

HISTORICAL SYNTAX

The syntax of Sanskrit compounds

JOHN J. LOWE

University of Oxford

Classical Sanskrit is well known for making extensive use of compounding. I argue, within a lexicalist framework, that the major rules of compounding in Sanskrit can be most appropriately characterized in syntactic, not morphological, terms. That is, Classical Sanskrit ‘compounds’ are in fact very often syntactic phrases. The syntactic analysis proposed captures the fact that compound formation is closer to a morphological process than other aspects of syntax, and so permits some acknowledgment of the gradient nature of the word–phrase divide, even within a strictly lexicalist theory.*

Keywords: compounds, syntax, lexicalism, Sanskrit, lexical-functional grammar (LFG)

1. INTRODUCTION. Classical Sanskrit is well known for making extensive use of compounding; indeed, modern Western linguistics has adopted the Sanskrit grammatical terms for some compound types, such as *bahuvrīhi* and *dvandva*. In this article, I discuss the morphosyntactic status of Classical Sanskrit compounds. I argue that the major rules of compounding in Sanskrit can be most appropriately characterized in syntactic, not morphological, terms: that is, as syntactic phrases rather than as complex words.

The position of compounds on the cline between syntactic phrase and lexical unit, in particular in English but also crosslinguistically, has been the subject of considerable debate and is a prominent topic in morphological work on the definition of a word. The topic has not been so widely treated in syntactic literature. Within strict lexicalist theories in particular, there is relatively little work on compounds and, perhaps, little motivation to treat them, since a lexical analysis is always possible. I argue within a strict lexicalist framework that, while some Classical Sanskrit compounds are indeed lexical, the productive rules of compounding must be considered part of the syntax of the language. At the same time, the syntax of Sanskrit compounds is in certain respects different from the ‘ordinary’ syntax of the Sanskrit clause and is, arguably, closer to a morphological/lexical process. I present an analysis, based within a strict lexicalist syntactic theory, that captures, at least to an extent, the fact that compound formation is closer to a lexical process than other aspects of syntax. It therefore permits some acknowledgment of the gradient nature of the word–phrase divide, even within a strictly lexicalist theory, and is particularly relevant for analyzing diachronic processes such as univerbation, as well as grammaticalization and degrammaticalization.

2. SANSKRIT SYNTAX AND COMPOUNDING. Sanskrit is an old Indo-Aryan language that has been widely used on the Indian subcontinent for over three thousand years. It was and still is used primarily as a literary and academic language, but also as a spoken language; for a long time it was the main lingua franca in use across South Asia.¹ Classical Sanskrit is a relatively standardized form of Sanskrit, codified on the basis of a native grammatical tradition. The earliest surviving, and most important, work in this

* I am very grateful to Jim Benson, Mary Dalrymple, and Oleg Belyaev for helpful discussion and comments on earlier versions of this work, and also to the audience at SE-LFG15 for their attention and comments, in particular Andy Spencer. I am also very grateful to Brendan Gillon and two anonymous referees for their comments on this article, which enabled me to correct a number of problems. This work was undertaken while in receipt of an Early Career Research Fellowship funded by the Leverhulme Trust.

¹ On the position of Sanskrit in pre-Modern India, see Pollock 2006, and on its demise as a living language and its current status, see Pollock 2001.

tradition is the *Aṣṭādhyāyī*, written by the grammarian Pāṇini in around the fourth century BC; it is primarily on the basis of this work that Classical Sanskrit was codified. Due to the strong influence of the grammatical tradition and Pāṇini's grammar, Classical Sanskrit underwent relatively little linguistic change after the fourth century BC, at least in terms of phonology and morphology.

In terms of basic grammatical features, Sanskrit is a highly inflectional language with considerable freedom of word order, although the predominant order of major constituents is SOV.² Example 1 illustrates a simple Sanskrit sentence: the arguments of the verb are inflected for number and case (determined by the selectional properties of the verb), and also according to gender; the verb is inflected for tense, mood, and diathesis (voice), as well as for number and person in agreement with the subject.³

- (1) devadatto bhikṣava odanam adāt
 Devadatta.NOM.SG mendicant.DAT.SG porridge.ACC.SG give.AOR.IND.ACT.3SG
 'Devadatta gave porridge to the mendicant.'

In these respects its syntax is relatively similar to that of related old Indo-European languages such as Latin and Ancient Greek, and even to some modern languages like Russian. However, Classical Sanskrit is notably different from all of these in respect to its extremely free use of compounding.⁴

Compounding in Classical Sanskrit is so productive that almost any meaning that can be expressed using two or more separate words can also be expressed using a single compound, and it is so widespread that almost no Sanskrit sentence is encountered that does not contain at least one compound. For example, the sentence in 1 can be paraphrased as in 2, with the entire verb phrase reformulated as a single compound 'word'.⁵

- (2) devadatto [bhikṣu- datta- odanaḥ]
 Devadatta.NOM.SG mendicant- given- porridge.NOM.SG.M
 'Devadatta gave porridge to the mendicant.' (lit. 'Devadatta is (one by whom) porridge (was) given (to the) mendicant.')

As can be seen in this example, the last element in a compound inflects like an ordinary Sanskrit word, distinguishing number, gender, and case as required. Nonfinal elements in a compound, in contrast, appear in what is called 'stem' form, that is, without any inflectional ending. The 'stem' form of words (which is also the citation form) cannot otherwise be used without some kind of inflectional or derivational material following.

There is a large variety of compound types in Sanskrit, though many are relatively rare.⁶ In this article I discuss the most important subtypes of the five major compound

² That SOV is the most commonly attested constituent order in Sanskrit is clear enough from reading almost any prose (and many verse) texts; the point is made by, for example, Delbrück (1878), Speyer (1896), and Gonda (1952). That is not to say, however, that Sanskrit is an SOV language: its word order is nonconfigurational, or more accurately discourse-configurational, the order of constituents being determined by information structure rather than grammatical function. For two rather different accounts of Sanskrit word order that nevertheless share this intuition, see Gillon & Shaer 2005 and Lowe 2015a:37–46.

³ Abbreviations used in this article: 3PL: 3rd person plural, 3SG: 3rd person singular, ABL: ablative, ACC: accusative, ACT: active, ADV: adverb, AOR: aorist, CL: clitic, DAT: dative, DEF: definite, DU: dual, F: feminine, GEN: genitive, INS: instrumental, LOC: locative, M: masculine, NOM: nominative, PASS: passive, PL: plural, QUOT: quotative, RELPRO: relative pronoun, SG: singular.

⁴ For compounds in earlier stages of Sanskrit, see §11.

⁵ In this example, and throughout this article, I represent the constituent members of compounds separated, with the whole compound surrounded by square brackets. This example is my own, but its structure parallels attested examples; see, for example, Coulson 1992:122.

⁶ For an overview of compounding in Sanskrit, see, for example, Whitney 1896:424–56.

types.⁷ These are the *dvandva*, the *tatpuruṣa*, the *karmadhāraya*, the *bahuvrīhi*, and the *avyayībhāva*.⁸ In order to keep this article to a reasonable length, I do not treat the full variety of subtypes of these categories, but focus on the most common. It is possible to distinguish up to thirty different types of compound in Sanskrit, depending on how fine are the distinctions one wishes to make (cf. e.g. the slightly different set of distinctions made by Gillon (2009:101) and Tubb and Boose (2007:85–145)). Most of these are subtypes of *tatpuruṣa*, *karmadhāraya*, or *bahuvrīhi*. The types treated in this article are undoubtedly the most commonly attested, almost certainly constituting the majority of Sanskrit compound phrases, and are representative of the range of possibilities for compound expression in Sanskrit; as for the types not treated, all are amenable to analysis along the same lines as proposed below, and I hope to demonstrate this explicitly in future work.

Before moving on, I briefly introduce these main compound types that are the subject of analysis in this article. *Dvandvas* are essentially compounds of coordinate nouns (or compound noun phrases). For example, the noun *gaja-* ‘elephant’ and the noun *aśva-* ‘horse’ can form a *dvandva* [*gaja- aśva-*] ‘elephant(s) and horse(s)’.⁹ *Dvandvas* are the only regular type of compound for which there is (arguably) no requirement of binarity, and no head (or rather no nonheads); that is, any number of nouns can be compounded in parallel, just as there is no requirement for binarity in full phrasal coordination. For example, the noun *ratha-* ‘chariot’ can be compounded with the nouns for ‘elephant’ and ‘horse’ to create [*ratha- gaja- aśva-*] ‘chariot(s), elephant(s), and horse(s)’. In this compound, the only division is three ways; there is no sense in which [*ratha- gaja-*] or [*gaja- aśva-*] form subconstituents.

