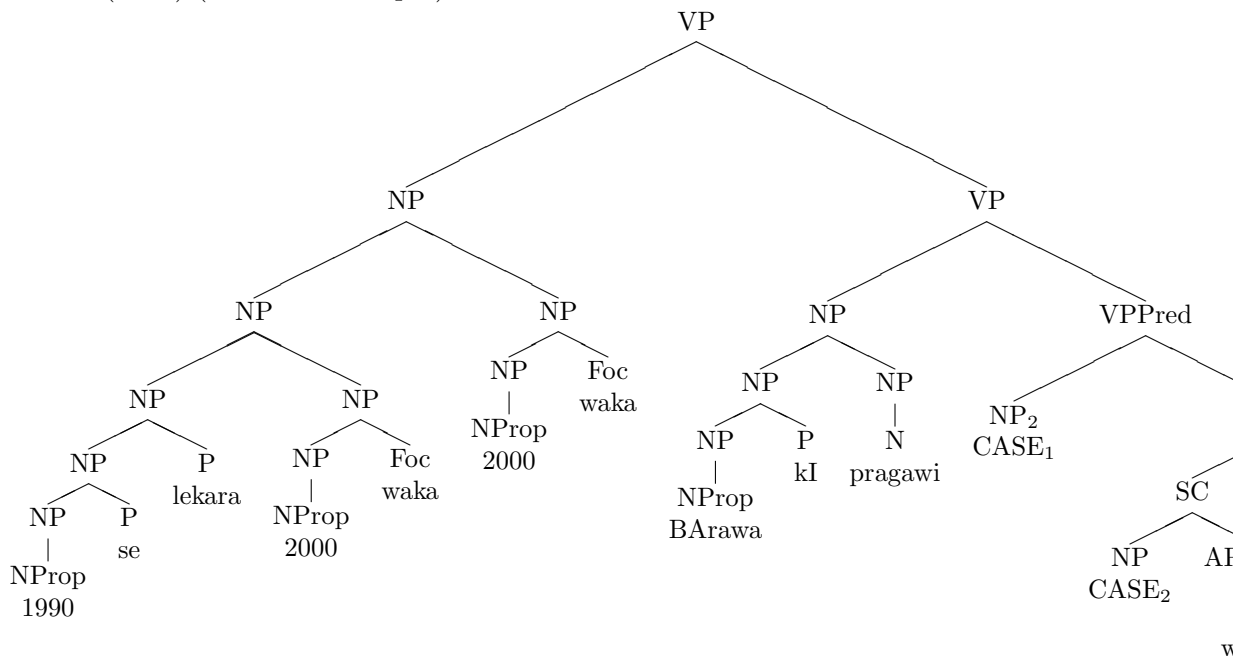
In the PS, we distinguish proper names from other nouns through use of the POS tag *NProp*. The *NProp* label does not project up to the maximal phrasal level, as shown below.

(182)
$$\begin{array}{c} \text{NP} \\ | \\ \text{NProp} \\ | \\ \text{rAma} \end{array}$$

(183) a. Tree for (181a) (Relation-DS-jk1-2)



b. Tree for (181b) (Relation-DS-rsp-1)



While the above distinction between “proper” nouns and all other nouns is largely irrelevant from the perspective of PS, we annotate it for ease of conversion. We note that eliminating

this tagging distinction, and labeling all nouns N, would be preferable in a version of the PS guidelines intended to be used independently of the other treebank levels.

Chapter 37

Foreign Words

Foreign words and loanwords are used in Hindi, particularly as the nonverbal element in N/V-V compound verbs.

