Constituent order in the Tibetan noun phrase

Edward Garrett and Nathan W. Hill
eg15@soas.ac.uk and nh36@soas.ac.uk

1. Introduction

The majority of Tibetan grammars lack explicit discussion of constituent order inside the Tibetan noun phrase (e.g. Schwieger 2006). Beyer's grammar is an exception (1992: 204-233). He distinguishes seven positions in a noun phrase: nominal, determiner, reflexive, numeral, plural, totalizer, and selector. These terms he defines semantically, e.g. "[s]electors, unlike plurals, do not specify simply that there is more than one entity referred to; instead, selectors specify what we can call the RANGE of entities referred to in the set denoted by the nominal" (1992: 232 emphasis in original). Because his analysis is illustrated anecdotally it is often difficult to know how Beyer would analyze a particular Tibetan NP. In some details Beyer's analysis is clearly wrong. For example, he treats *dag* as a 'plural' but *rnams* as a 'selector'.[1] In support of this analysis Beyer claims that *rta dag rnams* 'horses' is a possible order but not **rta rnams dag* 'horses' (1992: 205). However, Michael Hahn draws attention to the occurrence of both the orders *rnams dag* and *dag rnams* in the Tibetan translation of the *Buddhacarita* (2003[1978]). Thus, Beyer's syntactic argument for placing these two words in distinct part-of-speech categories does not hold. This case compels one to fear that Beyer arrived at his analysis impressionistically rather than empirically. A part-of-speech tagged corpus of Tibetan texts provides a convenient way to empirically investigate Tibetan noun phrase structure.

2. The Tibetan corpus and tag set

This paper gives a schematic presentation of the order of constituents in the Tibetan noun phrase, as revealed by an investigation of a part-of-speech tagged corpus of Tibetan texts. The corpus was produced by a project based at the School of Oriental and African Studies, University of London,[2] taking advantage of the profusion of digitised Tibetan texts that are now available online.[3]

The corpus can be divided into two parts, the reference corpus and the extended corpus. The reference corpus includes only highly secure analyses, namely those materials that have been hand-tagged and checked by a human being. By contrast, the extended corpus includes additional materials that have been tagged automatically by a computer. On the basis of insights gained by studying the reference corpus, we have created software that assigns a part-of-speech tag to each word in an unanalysed Tibetan text, after first segmenting the text into words. Garrett et al. (2014) describes Version 1.0 of the rule-based part-of-speech tagger, designed for Classical Tibetan materials.

---

[1] We tag both *dag* and *rnams* as [d.plural].

[2] The project 'Tibetan in Digital Communication' is funded by the U.K's Arts and Humanities Research Council.

[3] Available digital corpora include the entire Derge Kanjur (www.thlib.org/encyclopedias/literary/canons/kt/catalog.php#cat=d on 10 June 2014) and over 60 Dunhuang documents (otdo.aa.tufs.ac.jp on 10 June 2014).

For a complete description of the tag set used in the corpus, including a discussion of nominal heads and case markers, see Garrett et al. (2015). The following list focuses only on the part-of-speech categories for modifiers that occur within the noun phrase.

[adj]   Adjectives: Those words that can occur in attributive position and that are not verbal nouns, we define as adjectives. Tibetan adjectives are almost always multisyllablic, with the paragon varieties ending in *-po*, *-mo*, *-pa*, *-ma*, *-can*, *-ldan*, and *-med*. Etymologically *-can*, *-ldan* 'with', and *-med* 'without' are verbs, but when they occur in forms that are functioning nominally they cannot be analyzed as verbs. So, we treat them together with the preceding syllable as an adjective. Although words such as *nag* 'black', *gsar* 'new', and *che* 'big' are frequently treated as adjectives for pedagogical purposes, a single syllable in predicate position before verbal suffixes is a verb. These words may appear to occur attributively, but we see this as an instance of the formation of compounds. For example, the compound *blon-chen* 'prime minister' contrasts with noun and adjective pair *blon-po chen-po* 'great minister'.