The term *tatpuruṣa* covers a wide range of compound types, but the canonical type (the ‘*vibhakti-tatpuruṣa*’) involves the implication of a case relation between the first element, which is invariably a noun, and the second element, which may be a noun or an adjective. For example [*svarga- patita-*], lit. ‘heaven-fallen’, can be analyzed in terms of an ablative case relation between the adjective and noun: ‘fallen from heaven’ (*svargāt patita-*, where *svargāt* is the ablative singular of *svarga-*). Likewise, [*rāja- bhāryā-*], lit. ‘king-wife’, can be analyzed in terms of a genitive case relation between the nouns: ‘the wife of the king’ (*rājño bhāryā-*, where *rājño* is genitive singular of *rājan-* ‘king’).¹⁰

The term *karmadhāraya* likewise applies to a number of distinct compound types. The most common types, which I analyze in this article, are: compounds of adjective + noun where the adjective modifies the noun, for example, [*priya- vayasya-*], lit. ‘dear-friend’ (= ‘dear friend’); compounds of adjective + adjective or adverb + adjective, where the first member modifies the second, for example, [*udagra- ramaṇīya-*], lit. ‘intensely-lovely’ (= ‘intensely lovely’); compounds of noun + noun where both nouns refer to the

⁷ How these types fit into modern classifications of compound types is not important for the present purposes; many authors essentially follow the classifications of the Sanskrit tradition (as e.g. Olsen 2000); for an alternative see Bisetto & Scalise 2005. For an introduction to the analysis of compounding within the Indian grammatical tradition, see Joshi 1974:i–lxix.

⁸ The Indian grammatical tradition treated *karmadhārayas* as a type of *tatpuruṣa*, but here, as is often done, I treat them as separate compound types for descriptive purposes (since there are a large number of subtypes of both *karmadhāraya* and non-*karmadhāraya* *tatpuruṣas*).

⁹ Rules of sandhi apply between members of compounds, so that, for example, [*gaja- aśva-*] always surfaces as *gajāśva-* (by coalescence of two adjacent short *a* vowels into one long *ā* vowel). Throughout this article I give compounds in presandhi form, so as to avoid obscuring the divisions between elements.

¹⁰ Formally, [*rāja- bhāryā-*] is ambiguous: given the right context, it could also be interpreted as a *dvandva* ‘king(s) and wife/wives’.

same entity or entities, for example, [*rāja- ṛṣi-*], lit. ‘king-seer’ (i.e. a seer who is also a king), or [*amātya- devadatta-*], lit. ‘minister-Devadatta’ (= ‘Devadatta the minister’).

Bahuvrīhis are often described as ‘exocentric’ compounds; essentially, a bahuvrīhi functions as a kind of reduced nominal clause, modifying an external element. So, the most common, and the simplest, type of bahuvrīhi is a two-part compound consisting of an adjective followed by a noun, as in the following example.

- (3) [dīrgha- karṇa-]
 long- ear-
 ‘long-eared’

Bahuvrīhis are functionally adjectival, attributing a property to an entity external to the compound. Like any adjectival construction in Sanskrit, bahuvrīhis can function as nouns in the absence of an explicit modificand. So, if the noun with which it agrees is absent, [*dīrgha- karṇa-*] can mean ‘the one with long ears’.

Bahuvrīhis are often translated as relative clauses, for example, ‘to/of whom the ears are long’. This analysis originates with the Pāṇinian grammatical tradition, where bahuvrīhis are explicitly glossed as relative clauses; for example, the nominative singular masculine of the compound in 3 is traditionally glossed as follows.

- (4) dīrghau karṇau yasya saḥ
 long.NOM.DU.M ear.NOM.DU who.GEN.SG.M he
 ‘he whose ears are long’

Avyayībhāvas are compounds involving a preposition and a governed noun, functionally equivalent to an adverbially used prepositional phrase. For example, [*bahir-grāma-*], lit. ‘outside-village’, which can be analyzed in terms of a prepositional phrase ‘outside the village’ (*bahir grāmāt*, where *grāmāt* is the ablative singular of *grāma-*, ablative being the case form required by the preposition *bahir*). When avyayībhāva compounds are embedded inside another compound, the final member naturally appears in stem form, but when not embedded the final member appears in an invariant adverbial form.

- (5) [abhi- agni] śalabhāḥ patanti
 to- fire.ADV moth.NOM.PL fly.3PL
 ‘Moths fly toward the fire.’

(*Kāśikā ad Aṣṭādhyāyī* 2.1.14)

The compound [*abhi- agni*] here is equivalent to the prepositional phrase *abhi agnim* ‘to fire.ACC’. The form *agni* is not a genuine case form of the noun *agni-* ‘fire’; the form corresponds to what would be expected for an accusative singular neuter, but the noun is masculine.¹¹ In [*bahir- ātmām*] ‘outside one’s self’, *ātmām* not only does not correspond to any possible case form of the masculine *ātman-* ‘self’, but does not even correspond to the theoretical accusative singular neuter.

All of these compound types are fully productive in Classical Sanskrit, in that in principle a compound of one of these types can be freely formed from any two words of the appropriate categories. In addition, the rules of compound formation are recursive, such that a compound of one type may be embedded within a compound of the same or another type, and the resulting compound may itself be further embedded in a compound structure, and so on.¹²

¹¹ This is also true for *grāmam*, although here the ‘accusative neuter’ is identical to the actual accusative singular (masculine) form of this noun.

¹² Examples of all the major compound types embedded inside larger compounds appear below, except for avyayībhāvas, but only because these are comparatively rare compared with the other categories. Examples of avyayībhāvas embedded in larger compounds are therefore less widespread, but still well attested and cer-

It is important to note that the arguments made in this article apply only to compounds that are (at least potentially) not listed, that is, that are (at least potentially) freely formed according to the morphosyntactic rules of the language, and are not stored as complete sequences in a speaker's mental lexicon. The compounding processes that are of interest here are fully productive, and I take full semantic compositionality to be a necessary feature of productively formed syntactic phrases and morphological words. Semantic noncompositionality implies listedness (though the converse implication does not necessarily hold); there do exist very many compounds in Classical Sanskrit with noncompositional meanings and that therefore must be treated as listed, but the existence of such compounds does not prejudice the status of nonlexicalized compounds.¹³ An example is *kṛṣṇa-sarpa-*, which in purely compositional terms would mean simply 'black snake', but which refers specifically to 'the black cobra'; it is therefore fully parallel to English *blackbird*. While such compounds can be analyzed, at least in principle, as formed according to the same compounding rules that I treat in this article, they are necessarily listed (whether as idiomatic phrases or lexemes) and so cannot be used as evidence for the status of the productive compounding rules under discussion. In addition, a finite number of forms that were analyzed as compounds in the Indian grammatical tradition are unambiguously lexicalized (and therefore listed), since there is no regular relation between the forms of the 'compounds' and the words from which they are supposedly formed. For example, *balāhaka-* 'cloud' is traditionally analyzed as a compound of *vāri-* 'water' and *vāhaka-* 'bearer' (Tubb & Boose 2007:122), but both words have to be assumed to appear in idiosyncratic forms that, even if a compound analysis is reasonable, necessitates the conclusion that the compound is listed. All such forms are disregarded in the rest of this article.

I now move on to discuss the evidence regarding the morphosyntactic status of these compound types in Classical Sanskrit.

3. EVIDENCE FOR SYNTACTIC STATUS. The most widespread analysis of compounding found in both morphological and syntactic literature is as a morphological or lexical—that is, nonsyntactic—process, though some authors have proposed strictly syntactic analyses of compound formation in some languages. Of course, what one means by a morphological, lexical, or syntactic process varies considerably depending on one's theoretical persuasion, and some morphological analyses are hardly different from syntactic analyses based on different theoretical assumptions. I discuss previous and alternative approaches to compounding in detail in §4 below. But at this point, it is important to note two things.

First, my perspective here is explicitly lexicalist, since it is really the lexicalist approach to the syntax–morphology divide that is most challenged by compounding phenomena. The lexicalist approach that I adopt assumes a strict modularity between syntax and morphology (hence the dichotomy between morphological and syntactic processes in the first sentence of this section). I use the term SYNTAX to refer to the component of grammar that deals with linear, functional, and in particular hierarchical relations between WORDS. Words are the minimal units of syntax and are stored in the LEXICON. I use the term MORPHOLOGY to refer to the component of the grammar that deals with the internal structure of words.

tainly no less grammatical, for example, [*bahir-grāma-pratiśraya-*], lit. 'outside-village-dwelling-' = 'whose dwelling is outside the village' (*Mānavadharmasāstra* 10.36).

¹³ Lexicalized compounds are recognized as a distinct category in the Indian grammatical tradition, under the heading of *nitya-samāsa* 'obligatory compounds', so they are relatively easy to isolate from nonlexicalized compounds.

Second, my aim in this article is not to argue that all processes of compound formation attested crosslinguistically should be treated syntactically, nor necessarily to claim that Sanskrit compounding, as a syntactic phenomenon, is fundamentally different from compounding processes in other languages. My aim is more restricted: I argue purely for the status of compounding in Classical Sanskrit. In §5, I present a means of analyzing Sanskrit compounding that has considerable potential for modeling the intermediate status of compounds, between syntax and lexicon, crosslinguistically, but I make no specific claims as to the applicability of my proposals to any particular compounding process in any other language.