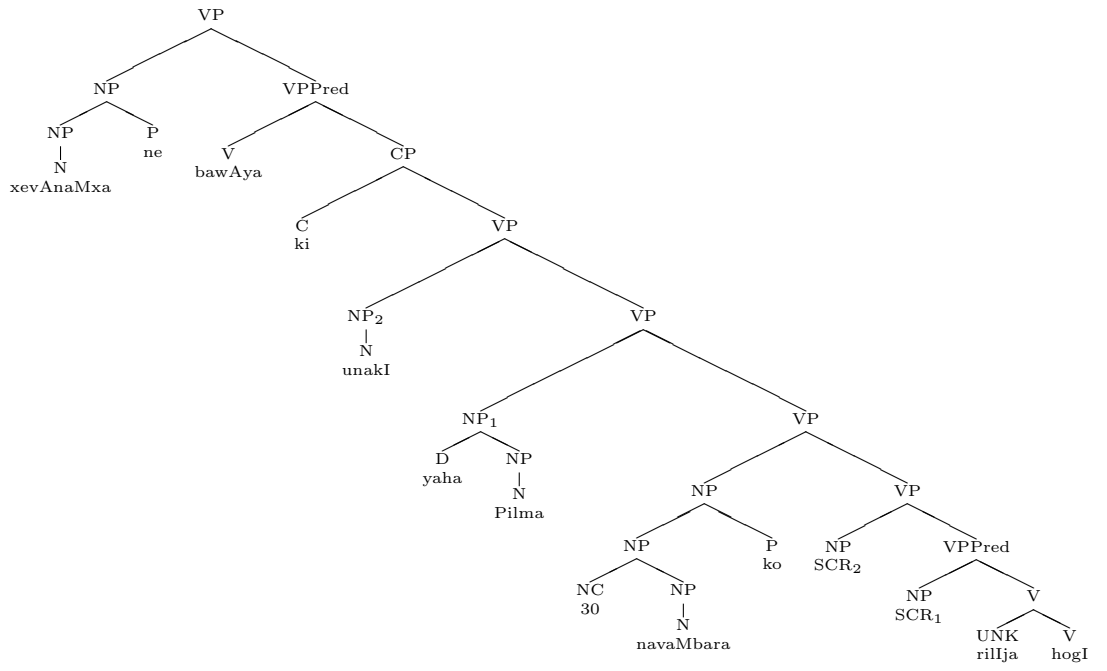
(184) (UNK-PS-1)

xevanamxa ne bawaya ki unaki yaha pilma 30 navambara ko rillja
xevAnaMxa ne bawAyA ki unakI yaha Pilma 30 navaMbara ko **rillja**
Dev.Anand Erg said that he.Gen.f this film 30 November on **release**
hogi
hogI
be.Fut.F

‘Dev Anand said that this film of his will be released on November 30.’

Non-Hindi-Urdu loanwords are assigned the POS tag *UNK*. When they occur as part of a compound verb, the UNK node is head adjoined to the verb. The tree below illustrates for the sentence in (184).

(185) Tree for (184) (sent-UNK-PS-1) Foreign word used in compound:



Part IX

Non-Syntactic Phenomeona

Chapter 38

Punctuation

Nothing here yet.

Chapter 39

Other Formatting in Written Language

Sometimes, written language has nonstandard formatting which we want to include in the syntactic trees.

39.1 Speaker Marking in Dialog

These are examples such as:

Rahul — Your tears are moistening your breath!

DS has the dash as head, and PS uses a PunctP

Chapter 40

Dates and Expressions with Numbers

This is about amounts of money, scores, etc.

Part X
Appendix

Appendix A

Confusion Areas

Appendix B

Information about the Treebank

Appendix C

Conversion

Appendix D

Summary of Empty Categories

ocr This is from the 2010 LREC paper verbatim and needs to be updated and checked!

| | Description | Type | Where Inserted | Label | Coindexed in PS? | Sec |
|----|--|--------|-------------------|-----------|------------------------------------|-----|
| 1 | Empty (Head of an) NP (ellipsis) | silent | DS | *HEAD-NP* | no | |
| 3 | Empty subject with predicative adjective and “ki” complement clause | silent | DS | *pro* | no | |
| 5 | Empty subject or object (regular pro-drop) | silent | PB | *pro* | no | Sec |
| 6 | Empty subject of a non-finite clause (control) | silent | PB | *PRO* | reference, but only if explicit | |
| 7 | Empty relative pronoun | silent | PB | *RELPRO* | no | |
| 8 | Trace of phrase undergoing extraposition (rightward movement) | trace | PS | *EXTR* | movement | |
| 9 | Trace of phrase moving to case position in verbal projection | trace | PS | *CASE* | movement | |
| 11 | Trace of scrambling (leftward movement) | trace | PS | *SCR* | movement | |
| 12 | Trace of head moving to incorporate | trace | PS | *HEAD* | movement | |
| 13 | Causative head | silent | PS | *CAUS* | no | |
| 14 | Complementizer | silent | PS | *COMP* | no | |

Table D.1: Empty category table. In column “Type”, *trace* means that the empty category shows the original location of a moved element while *silent* indicates an unmoved category that just happens to lack an overt phonological realization. Column “Coindexed in PS?” shows whether the empty category has co-indexation with another element in the PS, and if so, whether the co-indexation is due to movement or reference.