[d.dem] Demonstratives: This tag is used for the demonstratives *ḥdi* 'this' and *de* 'that'.

[d.det] Determiner: The most frequent determiner is *gźan* 'other'. In addition, we identify *ya-re* 'each one (of two)' as a determiner on the basis of the following sentence: *Brgya-byin daṅ Tshaṅs-paḥi rgyal-pos lag-pa ya-re nas zin te* 'The kings Indra and Brahma each took him by one of his hands'. We reckon *ḥbaḥ* 'sole' as a determiner on the basis of sentences such as *rus-pa daṅ khrag ḥbaḥ źig gis sa rtsog-rtsog ltar ḥdug-pa mthoṅ* 'They saw the ground besmirched with only bone and blood'.

[d.emph] Emphasis: This category was initially created for *ñid* in phrases such as *rgyal-po ñid* 'that very king' or *lus ñid* 'this body'. This syntactic use of *ñid* must be distinguished from the use in Buddhist terminology of *-ñid* inside words, e.g. *stoṅ-pa-ñid* 'emptiness'. Apart from *ñid*, we have categorized *kho-na* 'the very, same' and *re-re* 'each' as emphatics. This use of *kho-na* should not be confused with its function as a third person pronoun in Old Tibetan.

[d.indef] Indefinite: This category is used for the allomorphs of the indefinite marker *cig*, *źig*, and *śig* as in *pho-ña cig* 'a messenger'.

[d.plural] Plurals: The plural markers *rnams*, *dag*, *kun*, *thams-cad*, *ḥo-cog* (and its variants) and *tsho* are tagged as their own category 'plural'. However, plural pronouns (*bdag-cag, khyed-cag, ḥu-bu-cag*) are treated as one word. The plural marker *-cag* is not removed because to do so would result in pronominal stems which are not mutually comparable (viz. *bdag* is a singular pronoun, *khyed* a plural pronoun, and *ḥu-bu* has no independent life outside of *ḥu-bu-cag*).

[num.card] Cardinal numbers: We distinguish cardinals (*gcig* 'one', *gñis* 'two', *gsum* 'three', etc.) and ordinals (*daṅ-po* 'first', *gñis-pa* 'second', *gsum-pa* 'third', etc.). Other derivatives of numerals are treated according to their respective syntax, thus *gcig-pa* 'sole' is an adjective, *gñi-ga* 'both' is an indefinite pronoun, etc. In higher numbers each component digit is tagged separately.

[num.ord] Ordinal numbers: As just discussed, in numbers we distinguish cardinals (*gcig* 'one'*, gñis* 'two'*, gsum* 'three', etc.) and ordinals (*daṅ-po* 'first', *gñis-pa* 'second', *gsum-pa* 'third', etc.). Some adjectives can be distinguished from ordinal numbers only in context. For example, in the phrase *rdo-rje rtse-lṅa-pa* 'a five-pronged vajra' the noun phrase syntax and the overall meaning of the passage dictates that *lṅa-pa* is part of the word *rtse-lṅa-pa* 'five-pronged' rather than a word 'fifth'.

To explore the relative order of constituents within the Tibetan noun phrase, we queried the corpus for sequences of words with particular part-of-speech categories, irrespective of the orthographic content of the words themselves. In order to broaden our investigation beyond the patterns noticed by human taggers, we used the machine-tagged extended corpus as our dataset. By setting the 'exclude ambiguous tags' option in the search interface, we excluded from our search results all words of insecure part-of-speech assignment, that is, those cases where, unsure what the tag for a word should be, the rule tagger gives an ambiguous answer.[4] The resulting 'shingles' or sequences of parts-of-speech were then analysed and compared.[5]