There is no established set of criteria for distinguishing the minimal units of phrasal syntax from units of morphology or, to put it another way, for distinguishing words from parts of words, despite the fundamental importance of the distinction for lexicalist theories of syntax. Nevertheless, a number of criteria have been proposed, and are widely used, in discussions of this kind. In this section I discuss a number of criteria that are relevant to the status of productive compounding processes in Classical Sanskrit.¹⁴

3.1. 'ASAMARTHA' COMPOUNDING. *Asamartha* is the traditional term used to describe a construction in which a word external to a compound bears a syntactic relation to a word inside a compound, and not to the compound as a whole. This phenomenon is, strictly, not permitted according to the prescriptive Pāṇinian grammar (Tubb & Boose 2007:189–90), but is nevertheless well attested.¹⁵ A direct syntactic relation between a word outside a compound and an element embedded within a compound provides evidence that such compounds are syntactic phrases, at least from a lexicalist perspective. Perhaps the fundamental assumption of lexicalist syntax is that words can stand in syntactic relations to other words, but not to parts of words. As Lapointe (1980:8) defined his GENERALIZED LEXICALIST HYPOTHESIS, an early and particularly clear statement of what is more usually known as the STRONG LEXICALIST HYPOTHESIS: 'No syntactic rule can refer to elements of morphological structure'. The example in 6 is taken from Tubb & Boose 2007:189–90.¹⁶

- (6) jagato [[janma- ādi-]_{bv} kāraṇaṃ]_{tp} brahmādhigamyate
 world.GEN origin- etc.- cause brahman=understand.PASS.3SG
 'Brahman is understood to be the cause of the origin of the world, etc.'
 (*Brahmasūtrabhāṣya* 1.1.3)

Here, the genitive noun *jagataḥ* 'of the world' depends on the first element of the bahuvrīhi compound [*janma- ādi-*] 'origin etc.', which is itself embedded in a tatpuruṣa compound. The genitive cannot be construed with the tatpuruṣa, and cannot easily be construed with the bahuvrīhi. Likewise, in the following example the locative noun *pāṇḍaveṣu* 'among the Pāṇḍavas' is functionally dependent on the tatpuruṣa compound [*eka- puruṣa-*] 'one man', which is embedded in a tatpuruṣa, embedded within a bahuvrīhi. The locative cannot be construed with the bahuvrīhi, nor the outer tatpuruṣa, but only with the doubly embedded tatpuruṣa.

¹⁴ Haspelmath (2011) argues forcefully that none of the criteria used in discussions of wordhood, nor any combination of such criteria, are satisfactory for establishing a definition of a 'word' as a crosslinguistically valid concept. His arguments are in large part convincing, though I do not share his negative view on the status of 'word' as a crosslinguistic concept. In any case, my aim here is not crosslinguistic, but language-specific, and Haspelmath does not deny the possibility of words as language-specific concepts.

¹⁵ Gillon (1994, 2009) discusses *asamartha* compounding in more detail.

¹⁶ From hereon I show the internal structure of complex compounds in examples, and mark the type of compounds using the following subscript abbreviations: bv = bahuvrīhi, dd = dvandva, kd = karmadhāraya, tp = tatpuruṣa.

- (7) pāṇḍaveṣv [[[eka- puruṣa-]_{tp} vadha-]_{tp} artham]_{bv} amogham astram
 Pāṇḍava.LOC.PL one- man- killing- purpose unfailing weapon
 vimalā nāma śaktir
 Vimalā name spear
 ‘This spear, named Vimalā, is an unfailing weapon to be used for the purpose of killing one of the men among the Pāṇḍavas.’
 (Kāṇabhāra flg. vs. 23)

It is even possible for more than one element external to a compound to stand in relation to an element within the compound. Gillon (1994:120) provides the following example (translation his).

- (8) dṛḍham khalu tvayi [baddha- bhāvā]_{bv} ūrvaśī
 firmly indeed you.LOC fixed- affection Ūrvaśī
 ‘Indeed, Ūrvaśī is one whose affection is firmly fixed on you.’
 (Vikramorvaśīyam 2.134)

Here, both the adverb *dṛḍham* ‘firmly’ and the locative pronoun *tvayi* ‘in you’ are to be construed with *baddha-* ‘fixed’, the first element of the bahuvrīhi compound. The compound-external elements need not be individual words, but can be a syntactic phrase of more than one word. In the following example, the phrase *tat- abhāve sarvatra* ‘in every absence of it’ functions as an adjunct modifying the first element of the compound *abhāva- asiddeḥ* (following the interpretation of Gillon 1994:131, whence the example).

- (9) apratibaddhasya [tat- abhāve]_{tp} sarvatra [abhāva-
 unconnected.GEN that- absence.LOC everywhere absence-
 asiddeḥ]_{tp}
 nonestablishment.ABL
 ‘due to the nonestablishment of the absence of a thing that is unconnected
 in every absence of it’
 (Pramāṇavārttikasavṛtti 12.23)

Gillon (1994) shows that a wide range of syntactic relations are possible between compounded words and noncompounded words/phrases, all of which are also possible between noncompounded words. He also shows that *asamartha* compounding is no less frequent than regular syntactic constructions such as indirect questions or relative clauses; it must therefore be treated as a productive part of the grammar of Classical Sanskrit, even though it is not permitted by the prescriptive grammar.

In lexicalist approaches to grammar, words are the minimal units of syntax; it should therefore be impossible for syntactic relations to hold between subparts of morphologically formed compounds and words external to the compound. The fact that in Classical Sanskrit such relations are relatively common provides strong evidence for the syntactic status of Sanskrit compounding processes, at least within an approach that seeks to maintain the fundamental assumptions of lexicalism. Words can also stand in semantic, anaphoric, relations to parts of compounds, as discussed in the next section.

3.2. ANAPHORIC (NON)ISLANDHOOD. It is usually assumed, following Postal (1969), that words are anaphoric islands, that is, that it is not possible to refer anaphorically to parts of a word (‘outbound’ anaphora). In fact, constraints on outbound anaphora are primarily pragmatic and are not due to ungrammaticality (Ward et al. 1991). ‘Inbound’ anaphora, that is, anaphoric reference from part of a word to an element outside that word, is common with inflectional affixes, but crosslinguistically rare as part of compounds or derivational formations (Haspelmath 2011:51). Again, constraints on inbound anaphora may be more pragmatic than grammatical and may differ across languages; Harris (2006) shows that inbound anaphora does occur in Georgian.

Nevertheless, anaphora of both kinds are at the least highly constrained in most languages, and this is true also for unambiguously lexicalized compounds and derivational

formations in Sanskrit. Crucially, however, there are no constraints on inbound or outbound anaphora in relation to productively formed compound structures in Classical Sanskrit: anaphoric reference into, out of, and even within compounds functions entirely parallel to anaphoric reference within and between noncompound phrases and clauses.

Inbound anaphora is particularly common. All Sanskrit pronouns have forms for use in compounds, including the demonstrative and anaphoric pronouns, which usually refer back to elements outside of the compound in which they appear. The compound form *tat* of the anaphoric third-person pronoun is particularly common.¹⁷

- (10) *naivaṃ vākyaṇi, dṛśya-viśeṣa-tvāt,*
 not=thus sentences_i observable-distinct.features-ness.ABL
adrśyatve 'py [adrṣṭa- viśeṣāṇām]_{bv}
 unobservableness.LOC even unobserved- distinct.features.GEN
[[vijāyīyatva- upagama-]_{tp} virodhāt]_{tp}, [tat-
 heterogeneity- accepting- contradiction.ABL them_i-
viśeṣāṇām]_{tp} anyatrāpi śakya- kriyatvāt ...
 distinct.features.GEN elsewhere=too possible- creation.ABL
 ‘Sentences_i are not like this, because their distinct features are observable, and even if they were unobservable, due to the contradiction of accepting heterogeneity of things with unobserved distinct features (from those without), and because of the possibility of the creation of their_i distinct features elsewhere too ...’
(Pramāṇavārttikasvavṛtti 16.1)

Here the first element of the tatpuruṣa compound [*tat- viśeṣāṇām*] ‘their distinct features’ refers back to the noun *vākyaṇi* ‘sentences’. Likewise, in example 9, repeated as 11, the first element of the compound [*tat- abhāve*] ‘in its absence’ refers back to the preceding word, which is not part of the compound.

- (11) *apratibaddhasya [tat- abhāve]_{tp} sarvatra [abhāva-*
 unconnected.GEN_i that_i- absence.LOC everywhere absence-
asiddeḥ]_{tp}
 nonestablishment.ABL
 ‘due to the nonestablishment of the absence of a thing_i, that is unconnected in every absence of it_i’
(Pramāṇavārttikasvavṛtti 12.23)

Example 12 illustrates simultaneous inbound and outbound anaphora. The first element of the tatpuruṣa compound [*tat- anyena*] ‘one different from it’ refers back to the first constituent part, [*eka- dharmā-*] (itself a compound), of the preceding compound [[*eka- dharmā-*] *sadbhāvāt*].¹⁸

- (12) *[[eka- dharmā-]_{tp} sadbhāvāt]_{tp} [tat- anyena]_{tp} api bhavitavyam iti*
 [one- property-_i existence.ABL that_i- different.INS too must.exist QUOT
[niyama- abhāvāt]_{tp}
 constraint- absence.ABL
 ‘due to the absence of a constraint that (just) because one property_i exists, one different from it_i must exist also’
(Pramāṇavārttikasvavṛtti 17.23)

These anaphoric possibilities are usually prohibited in compounding, even in languages with productive compounding patterns such as English. The equivalent of 12 in English,

¹⁷ The first three examples in this section are taken from Gillon 1994; Gillon is to my knowledge the first to have discussed the anaphoric possibilities of Sanskrit compounds, and the relevance of these possibilities for their morphosyntactic analysis.