D.1 The Role of Empty Categories

D.1.1 Types of Empty Categories

We make a high-level distinction between different kinds of empty categories on the basis of whether they are postulated to mark displacement or not. Empty categories that mark displacement are called **traces**. In theories that use movement as a device to derive valid syntactic representations, traces are created as a result of movement. A trace is always co-indexed with another constituent in the sentence, and the interpretation is that the co-indexed constituent was, at a previous phase of the derivation, in the location of the trace (or, in some theories, at the same time).

All other empty categories are grouped under the label **silent**. These empty categories do not mark displacement; rather they represent undisplaced syntactic elements which happen to lack phonological realization. The label **silent** covers silent proforms as well as silent heads. Silent proforms are like other proforms (such as pronouns): they are co-referential with another element in the sentence or elsewhere in discourse, or implicit in the discourse situation. Examples include empty subject pronouns (*Are you hungry? Don't know.*), or empty clauses in Null Complement Anaphora (*A: Bill will be late again. B: I know.*). Silent heads resemble their non-silent counterparts. They differ from silent proforms in that they are not co-referential with other heads in the sentence or in the discourse context; they have their own reference. Examples include silent conjunctions, silent causative heads, and silent complementizers.

A different way of classifying empty categories emerges if we examine whether they can or must be co-indexed, and if they are what does the co-indexation indicate? As mentioned above, **traces** form a uniform class and always represent the movement relationship that forms them via co-indexation. The **silent** class is heterogenous with respect to co-indexation. The empty categories in this class are not a product of movement and co-indexation with elements of this class indicates a semantic dependency instead. However, we do not represent all semantic dependencies in the phrase structure; only those semantic dependencies which are completely determined by the syntax. Pronominal coreference (e.g. the relationship between *John* and *he* in *When John came back, he was upset*) is not fully determined by the syntax and for this reason it is not represented in the phrase structure. The referent of the silent PRO subject of an infinitival clause is sometimes fully determined by the syntax (*John wanted [PRO to dance]*), in which case it is represented via co-indexation; but in other cases the referent is left unspecified (*[PRO to dance] is fun*), and is not co-indexed.

D.1.2 Linguistic Motivation for Empty Categories

The primary reason why linguistic theories postulate ECs is that it allows for simpler descriptions. Often we find that sentences with certain empty elements have essentially the same properties (interpretation, case-marking, agreement) as the corresponding sentences where the element is overtly realized. In such cases, assuming that the putatively empty element is in fact realized by an empty category allows for simpler analyses; without such a

postulation we would need two sets of rules - one where the element is overtly realized and one where it is not overtly realized. If the trees in a treebank are intended to follow the principles of some linguistic theory, then if the linguistic theory in question assumes ECs, so does the treebank.

Traces are the result of choosing a linguistic representation which includes movement, as we do for PS. While the result of movement (typically) reflects surface order (such as a fronted *wh*-element), the underlying position reflects other important information (such as licensing by a lexical predicate, i.e., lexical predicate-argument structure). Since we want to represent both surface order and underlying position, we use traces.

D.1.3 Empirical Motivation for Empty Categories

A different kind of motivation for postulating ECs comes from the demands of natural language processing, in particular information extraction, question-answering, and related semantic tasks. The more precise and detailed our predicate argument structures are (including empty categories), the more complete our event descriptions will be, and therefore the more effective our semantic processing techniques will be. As argued above, in certain representations, traces provide information as to the true governing head of the moved argument, which can also be used in appropriate semantic role labeling and semantic processing. Consider the difference between *[Which candidate]_i do you expect t_i to win?* and *[Which prize]_i do you expect to win t_i?* The position of trace and the coindexation between trace and its antecedent indicate different readings of two sentences with very similar wordings.