3. Summary of Findings

There is a convenient way to find the attested patterns of three element noun phrases (and shorter NPs) by specifying that the fourth element of the shingles window includes those elements that are known to end noun phrases. The elements we chose to require as the fourth word in the search window are case markers [case.xxx] and clitics [cl.xxx]. Case markers end a noun phrase. Clitics come even after case markers. The reason to specify clitics as one option as the fourth constituent is that a noun phrase in the absolutive case (i.e. is not followed by an overt case marker) that is followed by a clitic will be specified by such a search. The shingles interface allows search from the right. We put [+case,cl] into the search window, searching from the right, with a window of four words and returning 500 shingles (cf. Illustration 1). This search captures a series of four words in which the fourth word is a case marker or a clitic. From the results of such a search it is necessary to disregard cases, converbs, clitics, negation, finite verbs,

---

[4] This has disadvantages such as precluding exploring the relative frequency of the orders *rnams dag* and *dag rnams* because the sequence of letters *dag* is ambiguous between the plural marker and a verb 'be pure'. Nonetheless, the corpus permits such investigations, which we shall benefit from returning to with an improved understanding of Tibetan noun phrase structure on the basis of words of unambiguous part-of-speech.

[5] Normally, shingle search returns POS tags only. By selecting the "show word forms" option, word forms are returned instead. This is important in order to see the specific words in context that underlie the generalized part-of-speech patterns. Preceding a POS item with "+" indicates that only a partial match is required. [+v] will match every tag that starts with "v", and [+v,n.v] will match every tag that starts with either "v" or "n.v". Note that both [+v,+n.v] and [v,n.v] are ill-formed searches.

and punctuation occurring among the first three constituents because these elements are not part of a noun phrase. For example, the most frequent pattern is [punc] [punc] [n.count] [case.gen], which for the purposes of analyzing noun phrase structure counts simply as [n.count].

Illustration 1: rightward rightward search for [+case,cl] (1 January 2015)

The process of disregarding irrelevant entries, combining compatible patterns, and supplementing with additional searches is described in the appendix. Table 1 presents the overall patterns discovered.

Table 1: Summary of Tibetan noun phrase structures

[n.count]	[adj]	[d.det]	[d.plural]	[d.dem]	[d.emph]
[n.count]	[adj]	[d.det]	[d.plural]	[d.plural]	[d.emph]
[n.count]	[adj]	[d.det]	[d.indef]		
[n.count]	[adj]	[num.card]	[d.indef]		
[n.count]	[num.card]	[num.card]	[num.card]	[num.ord]	
[n.v.invar] [n.count]	[n.prop]	[n.count]	[d.dem]		
	[p.pers]	[n.count]			
	[p.pers]	[p.refl]	[d.plural]		

These patterns fall into three groups: noun phrases headed by common nouns, noun phrases headed by proper nouns, and noun phrases headed by personal pronouns.

Common nouns: The first five rows of Table 1 give patterns headed by common nouns. One might be tempted to further combine the patterns [d.plural] [d.dem] [d.emph] and [d.plural] [d.plural] [d.emph] as [d.plural] [d.plural] [d.dem] [d.emph], but this is not possible because [d.plural] [d.plural] [d.dem] is not an attested pattern. If we look in more detail at the pattern [d.plural] [d.plural] it becomes clear that certain words sit more comfortably as the first of two plurals whereas others sit more comfortably as the second of two plurals. In particular, the patterns *rnams thams-cad*, *rnams kun*, and

rnams sna-tshogs suggest that the plural marker *rnams* is of quite a different nature than these others words, which imply something about quantification. Thus, Beyer may be right to divide 'plurals' from 'selectors' and to put *thams-cad* and *kun* in the latter category (1992: 232). If so, *rnams* should count as a 'plural' and not a 'selector' as Beyer has it. The rare patterns *thams-cad kun* and *sna-tshogs kun* suggest that even a category of 'selectors' would not be homogenous.

A surprising omission in Table 1 is the pattern [d.dem] [d.plural], which occurs in the corpus 399 times, much more frequently than the 72 attestations of the pattern [d.plural] [d.dem]. The reason that this pattern was not included in Table 1 is that it did not combine felicitously with the other patterns. Noting that 154 of the occurrences of pattern [d.dem] [d.plural] occur after the *śad* punctuation mark, it seems likely that this is a pattern typical of pronominal use of *de* and *ḥdi*.