¹⁸ In this example, we also have *asamartha* compounding: the clause headed by the quotative particle *iti* (everything from *eka-* to *iti*) is dependent on *niyama-*, which is part of the larger compound [*niyama-abhāvāt*].

were it possible, would be a noun phrase like **a beer drinker and a wine one*, where *wine one* were itself a compound just like *beer drinker*. It is even possible for pronouns such as *tat* to refer to another element within the same compound. This is a regular strategy for disambiguating otherwise ambiguous compounds, for example, where a dvandva might be mistaken for a tatpuruṣa. Example 13, from Tubb & Boose 2007:191, is a compound consisting of a dvandva of two tatpuruṣas. The first element of the second tatpuruṣa refers anaphorically back to the first element of the first tatpuruṣa.

- (13) [[adhyāsa- svarūpa-]_{tp} [tat- sambhāvanāya]_{tp}]_{dd}
 superimposition_i- nature- it_i- possibility.DAT
 ‘for (the sake of) the nature of superimposition_i and the possibility of it_i’
 (Pañcapādikā p. 33)

We therefore have simultaneous inbound and outbound anaphora again, but this time all inside the same compound. The compound in example 13 is essentially a disambiguated version of the same compound without the pronoun, which could be interpreted in two ways.

- (14) a. [adhyāsa- [svarūpa- sambhāvanāya]_{dd}]_{tp}
 superimposition- nature- possibility.DAT
 ‘for (the sake of) the nature, and the possibility, of superimposition’
 b. [[adhyāsa- svarūpa-]_{tp} sambhāvanāya]_{tp}
 superimposition- nature- possibility.DAT
 ‘for (the sake of) the possibility of the nature of superimposition’

The construction in example 13 thus serves as a way of enforcing the interpretation in example 14a. Despite its specifically disambiguating function, example 13 is entirely regular and formed according to the same productive rules as all of the other compounds treated in this article. I know of no morphological parallels that might support treating the sort of internal reference in example 13 as a word-internal, rather than phrase-internal, phenomenon.

Outbound anaphora can also involve relative pronouns, which like other pronouns can appear within Classical Sanskrit compounds. Correlative structures are common in Sanskrit; Tubb and Boose (2007:192) provide an example in which a compounded relative pronoun is referred back to by an uncompounded correlative.

- (15) [yad- viṣayā]_{bv} buddhir na vyabharati tat sat
 RELPRO_i- object.NOM cognition.NOM not be.in.error.3SG that_i real
 ‘A thing_i which, when a cognition_j that has it_i as its_j object is not in error,
 that thing_i is real.’ (lit. ‘A cognition having-which_i-as-its-object is not in
 error, that_i is real.’)
 (Bhagavadgītābhāṣya 2.16)

Inbound anaphora involving correlative structures is also possible. In the following example, the first element of the compound [*tadā- prabhṛti*] ‘since then’ functions as the correlative to the preceding relative adverb *yadā* ‘when’.

- (16) atas tatrābhavān avimārakaḥ ... yadā [hasti- sambhrama- divase]_{tp}
 so his.highness Avimāraka when_i elephant- disturbance- day.LOC
 [kuntibhoja- duhitā]_{tp} dṛṣṭā, [tadā- prabhṛty]_{tp} anyādṛśa iva
 Kuntibhoja- daughter seen then_i- since like.another like
 saṃvṛttaḥ
 become
 ‘So his highness Avimāraka ... has become like another person since he
 saw the daughter of Kuntibhoja on the day of the disturbance with the
 elephant.’
 (Avimāraka II)

Relatives and demonstratives are not the only types of pronouns that can appear in compounds; there are no restrictions. For example, interrogative pronouns can appear inside ‘compounds’, and they retain clausal scope, marking the whole clause as a question.

- (17) [kim- lakṣaṇam]_{bv} punas tad brahma
 what- definition but that brahman
 ‘But what is the definition of that brahman?’ (lit. ‘But having-what-as-its-
 definition is that brahman?’) (*Brahmasūtrabhāṣya* 1.1.3)

The anaphoric and syntactic possibilities found with Sanskrit compounds therefore go well beyond the standard possibilities of morphological or lexical units. But they are precisely what we would expect of syntactic structures and provide the strongest evidence that the productive rules of compounding in Sanskrit should be treated as fundamentally syntactic rules.

3.3. LENGTH. The features discussed in the previous sections provide strong positive evidence for the syntactic status of Classical Sanskrit compounds. There are a variety of other features that cannot be taken as providing such strong positive evidence, but that are at least consistent with a syntactic analysis.

The first of these is a feature not usually discussed in relation to the status of compounds, but that is, I believe, relevant to the status of Classical Sanskrit compounds. Famously, Classical Sanskrit compounds can be extremely long. According to a number of popular authorities, including Guinness World Records,¹⁹ the longest ‘word’ ever attested in any language is a Sanskrit compound found in the *Varadāmbikā Pariṇaya Campū*, a literary work of the sixteenth century by Tirumalāmbā, a poet of the Vijayanagara empire of Southern India. The compound is given in example 18.²⁰

- (18) [nirantarā- andhakāritā- digantara- kandalad- amanda- sudhārasa- bindu-
 constantly- made.dark- quarters- sprouts- abundant- nectar- drop-
 sāndratara- ghanāghana- vṛnda- sandeha- kara- syandamāna-
 dense- thick.cloud- mass- delusion-making-trickling-
 makaranda- bindu- bandhuratara- mākanda- taru- kula- talpa-
 nectar- drop- more.charming- mango- tree- cluster-couch-
 kalpa- mṛdula- sikatā- jāla- jaṭila- mūla- tala-
 equivalent.to- soft- sand- lattice crested.with- foot- base-
 maruvaka- milad- alaghu- laghulaya- kalita- ramaṇīya- pānīya- śālikā-
 marjoram- mixing- thick- khas.root- made- pleasant water- shed-
 bālikā- kara- āravinda- galantikā- galad- elā- lavaṅga-
 maiden hand- lotus- pitcher- dripping- cardamom- clove
 pāṭala- ghanasāra- kastūrikā- atisaurabha- medura- laghutara-
 saffron- camphor- musk- excess.fragrance- thick.with- light-
 madhura- śītalatara- salila- dhārā- nirākariṣṇu- tadīya- vimala-
 sweet- cold- water- stream- shaming- their- bright-
 vilocana- mayūkha- rekhā- apasārita- pipāsā- āyāsa- pathika- lokān]
 eye- ray- series- alleviated- thirst- weariness- traveler- people
 ‘It was a place where travelers’ weariness due to thirst was alleviated by
 series of rays of the bright eyes of the girls, the rays that were shaming

¹⁹ See <http://www.guinnessworldrecords.com/world-records/longest-word> and http://en.wikipedia.org/wiki/Longest_words, but compare also http://en.wikipedia.org/wiki/Longest_word_in_English, particularly in relation to the naming of organic chemical compounds. (All weblinks accessed 10 April 2015.)

²⁰ The division of the compound into its constituent parts is correct here, unlike on the webpages referenced in n. 19.

the streams of light, sweet and cold water thick with the strong fragrance of cardamom, clove, saffron, camphor, and musk and flowing out of the pitchers in the lotus-like hands of maidens (seated) in the beautiful water-sheds, which were made of the thick roots of khas grass mixed with marjoram, (which were sited) at the foot, covered with heaps of couch-like soft sand, of the clusters of mango trees, which looked all the more charming on account of the trickling drops of nectar and caused the delusion of a mass of dense rain clouds, dense with drops of abundant nectar from the (new) sprouts, which constantly darkened the quarters of the sky.’

(op. cit., *Tuṅḍiradeśavarṇana*; text and translation based on Suryakanta 1970:18–19)

Guinness World Records classifies this as the longest ‘word’ ever recorded on the basis of the number of characters in the native devanagari script (195), while the Wikipedia page makes reference to the number of letters in the English transliteration (c. 430). Clearly, either is a poor basis on which to compare word length from different languages, but nevertheless it is worth noting that, according to the same popular authorities, the longest word in a language other than Classical Sanskrit is less than half the length by the same criteria. There is an invented compound in an Ancient Greek comedy by Aristophanes, which comes in at 173 ‘letters’ in English transliteration. The longest word listed for a modern language also has 173 letters, a compound involving numerals in Polish. A more reasonable measure of comparison might be number of syllables: the Sanskrit compound has 194 syllables, compared with seventy-nine in the Greek word, and eighty-six in the Polish. In terms of number of members (since all are compounds), the Sanskrit compound has sixty-three members, the Greek twenty-four, and the Polish nineteen.²¹

It is undeniable that compounds of even half the length of that seen in 18 are artificial, in the sense that they are high literary constructions, coined within a literary tradition that reveled in complexity, ambiguity, and interpretative opacity. Nevertheless, such compounds are formed according to the regular rules of compounding and are not, strictly speaking, ungrammatical. Artificiality is also a feature of the very long compounds in Ancient Greek and Polish mentioned above. But all are formed according to the regular recursive application of rules for compound formation. The relevant difference between Sanskrit compounds and those in Ancient Greek, Polish, or other languages is length: as stated, the longest attested Sanskrit compound is more than twice the length of the longest compound known from any other language.

Although it is often said that some languages in principle allow compounds of entirely unrestricted, potentially infinite length, in reality the longest compounds attested in such languages tend to be considerably shorter than the longest compounds attested in Sanskrit. Ørsnes (1996) discusses compounding in Danish, a language that theoretically admits infinite compounding, and shows that compound length is restricted by certain factors. In all such languages, of course, one must first of all establish whether ‘compound’ processes that permit potentially infinite sequences are lexical or syntactic. Likewise, some polysynthetic languages in principle permit words of infinite length, for example, by recursive noun incorporation, but first, words as long as the Sanskrit compound in example 18 do not in practice occur, and second, one could equally argue over

²¹ The best comparison of ‘length’ might be by number of morphemes, but determining the number of morphemes in a word depends on theory-specific attitudes toward questions concerning the divisibility of irregular stems and the existence of null morphemes, which would render such comparison problematic.

compounds simply involve the recursive application of regular rules, and in this sense are not any less grammatical than shorter compounds. An interesting parallel is the existence of word games in some polysynthetic languages, in which the aim is to coin the longest possible ‘word’ using the regular rules of the language. One might equally compare the existence of extremely long sentences in high literary and academic English, much longer than tend to occur in the spoken language, but no less grammatical.

3.4. MORPHOPHONOLOGICAL REGULARITY. A more widely used criterion for distinguishing words from phrases is that morphophonological idiosyncrasies are a feature of the combination of stems and affixes into words, but not of words with other words. This is one of the criteria proposed by Zwicky and Pullum (1983) for distinguishing clitic sequences from morpheme sequences. Although this generalization, like all the others, cannot be considered exceptionless (Haspelmath 2011:52–54), it is still important that the productive rules of compounding in Classical Sanskrit involve no morphophonological idiosyncrasies.²²

One feature of Sanskrit compounding that might be considered an ‘idiosyncrasy’ is the fact that, as noted above, all but the last element of a compound appears in the so-called ‘stem’ form, that is (in the case of nouns and adjectives), without the inflectional case/number/gender marking that is obligatory for noncompound forms of the word. This is idiosyncratic to the compound context, but is not lexically idiosyncratic; it is specifically the latter that Zwicky and Pullum’s (1983) criteria refer to. In fact, it is unproblematic to account for the stem forms found in compounds by treating them as specialized forms specified for use in particular syntactic contexts (i.e. in ‘compound’ syntax); the formal model of compound syntax advanced below provides a good account of this. Granted the use of the ‘stem’ form in Sanskrit compounds, it is significant that in phonological terms the juncture between elements of a compound is resolved according to the same rules of sandhi that apply between independent words (‘external’ sandhi), and not according to the rules that generally apply between stems and affixes (‘internal’ sandhi). Again, it may not be possible to treat this as positive evidence for the syntactic status of Sanskrit compounds, but it is at least consistent with such an analysis.²³

Theoretically, accent could be argued to provide evidence for the single-word status of Sanskrit compounds. This is because in general each independent word is assigned a single accent, and the rules for compound accentuation provide for only a single accent for a compound of any length. However, the rules of accentuation, as specified for example by Pāṇini, are based on the accent of the late Vedic period, when compounding was arguably (more) lexical (see §11). Importantly, the pitch accent described by Pāṇini was lost in the immediately following centuries, so that for the vast majority of the Classical Sanskrit period it is merely theoretical. Moreover, even in the Vedic period there was no one-to-one correlation between one morphosyntactic word and one pitch accent: some morphosyntactic words had no accent (mainly clitic particles, but also, for example, finite verbs in some contexts), while some incontestably lexical compounds had two accents. In fact, the Vedic pitch accent is most accurately analyzed as a feature

²² Unsurprisingly, some lexicalized compounds do show morphophonological irregularities, as discussed in §2, but these are irrelevant to the question of the productive rules of compounding.

²³ In fact, internal sandhi applies to only a subset of what are usually treated as word-internal junctures, so sandhi can really only provide positive evidence for the morphological status of a particular sequence. Under ‘internal’ sandhi rules I do not include the rules involving retroflexion of /s/ and /n/. These apply inside single words, but can also, optionally, apply between compound members. These are not significant for the status of compounds, however, since they are primarily a phonological-word phenomenon; they can apply, for example, inside clitic sequences (cf. Lowe 2014:20–23 with references).

of phonological words; that is, a phonological word in Vedic is characterized by (among other things) association with a single pitch accent; it is only usually, and not necessarily, the case that one phonological word corresponds to one morphosyntactic word.

Overall, the evidence of morphophonology is at least consistent with a syntactic analysis of Sanskrit compounds, but it cannot be considered to directly support it, not least because it is almost always possible to argue that morphophonological phenomena apply to phonological, not morphosyntactic, words.

3.5. SECONDARY DERIVATION. The evidence I have presented above either speaks in favor of, or is at least consistent with, a syntactic analysis of Sanskrit compounding. One piece of evidence that might be taken to favor a morphological or lexical status for Sanskrit compounds is, however, the possibility of secondary derivation. It is in principle possible for a compound of any length to be the input to what are usually assumed to be secondary morphological processes such as derivational affixation. This is not restricted to lexicalized compounds, or even to frequent, easily lexicalizable compounds, but is a productive process that can at least theoretically apply to any newly coined compound.

In principle, any derivational process may be applied to any productively formed compound of the relevant grammatical category. A few morphemes are used particularly frequently to derive words from productively formed compounds.²⁴ For example, one affix commonly attached to compounds is the suffix *-tva-*, which forms abstract nouns from adjectival categories. So the bahuvrīhi [*dīrgha-karṇa-*] ‘long-eared’ (example 3) could be the basis of a word *dīrgha-karṇa-tva-* ‘long-eared-ness’, and from the karmadhāraya [*udagra-ramaṇīya-*] ‘intensely lovely’ an abstract noun *udagra-ramaṇīya-tva-* ‘intense loveliness’ could be formed.

Another such suffix is *-ka-*. This suffix is commonly attached to bahuvrīhis, either to disambiguate them from Adj + N karmadhārayas, or sometimes to simplify the inflection by turning the compound into an *a*-stem.²⁵ In these uses *-ka-* is essentially semantically empty, and it does not affect the category of the compound to which it attaches.²⁶ For example, *dīrgha-karṇa-ka-* means ‘long-eared’ just as the bahuvrīhi [*dīrgha-karṇa-*] does, but the latter, and not the former, could equally be interpreted as a karmadhāraya meaning ‘long ear’. The attachment of suffixes such as *-tva-* to compounds of more than a few members is relatively rare, but in principle it is unrestricted, and any rarity may simply be a combination of the fact that longer compounds are rarer than shorter ones, and secondary derivation from compounds is itself less common than its absence. Elements such as *-tva-* and *-ka-*, which can attach to productively formed compounds, are not specialized compound affixes but can also attach to noncompound

²⁴ It is possible that, in practice, there are constraints on the application of some, or even many, derivational processes to productively formed compounds, but to my knowledge no study of the question exists. One difficulty in assessing how freely productively formed compounds could be input to derivational processes is that one must first distinguish listed compounds from those that are not; assuming that listed compounds are generally lexicalized, their being subject to secondary derivation is not surprising. For example, a referee quotes the form *aikāntika-* ‘absolute, complete’ as an example of a somewhat complex derivational process affecting the compound [*eka-anta-*]_{IP}, but this is precisely an example of a listed compound: its meaning (at least the meaning that is relevant here) is ‘absoluteness, completeness’, not the compositionally regular ‘one end’, which we would expect if the compound were not listed. For the present purposes I assume that in principle any derivational process could be applied to any productively formed compound, but I discuss only those common derivational processes for which examples are easily found.

²⁵ *a*-stems are the most common inflectional class in the Sanskrit nominal system, and forms for all genders exist, which is not the case for all other inflectional classes.

²⁶ The use of *-ka-* as a semantically null bahuvrīhi suffix is briefly discussed by Gillon (1991, 2007), and compare also my analysis of bahuvrīhi phrase structure below.

stems. Traditionally, these elements are treated as derivational morphemes, which is why I refer to them as affixes. However, there is relatively little evidence against treating affixes such as *-tva-* and *-ka-* as independent lexical or functional words, clitics perhaps, rather than morphemes, at least when used at the end of compounds. This would imply that these elements, at least in compounding, have undergone a degrammaticalization from affixes to words/clitics, since they were uncontroversially affixes at an earlier period, and may still be affixes when attached to noncompounded words. What little evidence there is may favor the clitic analysis: all of the evidence discussed above regarding the syntactic status of Sanskrit compounds applies equally to compounds that have undergone further derivation by one of these productively used ‘affixes’. For example, the anaphoric possibilities into and out of compounds are not affected by secondary derivation (at least with the most common types of secondary derivation, that is, those for which sufficient data is available). Likewise, *asamartha* compounding is found with secondary derivatives from compounds, as in example 21.²⁷

- (21) [sādhya- abhāve]_{tp} [asattva- vacana-]_{tp}-vat
 to.be.established- absence.LOC nonpresence- statement-like.ADV
 ‘like the statement of (its) nonpresence in the absence of what is to be
 established’
(Pramāṇavārttikasvavṛtti 2.4)

Here the adverb-forming suffix *-vat*, which indicates similarity, is attached to the compound [asattva- vacana-] ‘statement of nonpresence’; the preceding compound, [sādhya- abhāve], is dependent on *asattva*.