The first four rows of Table 1 make clear that noun phrases are either marked as indefinite, with [d.indef], or they are marked as definite, with a plural marker [d.plural], a demonstrative [d.dem] or both. Table 1 might also hasten the conclusion that that [d.emph] is only compatible with definite marked noun phrases, but the absence of the pattern [d.indef] [d.emph] may be an accidental gap.

As a final remark about noun phrases headed by common nouns, the patterns [adj] [num.ord] and [adj] [num.card] [num.ord] are not attested. This potentially suggests that ordinal numbers are syntactically treated the same as adjectives.

Proper nouns: The sixth row of Table 1 gives the one overarching pattern of noun phrases headed by a proper noun, viz. [n.v.invar] [n.count] [n.prop] [n.count] [d.dem]. Although though modifiers typically follow their heads in Tibetan noun phrases, as seen in this pattern, modifiers both precede and follow proper nouns. The initial [n.v.invar] is typically *ḥphags-pa* 'exhaulted' or *dam-pa* 'sacred'. The remaining elements or the pattern are illustrated by the phrases *pha-rgan Mar-pa lo-tstsha* 'old father, Marpa the translator' and *Mar-pa lo-tstsha ḥdi* 'this Marpa the translator'.

Personal pronouns: The final two rows of Table 1 give the two patterns of noun phrases headed by personal pronouns: [p.pers] [n.count] as seen in *ṅed ma-smad* 'we, mother and children' and [p.pers] [p.refl] [d.plural] as seen in *ṅed raṅ tsho* 'we'.

Although this study has merely touched upon the mysteries of the Tibetan noun phrase a few tentative conclusions are possible. First, the words that we tag [d.plural] can be further sub-classified with *rnams* in one category and *thams-cad, kun*, and *sna-tshogs* in another category. Second, proper nouns are peculiar in allowing preceding modifiers. Third, adjectives and ordinals appear to be incompatible.

References

Beyer, S. (1992) *The Classical Tibetan language*. New York: State University of New York.

Garrett, E., N. W. Hill, and A. Zadoks (2014) A Rule-based Part-of-speech Tagger for Classical Tibetan. *Himalayan Linguistics* 13.1, 9-57.

Garrett, E., N. W. Hill, A. Kilgarriff, R. Vadlapudi, and A. Zadoks (2015) The

contribution of corpus linguistics to lexicography and the future of Tibetan dictionaries. *Revue d'Etudes Tibétaines* 32, 51-86.

Hahn, M. (2003[1978]) On the function and origin of the particle dag. *Schluessel zum Lehrbuch der klassischen tibetischen Schriftsprache und Beitrage zur tibetischen Wortkunde* (*Miscellanea etymologica tibetica I – VI*). Marburg: Indica et Tibetica Verlag, pp. 95-104.

4. Appendix 1: early stages in the analysis

We isolated the patterns presented in Table 1 (above) by querying the 'shingles' search interface of the corpus. However, the limits of this interface and the presence of patterns including more than one noun phrase in the search results meant that a certain amount of cleanup was necessary in order to isolate the patterns that drew our attention. We put [+case,cl] into the search window, searching from the right, with a window of four words and returning 500 shingles. This search captures a series of four words in which the fourth word is a case marker or a clitic. From the results of such a search it is necessary to disregard cases, converbs, clitics, negation, finite verbs, and punctuation, occurring among the first three constituents because these elements are not part of a noun phrase. For example, the most frequent pattern is [punc] [punc] [n.count] [case.gen], which for the purposes of analyzing noun phrase structure counts simply as [n.count]. When such entities are disregarded and the frequency calculations are recalculated, the results achieved are what can be seen in Table 5 in appendix 2. These data still contain many patterns that are not legitimate noun phrases. We propose to simplify this picture by disregarding two types of phenomena.