Example 21 shows unambiguously that syntactic dependencies may exist between elements embedded within a compound that has undergone ‘derivation by affixation’ and elements external to the compound. There are two possibilities for the analysis of these ‘affixes’ and the compounds to which they are attached. Given the evidence for the syntactic status of Classical Sanskrit compounding, in order to treat these elements as affixes we would have to admit that syntactically formed compound ‘phrases’ could be productively lexicalized and input to further derivation, but then the problem would be accounting for the evidence for syntactic status of these derivatives. Alternatively, we could treat these ‘suffixes’ as independent syntactic elements, perhaps clitics, attaching to the right edge of compound ‘phrases’. In the latter case, a ‘derived’ abstract such as *asattva-vacana-vat* would be no less a syntactic phrase than the *bahuvrīhi* from which it is formed.

There is some evidence in favor of this second analysis, though it cannot be considered conclusive. In a few compounds a ‘derivational suffix’ must be interpreted as applying separately to two or more members of the compound. For example, the noun *pada-ka-* means ‘an adept in the *pada* (mode of Vedic recitation)’, and the noun *krama-ka-* means ‘an adept in the *krama* (mode of Vedic recitation)’; the dvandva corresponding to a compound of these two words is not, as we might have expected, [*pada-ka- krama-ka-*], but rather [*pada- krama- -ka-*] ‘an adept in the *pada* (mode of Vedic recitation) and an adept in the *krama* (mode of Vedic recitation)’. This does not mean ‘someone/people who is/are adept in both the *pada* and *krama*’, which is what we would expect if this were formed by affixation of *-ka-* to a dvandva [*pada- krama-*]; rather, it refers to two distinct people (or sets of people), one of whom is a ‘*pada-ka-*’ and one of whom is a ‘*krama-ka-*’. Therefore it can only be understood by taking the ‘suffix’ *-ka-* with both *pada-* and *krama-* separately.²⁸

²⁷ This example is taken from Gillon 1994:126–33; Gillon provides a number of other examples that he analyzes in this way, but they are less clear.

²⁸ *pada-* alone cannot mean ‘an adept in the *pada*’.

A clitic analysis is only one possibility for analyzing such a construction; it is reminiscent of ‘suspended affixation’, which is often analyzed with recourse to affix ellipsis (e.g. Erschler 2012) or, within the grammatical framework assumed in this article, more commonly with recourse to the theory of LEXICAL SHARING (e.g. Broadwell 2008, Belyaev 2014), which I discuss and make use of in my analysis of bahuvrīhi and avyayībhāva compounds below (§§9–10). In addition, it might be possible, at least in this case, to assume that we are dealing with a lexicalized, and therefore irregular, form, such that this could not be used as evidence for the nonaffixal status of *-ka-*. But then it would be difficult to prove that any relevant sequence were not lexicalized.

Taking the evidence as uncertain, I therefore make no definite claim either way as to the status of the ‘affixes’ such as *-ka-* that can be productively attached to productively formed compounds. Within the strictly lexicalist theory in which my analysis is couched, a clitic analysis would be more consistent, but it would be possible to assume productive lexicalization of compound sequences in order to admit an affixal analysis.²⁹

In any case, what the phenomenon may, at least, demonstrate is that Sanskrit compounding is closer to a lexical process than other unambiguously syntactic processes. That is, while I have argued that Classical Sanskrit compounding is fundamentally syntactic, it is in some respects undeniably less syntactic than noncompound syntax. At least superficially, the same derivational processes that can apply to single words can also apply to compound sequences, whether this is the result of lexicalization of the compound, or degrammaticalization of the affix (or in some cases one, in some cases the other). A fully descriptive formal account of Sanskrit compounding should take account of this somewhat intermediate status of compounds, and indeed the analysis I propose below does explicitly capture the less than fully syntactic status of these compounds.

3.6. SUMMARY. Classical Sanskrit compounds show a variety of properties that speak for, or are consistent with, a syntactic analysis, while the only one that may support a lexical analysis, secondary derivation, is rather unclear. It is worth remarking that none of the data discussed here is unique to Sanskrit compounds, and much of it is not even necessarily unusual in crosslinguistic terms. Perhaps the closest parallels with the Sanskrit data discussed here are found in incorporating languages. The data discussed by Sadock (1980, 1986) from Greenlandic Eskimo, for example, show some of the same anaphoric possibilities and the equivalent of the Sanskrit *asamartha* phenomena.³⁰ Lexical accounts of noun incorporation are found, of course (e.g. Mithun 1984, Malouf 1999), and a lexical account of Sanskrit compounding would no doubt also be possible. But the fact is that within strictly lexicalist approaches to syntax, a lexical account of problematic phenomena is always possible, to the extent that assigning a particular phenomenon to the lexicon has little explanatory value. I take the data presented above as sufficient evidence that a syntactic analysis of Classical Sanskrit compounding should be sought, even within a lexicalist framework.

²⁹ Many authors working within a strictly lexicalist theory of syntax assume the possibility of productive lexicalization, that is, the possibility that any syntactic sequence could be innovatively treated as a lexical sequence and input to morphological processes (e.g. Bresnan & Mchombo 1995). It could be objected that such an assumption is a somewhat stipulative attempt to preserve lexicalism in the face of clear counterevidence, and it does introduce a degree of circularity into the definition of the syntax–morphology divide. But at least for this case, the analysis of Sanskrit compound syntax proposed below could provide some rationale for such an explanation, if it were needed, by recognizing the close relation between compound ‘phrases’ and single words.

³⁰ I use the term ‘incorporating language’ in the loosest possible sense; it is relatively controversial whether Greenlandic Eskimo truly attests noun incorporation.

This is what I aim to do in §5; I preface this, in the next section, with a discussion of previous approaches to compounding phenomena in the morphological and syntactic literature.

4. APPROACHES TO COMPOUNDING. A range of theoretical treatments of compounding have been proposed, varying according to the particular language and phenomena under consideration, and even more according to the particular theoretical concerns of the authors. In this section I provide a brief discussion of this range.

Relatively little work exists on the theoretical treatment of compounding in Sanskrit itself, and the status of Sanskrit compounding as syntactic, morphological, or lexical has never been argued for in detail. The most important work on Sanskrit compounds has been done by Gillon (1991, 1994, 1995, 2007, 2009). Gillon's analyses are broadly couched within the framework proposed by authors such as Selkirk (1982) and Di Sciullo and Williams (1987), in which the structural similarity between syntactic and morphological structure is emphasized by treating derivational morphological processes, and also compounding, by means of 'lexical syntactic' context-free rules. Within such a framework, compounding is usually assumed to be a word-level process, but the rules by which compounds are formed are relatively close to the sorts of rules used for phrasal syntax. There also exists some work on the computational processing of Sanskrit compounds, for example, Kumar et al. 2009, but this has no firm theoretical basis.

Most work on the status of compounds has, of course, been based on English, in particular the productive N + N compounding in English. A valuable and balanced discussion of these is provided by Payne and Huddleston (2002:448–51); other valuable discussions on the status of compounds are by Bisetto and Scalise (1999), who discuss compounding in Italian, and ten Hacken (1994), who discusses compounding in Dutch.

It seems that all of the possible approaches to compounding are attested in recent literature. One possibility is to treat even productive compounding processes as fundamentally morphological or lexical (depending on one's view of the status of morphology vis-à-vis the lexicon); this is the usual approach in strictly lexicalist syntax, as discussed below, and is also essentially the approach of Ackema and Neeleman (2004) and Booij (2005, 2009), from a morphological perspective. Another approach is to treat at least some productive compounding processes as based in the syntax, while confining others to the lexicon. This is the position taken by Baker (1988) in regard to compounding/incorporation phenomena in polysynthetic languages, and following him (but extending to English) Snyder (2001). Similarly, Anderson (1992) assumes that the 'word structure rules' that are used to form compounds can apply in the lexicon or syntax, though he leans toward a more syntactic treatment. Di Sciullo and Williams (1987) take the somewhat extreme position that all exocentric compounds in English are formed in the phrasal syntax, a position that was comprehensively criticized by Anderson (1992:306–18). Other authors, for example, Lieber and Štekauer (2009a) and Böer and colleagues (2011), note the difficulties of providing an absolute categorization in either direction. These difficulties lead some authors to abandon the notion of a strict distinction between syntax and morphology (or syntax and the lexicon); this is the position taken by Giegerich (2004, 2005, 2009), for example, based on detailed analysis of English compounding; Haspelmath (2011) provides a comprehensive criticism of most tests used to distinguish syntax from morphology, and argues strongly that there is no good evidence for such a distinction crosslinguistically. In some theoretical frameworks, where no real distinction is made between syntax and morphology, the status of compounding is something of a moot point (cf. Harley 2009 on compounding in DISTRIBUTED MORPHOLOGY). The details of all these approaches are relatively theory-

specific, and do not necessarily transfer easily into the framework adopted here. What they show is the considerable variety of standpoints that can be taken on this question. In the following section, I discuss some explicit similarities between two previous approaches and my own proposals.