In the first case, in a sequence [n.count] [n.count] we disregard the first noun. The pattern [n.count] [n.count] is a very frequent pattern, occurring 348 times in the corpus. However, according to the protocols of the project this sequence of tags is only possible when two nouns are in apposition or when they belong to two different noun phrases. For example, the most common sequence of words that match the pattern [n.count] [n.count] [+case,cl] is *bcom-ldan-ḥdas rgyal-po ḥi*, occurring three times in the corpus. In this example *bcom-ldan-ḥdas* 'Bhagavan' and *rgyal-po* 'king' belong to two different noun phrases and *bcom-ldan-ḥdas* should be understood as in the absolutive case. A scroll through the results of the search [n.count] [n.count] [+case,cl] does not yield any obvious examples of apposition. Consequently, it is safe to assume that most of the 348 examples of [n.count] [n.count] in the corpus are best taken as examples of the noun phrase structure that consists of a single noun. Put more generally, in those patterns in Table 5 that yield a sequence of two nouns, the first noun can be omitted from analysis, similar to how we ignored nouns in position one followed by a case marker in position two.[6]

In the second case, we disregard all of those noun phrase structures in which the third element is a verbal noun [n.v.xxx]. Although the distinction between an adjective and a nominalized intransitive verb is often somewhat arbitrary, because our project does not encode a distinction between transitive and intransitive verbs, allowing [n.v.xxx] as the right constituent of a noun phrase includes many structures that are not wanted. The

---

[6]As an aside, because it is not strictly speaking incorrect to analyze an apposition of the structure [n.count] [n.count] – for example *yab rgyal-po* 'the father, the king' as an example of the noun phrase structure [n.count], there is in fact no problem at all with disregarding the first noun in such cases.

pattern [n.count] [n.v.invar] may seem like an ordinary noun phrase in a case like *chos dam-pa* 'sacred dharma' (cf. *rgyal-po chen-po* [n.count] [adj] 'great king'), but the pattern [n.count] [n.v.past] is not intuitively amendable to analysis as a noun phrase in a pattern like (*rgyal-po s*) *dgra bsad-pa* '(the king) killed the enemy'. Thus, patterns of the type [n.count] [n.v.xxx] are best addressed within the context of an overall investigation of the relationship of noun phrase structure with verb nominalization and subordination structures. These are issues that deserve attention, but are beyond the scope of this study.

Disregarding the first [n.count] in a sequence of two and all [n.v.xxx] in third position results in the noun phrase structures seen in Table 6 (appendix 2). Some of these patterns are unlikely to be actual noun phrase patterns. Specifically, there are six patterns that have [n.count] in the third position (cf. Table 2).

Table 2: noun phrase patterns with [n.count] in position three

32	[n.count]	[num.card]	[n.count]
22	[n.count]	[d.plural]	[n.count]
21	[n.count]	[adj]	[n.count]
15	[adj]	[d.plural]	[n.count]
10	[n.count]	[n.prop]	[n.count]

Since a noun often heads a noun phrase, these patterns look as if they may have more than one head. We have investigated the specific examples underlying each of these patterns, and many have a second constituent in the absolute case, e.g. *lan gcig yul du* 'once, in a land', *bya rnams nam-mkhaḥ la* 'birds, to the sky', *rgyal-po chen-po nad kyis* 'the great king, by illness', *maṅ-po rnams saṅs-rgyas kyis* 'many, by the buddha'.

The final pattern [n.count] [n.prop] [n.count] includes phrases that consist of two noun phrases, such as *dam-chos Rgya-gar yul nas* 'sacred dharma from the land [of] India' but also phrases that are a single noun phrase such as *pha-rgan Mar-pa lo-tstsha daṅ* 'Old father, Marpa, the translator, and'. Thus, we exclude from further consideration the first four patterns in Table 2, but maintain the fifth.

There are seven patterns that have [n.count] in the second position. These patterns seem *a priori* unlikely to be capturing genuine noun phrases, because an [n.count] typically heads a noun phrase and is followed by, rather than proceeded by, dependent constituents.

Table 3: noun phrase patterns with [n.count] in position two

12 [n.v.fut]	[n.count]	
9 [d.plural]	[n.count]	[d.plural]
8 [num.ord]	[n.count]	
8 [n.v.past]	[n.count]	[num.card]
10 [n.prop]	[n.count]	
10 [n.v.invar]	[n.count]	
9 [p.pers]	[n.count]	

Of the patterns in Table 3 the first three appear not to permit legitimate noun phrases. We provide a common example of each pattern: *bya-ba stan las* 'called, from the seat', *rnams yul-mi rnams daṅ* 'plr., neighbors, and', *daṅ-po theg-chen gyi* 'first, of the Mahāyāna'. [7] The examples underlying the fourth pattern [n.v.past] [n.count] [num.card] often do not consist of single noun phrases, e.g. *ḥkhrugs-pa lan gñis su* 'disturbed, a second time'. However, some cases, such as *bsdus-pa bam-po gsum daṅ* 'collected, as (?) section three' are perhaps amenable to analysis as a single noun phrase. Nonetheless, because such a noun phrase involves a nominalized verb that is governing additional constituents to the left, outside of the shingle frame, such examples are not considered here. The final three patterns in Table 3 do include cases of single noun phrases: *Rgya-gar yul du* 'in the land, India', *ḥphags-pa raṅ-saṅs-rgyas las* 'to the noble Pratyekabuddha', *ṅed ma-smad kyi* 'of us, mother and children'. However, the patterns [n.v.invar] [n.count] and [p.pers] [n.count] also include examples that must be analyzed as two separate noun phrases: *sogs-pa saṅs-rgyas kyis* 'etc., the buddha', *khyod stan la* 'you, on the seat'. Consequently, these patterns as such cannot be used to detect noun phrase boundaries.

A final pattern that came up in the search, but does not select a single noun phrase is [p.pers] [p.pers]. In all examples the two pronouns are different grammatical persons, e.g. *khyod ṅa la* 'thou, to me' or *bdag-cag khyod la* 'we, to thou', and thus obviously are not part of the same noun phrase. [8]

Table 7 (appendix 2) omits those patterns that appear incompatible with a single noun phrase, and also omits single word noun phrases; such words reveal nothing about noun phrase structure. One formalization of Tibetan noun phrase structure would simply consist of the structures present in Table 7 (appendix 2), however this presentation can be simplified, by presuming that trailing members of a noun phrase are optional; with this possibility in mind several of the two and three constituent patterns are combinable into one or more overarching four constituent pattern. For example, [n.count] [adj], [n.count] [d.plural], [n.count] [d.dem], [n.count] [adj] [d.plural], and [n.count] [d.plural] [d.dem] may all be conceptualized as sub-cases of the four constituent patter [n.count] [adj] [d.plural] [d.dem]. If all of the possibilities discussed so far are combined in this way the result is Table 4.

---

[7] It is so difficult to imagine the pattern [d.plural] [n.count] [d.plural] even not partially including two noun phrases; this pattern could thus be used to identify absolutive case marking after the first plural.

[8] The pattern [p.pers] [p.pers] could be used to isolate some noun phrases marked in the absolutive case.

Table 4: Summary of Tibetan noun phrase structures

[n.count]	[adj]	[d.det]	[d.plural]	[d.dem]	[n.rel]
[n.count]	[adj]	[d.det]	[d.plural]	[d.dem]	[d.emph]
[n.count]	[adj]	[d.det]	[d.plural]	[d.plural]	[d.emph]
[n.count]	[adj]	[d.det]	[d.indef]		
[n.count]	[adj]	[num.card]	[d.indef]		
[n.count]	[num.card]	[num.card]	[num.card]	[num.ord]	
[n.v.invar] [n.count]	[n.prop]	[n.count]	[d.dem]		
[n.count]	[p.interrog]	[n.rel]			
[p.pers]	[n.count]				
[p.pers]	[p.refl]	[d.plural]			