In this article, I take a strict lexicalist approach to the question of compound formation, in contrast to all of the authors mentioned above, and present an analysis based within the framework of LEXICAL-FUNCTIONAL GRAMMAR (LFG; Kaplan & Bresnan 1982, Bresnan 2001, Dalrymple 2001, Falk 2001). On a nonlexicalist approach to syntax, of course, the evidence presented above would mean relatively little, and a number of different analyses of Sanskrit compounds would be possible. In strict lexicalist syntax, by contrast, the analysis of compounding has been relatively neglected, and where it is treated a morphological or lexical analysis is usually assumed.³¹ The evidence presented above shows, however, that compounding in Sanskrit is a fundamentally syntactic process: it involves productive, structure-generating rules that interact with other syntactic processes in a way that could not easily be captured with a lexical analysis. The analysis proposed below shows that it is possible to analyze compounding as a syntactic process, even within a strictly lexicalist theory, while at the same time taking account of the differences between ‘ordinary’ syntax and the apparently more lexical syntax of compounds.

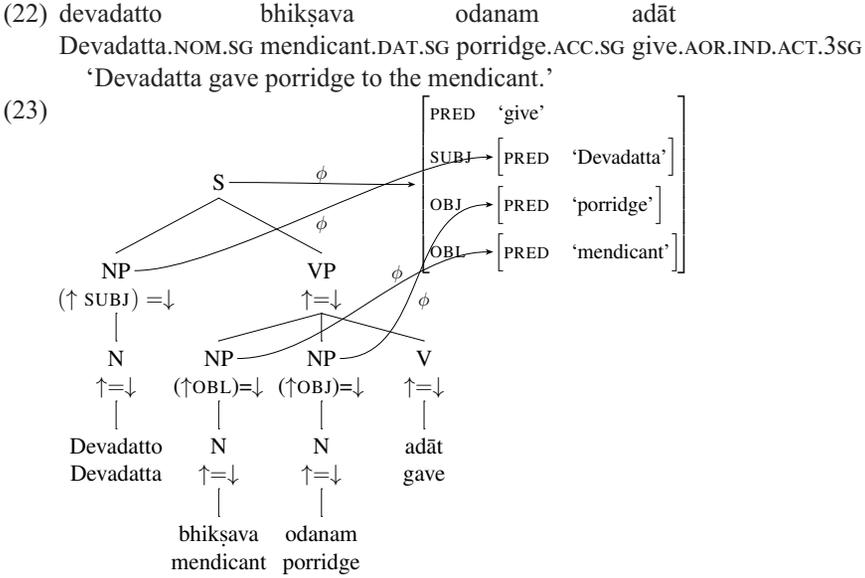
5. COMPOUNDS AND ‘NONPROJECTING’ CATEGORIES. I have argued that Classical Sanskrit ‘compounding’ should be treated as a fundamentally syntactic process and should be analyzed in syntactic terms. But it is clear enough that the syntax of Sanskrit ‘compounds’ is very different from the rules of syntax that apply at the larger sentence level, that is, between fully inflected words (including inflected compounds). In terms of sentential syntax, Classical Sanskrit permits a relatively large degree of freedom in constituent order, and also permits discontinuous constituents. Sanskrit is a highly inflected language, and extensive case marking and agreement mean that grammatical relations are usually clear regardless of the word order.

If we treat compounding as syntactic, its features are in stark contrast to the usual rules of Sanskrit syntax. There is no morphological marking of the relations between elements of a ‘compound’, since all but the final element appear in uninflected ‘stem’ form; the relations between elements are determined purely by the structure of the ‘compound’ syntax; that is, in syntactic terms Sanskrit compounds are fully configurational. It is therefore necessary to distinguish these two types of syntax, and also to ensure that the correct form of a word is used in the appropriate syntactic context: inflected forms at the end of compound sequences and outside of compounds, and uninflected ‘stem’ forms inside compounds. As noted above, it is also necessary to capture at least something of the fact that compound phrases, while phrases, are nevertheless closer to words than noncompound phrases.

As mentioned above, the analysis provided in this section is formulated in the framework of LFG. LFG is a strict lexicalist theory of syntax, and offers a strongly modular representation of grammar, whereby different types of grammatical information are represented separately and permitted to interact only via ‘projection’ functions between

³¹ See, for example, the following papers on compounding within the framework of LFG, all of which assume a fundamentally lexical status for the phenomena under discussion: Ørnesnes 1996, Baker & Nordlinger 2008, and Lee & Ackerman 2011.

modules.³² In particular, LFG distinguishes at least two syntactic components.³³ C(onstituent)-structure is represented using the familiar ‘tree’ diagrams, but represents purely surface phrasal configuration, and not abstract relations such as grammatical functions (subject, object, etc.), control relations, and unbounded dependencies, or abstract features such as tense or definiteness properties. Abstract grammatical relations and features are represented at a separate level of structure, the f(unctional)-structure. These structures are related by a function, labeled ϕ , that maps c-structure nodes to f-structures. As an example, I provide the c- and f-structure in 23, which represents the syntactic analysis of the sentence in example 1, repeated as 22.



The tree represents the hierarchical syntactic structure of the clause. The specific assumptions about Sanskrit c-structure that the tree implies are not important for the present purposes, only that the c-structure represents the surface phrasal configuration, with all words analyzed in their surface linear position.³⁴ The associated f-structure represents the abstract grammatical relations: the main predicate of the clause is the verb *adāt* ‘give’, and the three other words supply this verb’s subject, object, and oblique arguments. The relation between the c- and f-structures is formalized as the function ϕ , represented by the labeled arrows. This is constructed on the basis of the annotations on the c-structure nodes, for example $\uparrow=\downarrow$. The symbol \downarrow represents a function from the current c-structure node to the f-structure projected from that node, while \uparrow represents a function from the current c-structure node’s mother to its f-structure. The annotation $\uparrow=\downarrow$ therefore means that the f-structure projected from the current c-structure node is the same as the f-structure projected from the current c-structure node’s mother. Fol-

³² On the modularity of LFG, see Dalrymple & Mycock 2011 and Lowe 2016.

³³ Some models treat the ‘(syntactic) string’, or linearly ordered string of words, as a distinct level, separate from the c-structure, but this can be ignored for the present purposes. See Lowe 2015b for details and references on the string.

³⁴ For example, I know of no clear evidence for the existence of intermediate phrases (*X'*) in Sanskrit, and since nothing depends on it and it simplifies the representation, I assume only *XP* and *X*. For more detailed discussion of the syntactic structure of Sanskrit, see for example Gillon & Shaer 2005 and Lowe 2015a: 37–46.

lowing down the right-hand edge of the tree, the annotations specify that the f-structure projected from the VP node is the same as the f-structure projected from the S node, and is also the same as the f-structure projected from the V node. S represents the clause, so the f-structure projected from S represents the f-structure for the clause; since the f-structure projected from the V is the same f-structure, the V can supply the PRED value for the clausal f-structure (PRED 'give'). Following the left-hand edge of the tree, the annotation (\uparrow SUBJ) = \downarrow on the NP node specifies that the f-structure projected from the NP serves as the value of SUBJ in the f-structure projected from the S; \uparrow = \downarrow on the daughter N means that the f-structure for the N and NP are the same, meaning that the N can supply the value of SUBJ in the f-structure for the clause (PRED 'Devadatta'). The other annotations work in the same way.³⁵

C-structure, that is, the representation of surface syntactic relations, is of most concern for the present analysis, since it is here that the lexicalism of LFG is most apparent: only syntactic words can be associated with the terminal nodes of the c-structure tree; that is, sublexical units are inaccessible to the c-structure. F-structure is also important, however, since it is always necessary to show that the relevant abstract grammatical relations can be derived from the c-structure configuration assumed.

It is worth reconsidering for a moment what is meant by phrasal syntax, and what the consequences are of our definition. In representing c-structure, LFG makes use of a version of X' theory (Jackendoff 1977), one of the most widely adopted approaches to phrasal structure.³⁶ One way of understanding X' theory is as a claim that the syntax of all languages can be stated in terms of relations between phrases, which themselves consist of combinations of words and phrases.

Assuming X'' is the maximal phrase, there are at most two distinct types of phrase, XP (\equiv X'') and X', which can exist in any language (though some languages may have only one phrasal level), in addition to the word level, X⁰. Just as X' theory has been successfully applied to English and many other languages, so can it also successfully be applied to the noncompound 'sentential' syntax of Sanskrit (as is done, for example, in Lowe 2015a). It is important to note that, as mentioned above, within the theoretical framework adopted here, phrasal syntactic relations are taken to represent purely features of surface configurationality, and not more abstract, functional relations between words.

For this reason the ways in which X' theory is applied and modified in the analysis of any particular phenomenon are driven not by theoretical assumptions about the nature of the most basic, underlying relations between words, but purely by evidence for surface configurationality and constituency.

On a strict-ish version of X' theory, we can make the following statement about the concept of a syntactic word: a word is an element of language whose relations to other elements can be stated wholly by reference to the relations between phrases admitted by X' theory. That is, a word is any element that can be analyzed as heading a phrase, a phrase that can contain other phrases, and that can itself be contained within another phrase headed by another word. To put it another way, the relations between a word and other words in a clause are stated in terms of XP and/or X' PHRASES, and not directly in terms of the words themselves. That is, there is no direct syntactic relation between one word and another word under X' theory: words are directly related, in syntactic terms, only to phrases.

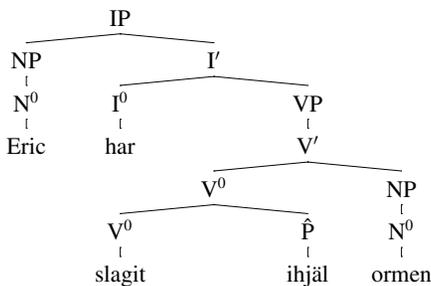
³⁵ For further introduction to the formalism of LFG, with specific reference to Sanskrit, see Lowe 2015a: 47–83.