This combined presentation takes advantage of some patters that were not revealed in the initial search, but only discovered during the process of attempting the over schematization. In particular the following were helpful: *ñis brgya rtsa bdun-pa* [num.card] [num.card] [num.card] [num.ord] '207th', *kha-che paṇ-chen de la* [n.prop] [n.count] [d.dem] [case.all] 'to that great-pundit Kashmiri', *btsun-mo chuṅ-ṅu gźan dag* [n.count] [adj] [d.det] [d.plural] 'other small queens', *dam-pa rje-btsun Mi-la* [n.v.invar] [n.count] [n.prop] 'holy lord Mila', *srin-po chen-po lṅa źig* [n.count] [adj] [num.card] [d.indef] 'a five great demons', *rnams thams-cad ñid* [d.plural] [d.plural] [d.emph].[9]

Table 1 (above) is a recapitulation of Table 4, excluding the two patterns that end with [n.rel], since it is convenient to see the relator noun as acting on rather than as a member of the preceding noun phrase.

5. Appendix 2: Large tables

6. Table 5: potential noun phrase structures of up to three elements

3620 [n.count]		
490 [n.count]	[d.plural]	
45 [n.count]	[d.plural]	[n.v.invar]
22 [n.count]	[d.plural]	[n.count]
12 [n.count]	[d.plural]	[d.dem]
12 [n.count]	[d.plural]	[d.plural]
9 [n.count]	[d.plural]	[n.v.past]
348 [n.count]	[n.count]	
91 [n.count]	[n.count]	[d.plural]
38 [n.count]	[n.count]	[d.dem]

---

[9] The patterns [adj] [num.ord] and [adj] [num.card] [num.ord] are not attested. This potentially suggests that ordinal numbers are syntactically treated the same as adjectives.

27 [n.count]	[n.count]	[num.card]
19 [n.count]	[n.count]	[adj]
13 [n.count]	[n.count]	[n.v.invar]
9 [n.count]	[n.count]	[n.v.neg]
9 [n.count]	[n.count]	[n.v.past]
8 [n.count]	[n.count]	[n.prop]

264 [n.count]	[n.v.past]
261 [n.count]	[n.v.invar]
83 [n.count]	[n.v.fut.n.v.pres]
54 [n.count]	[n.v.fut]
53 [n.count]	[n.v.neg]
41 [n.count]	[n.v.past.n.v.pres]
27 [n.count]	[n.v.cop]
17 [n.count]	[n.v.fut.n.v.past]

250 [n.count]	[d.dem]	
22 [n.count]	[d.dem]	[n.rel]
18 [n.count]	[d.dem]	[d.emph]

212 [n.count]	[adj]	
22 [n.count]	[adj]	[n.v.invar]
13 [n.count]	[adj]	[d.dem]
11 [n.count]	[adj]	[n.count]
10 [n.count]	[adj]	[d.plural]
10 [n.count]	[adj]	[n.count]

| 202 [n.count] | [n.prop] | |
| 10 [n.count] | [n.prop] | [n.count] |

172 [n.count]	[num.card]	
32 [n.count]	[num.card]	[n.count]
27 [n.count]	[num.card]	[num.card]
15 [n.count]	[num.card]	[n.v.invar]
13 [n.count]	[num.card]	[n.v.past]

| 42 [n.count] | [p.pers] |

| 20 [n.count] | [num.ord] |
| 16 [n.count] | [d.det] |

9 [n.count]   [d.det]   [d.plural]

14 [n.count]   [p.interrog]   [n.rel]

8 [n.count]   [d.indef]
8 [n.count]   [p.indef]

824 [n.v.invar]
10 [n.v.invar]   [n.count]
8 [n.v.invar]   [n.v.pres]
804 [n.v.past]
8 [n.v.past]   [n.count]   [num.card]
414 [n.rel]
396 [n.v.fut.n.v.pres]
8 [n.v.fut.n.v.pres]   [n.v.past]
164 [n.v.fut]
12 [n.v.fut]   [n.count]
127 [n.v.past.n.v.pres]
99 [n.v.pres]
9 [n.v.fut.n.v.past]

254 [n.prop]
19 [n.prop]   [d.dem]
10 [n.prop]   [n.count]

253 [p.pers]
10 [p.pers]   [n.v.invar]
9 [p.pers]   [n.count]
9 [p.pers]   [p.refl]
8 [p.pers]   [d.plural]
8 [p.pers]   [p.pers]

93 [d.dem]
167 [d.dem]   [n.rel]

70 [num.card]   [num.card]   [num.card]

62 [num.ord]
8 [num.ord]   [n.count]

48 [adj]			
15 [adj]	[d.plural]	[n.count]	
47 [d.det]			
37 [p.interrog]	[n.rel]		
26 [d.plural]			
9 [d.plural]	[n.count]	[d.plural]	

Table 6: noun phrase structures excluding [n.count] [n.count] and [n.v.xxx]

3968 [n.count]			
581 [n.count]	[d.plural]		
22 [n.count]	[d.plural]	[n.count]	
12 [n.count]	[d.plural]	[d.dem]	
12 [n.count]	[d.plural]	[d.plural]	
288 [n.count]	[d.dem]		
22 [n.count]	[d.dem]	[n.rel]	
18 [n.count]	[d.dem]	[d.emph]	
231 [n.count]	[adj]		
13 [n.count]	[adj]	[d.dem]	
21 [n.count]	[adj]	[n.count]	
10 [n.count]	[adj]	[d.plural]	
210 [n.count]	[n.prop]		
10 [n.count]	[n.prop]	[n.count]	
199 [n.count]	[num.card]		
32 [n.count]	[num.card]	[n.count]	
27 [n.count]	[num.card]	[num.card]	
42 [n.count]	[p.pers]		
20 [n.count]	[num.ord]		
16 [n.count]	[d.det]		
9 [n.count]	[d.det]	[d.plural]	
14 [n.count]	[p.interrog]	[n.rel]	

8 [n.count]	[d.indef]	
8 [n.count]	[p.indef]	
10 [n.v.invar]	[n.count]	
8 [n.v.past]	[n.count]	[num.card]
414 [n.rel]		
12 [n.v.fut]	[n.count]	
254 [n.prop]		
19 [n.prop]	[d.dem]	
10 [n.prop]	[n.count]	
253 [p.pers]		
9 [p.pers]	[n.count]	
9 [p.pers]	[p.refl]	
8 [p.pers]	[d.plural]	
8 [p.pers]	[p.pers]	
93 [d.dem]		
167 [d.dem]	[n.rel]	
70 [num.card]	[num.card]	[num.card]
62 [num.ord]		
8 [num.ord]	[n.count]	
48 [adj]		
15 [adj]	[d.plural]	[n.count]
47 [d.det]		
37 [p.interrog]	[n.rel]	
26 [d.plural]		
9 [d.plural]	[n.count]	[d.plural]

Table 7: noun phrase structures with all suspect and uninformative structures removed

581 [n.count]	[d.plural]	
12 [n.count]	[d.plural]	[d.dem]
12 [n.count]	[d.plural]	[d.plural]
288 [n.count]	[d.dem]	

22 [n.count]	[d.dem]	[n.rel]
18 [n.count]	[d.dem]	[d.emph]
231 [n.count]	[adj]	
13 [n.count]	[adj]	[d.dem]
10 [n.count]	[adj]	[d.plural]
210 [n.count]	[n.prop]	
10 [n.count]	[n.prop]	[n.count]
199 [n.count]	[num.card]	
27 [n.count]	[num.card]	[num.card]
42 [n.count]	[p.pers]	
20 [n.count]	[num.ord]	
16 [n.count]	[d.det]	
9 [n.count]	[d.det]	[d.plural]
14 [n.count]	[p.interrog]	[n.rel]
8 [n.count]	[d.indef]	
8 [n.count]	[p.indef]	
10 [n.v.invar]	[n.count]	
19 [n.prop]	[d.dem]	
10 [n.prop]	[n.count]	
9 [p.pers]	[n.count]	
9 [p.pers]	[p.refl]	
8 [p.pers]	[d.plural]	
167 [d.dem]	[n.rel]	
70 [num.card]	[num.card]	[num.card]
37 [p.interrog]	[n.rel]	