³⁶ Alongside the bare phrase structure of the minimalist program and related theories.

This is a strong claim about the nature of syntax and syntactic relations. While for that reason attractive, it is in fact likely to be too strong. One important argument for weakening X' theory in this respect is made by Toivonen (2003): there is evidence that some words do not project phrases. Toivonen (2003) argues, within the framework of LFG, that alongside the traditional projecting lexical categories, there exist also NON-PROJECTING categories, represented as \hat{X} , which adjoin to X^0 heads. Nonprojecting words do not head phrases, and so it is not possible for another phrase to stand in a specifier, complement, or adjunct relation to such a word. Words that do not project phrasal structure are often particles and/or clitics. Toivonen proposes the augmentation to X' theory shown in 24; she argues in detail that verb particles in Swedish are nonprojecting \hat{P} s. Example 25 is from Toivonen 2003:2.

(24) $X^0 \rightarrow X^0, \hat{Y}$

(25) Eric har slagit ihjäl ormen
 Eric has beaten to.death snake.DEF
 'Eric has beaten the snake to death'



To permit adjunction of nonprojecting categories to phrasal heads it may also be necessary to extend the proposal. Spencer (2005) argues for adjunction of nonprojecting words to XP in order to capture the properties of case clitics in Hindi, and if this is possible there is little reason to reject adjunction to X'.

(26) a. $X' \rightarrow X', \hat{Y}$

b. $XP \rightarrow XP, \hat{Y}$

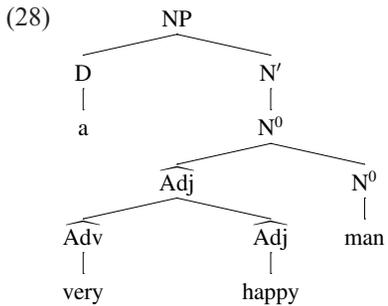
While the proposal that some words do not project phrases is empirically strong (see Toivonen 2003), it does somewhat undermine the strength of X' theory, at least as a claim that has relevance for the distinction between words (/morphology) and phrases (/syntax). If we consider only adjunction of \hat{X} to X^0 , it should be clear that the distinctions between word and morpheme, and between word and phrase, have been loosened. If phrases are defined as syntactic constituents that potentially contain more than one word, X^0 is no longer a nonphrasal category. In addition, the distinction between \hat{X} and morphemes must be clearly made on other grounds, since it is no longer entirely clear. For example, one could make an argument for treating most regular inflectional affixes as \hat{X} s adjoined to Y^0 , which would violate the most basic assumptions of strict lexicalism.

From a lexicalist perspective this could be seen as a bad thing. By contrast, it can equally be seen as a valuable augmentation to X' theory, which permits it to capture something of the ambiguity between syntax and morphology. Even on a strict lexicalist view of grammar, the indisputable fact of diachronic processes such as grammaticalization requires some acknowledgment of either a gradient between lexical element (word) and sublexical element (morpheme), or the possibility of ambiguous status. That is, since words can become morphemes, it must be possible that at some point the analysis of a particular form may be intermediate or ambiguous.

The theory of nonprojecting categories permits precisely that. An \hat{X} , just like an X^0 , is a lexical word. But \hat{X} s do not participate in phrasal syntax in the same way that X^0 s do, and so are somehow ‘less syntactic’.³⁷ Moreover, the syntactic grouping of an \hat{X} and an X^0 is of the same category, X^0 , as all lexical words. An \hat{X} adjoined to an X^0 , then, is a lexical word, but a lexical word that does not participate in phrasal syntax in the same way as other lexical words, and that is perhaps more liable than a projecting word to be reanalyzed as a morpheme. Many of the sorts of words most easily analyzed as nonprojecting—for example, verb particles, case-marking clitics, and so on—are in fact just the sorts of words that are somewhat less independent than other words, and that may be part way along the cline of grammaticalization toward morphemes.

An important extension to the theory of nonprojecting categories was proposed by Duncan (2007) and, more recently, by Arnold and Sadler (2013).³⁸ Arnold and Sadler base their proposals on the relatively familiar features of prenominal modification in English. Building on work by Poser (1992) and Sadler and Arnold (1994), they argue that prenominal modification in English should be analyzed in terms of nonprojecting categories (on the basis of evidence such as the fact that prenominal adjectives cannot take postmodifying phrases, unlike adjectives in other positions). But since prenominal modification is recursive, this requires that nonprojecting categories can be adjoined not only to X^0 (and XP and X'), but also to nonprojecting \hat{X} s. That is, we require a rule of the kind in 27; the analysis proposed by Arnold and Sadler (2013) for prenominal modification in English is shown in 28.

(27) $\hat{X} \rightarrow \hat{Y} \hat{X}$



³⁷ Although it is common to treat morphological relations in ways parallel to syntactic relations, for example, with morphemes functioning as heads of morpheme ‘phrases’ (stems) and governing dependent morpheme ‘phrases’ (stems), the evidence for such relations is almost entirely functional, or abstract, and not based on tests for surface constituency. Morphemes cannot generally be reordered or participate in any of the processes that are used to determine surface syntactic constituency, and so within a framework where phrasal structure is strictly distinguished from abstract functional structure, the sorts of phrasal relations possible in syntax cannot be transferred to morphology. That is, ignoring functional relations between morphemes and stems, there is actually little or no evidence for any kind of hierarchical structure within the word, and so the only relations we can assume between morphemes are direct linear relations between elements, if we assume any at all. (Assuming none at all would amount to saying that the separation of functional relations from hierarchical supports a realizational approach to morphology, as argued for by e.g. Kaplan and Kay (1994), Stump (2001), Sadler and Spencer (2001), Spencer (2003, 2006, 2013), and Beesley and Karttunen (2003).) There is, then, a clear difference between syntactic relations and morphological relations: the former involve hierarchy, and direct relations only between elements (words) and groups of elements (words); the latter involve, if anything, direct relations between elements (morphemes and stems). This is the justification for treating the inability to display ‘phrasal’ relations as a feature of ‘less syntactic’ status.

³⁸ Duncan (2007) suggests that it was already implied by Toivonen (2003), though it was not explicitly mentioned.

This proposal changes the nature of nonprojecting categories in a fundamental way, since it means that a nonprojecting word can now head an \hat{X} ‘phrase’—a phrase, undoubtedly, but a phrase of very different kind from, and considerably more restricted than, XP/X' phrases. While Toivonen’s original proposal is highly relevant to the potentially ambiguous distinction between clitics and affixes, the extension proposed by Duncan (2007) and Arnold and Sadler (2013) is directly relevant to the other great question mark over the syntax–morphology divide: compounding/incorporation phenomena.³⁹ In relation to English, for example, it captures the structural similarity between $Adj + N$ phrases and $Adj + N$ compounds, and may therefore help to account for why the former can so often become reanalyzed as the latter: the syntactic phrase *black bird* is an N^0 , just as the lexical compound *blackbird* is. All that is lost in the reanalysis of an $Adj + N$ phrase as an $Adj + N$ compound is the internal syntactic structure of the mother N^0 , and not any higher phrasal (X' or XP) structure. The productivity of $N + N$ sequences in English, the possibility of recursion of these sequences, and the fact that the status of these sequences as phrases or lexical compounds is often ambiguous likewise receive a natural explanation. So, the sequence *photo frame insert manufacturing specification* is an N^0 just as *photo frame* is, and just as *photo* is. The possibility arises of explaining the ambiguous status of these sequences in terms of the reanalysis of N^0 phrases as N^0 words.

So, nonprojecting categories are directly relevant to syntactic phenomena that are somewhat less syntactic than ‘ordinary’ projecting syntax, and to words and phrases on the border between syntax and morphology. For this reason, they are exactly what is needed for analyzing Classical Sanskrit compounding.

Essentially, I propose rules of the form in 29 for Sanskrit compound syntax. Classical Sanskrit compounds will therefore have a syntactic structure parallel to premodified English N^0 s, as in 28 above.

- (29) a. $X^0 \rightarrow \hat{Y} X^0$
 b. $\hat{X} \rightarrow \hat{Y} \hat{X}$

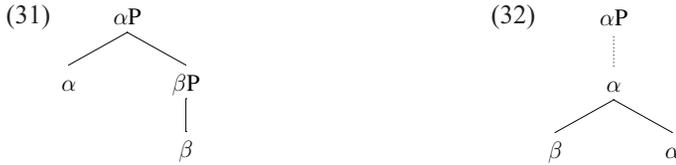
An analysis of this kind has a number of benefits. Recursive adjunction of \hat{Y} to \hat{X} captures the phrasal nature of Sanskrit compounds, while keeping the phrase structure rules for compounds fully distinct from the rules of noncompound syntax. The use of \hat{X} as the category for nonfinal members of compounds permits a clear distinction to be made between inflected words and the uninflected stem forms in compound syntax: stem forms are \hat{X} ; fully inflected words are X^0 . And the adjunction of \hat{X} to X^0 captures the fact that phrasally formed compounds are closer to lexical words than noncompound phrases. The details of my analysis for the four major types of compound discussed in this article are presented in the next section.

The proposals made here, and the analysis I have advanced for understanding nonprojecting categories in terms of the unclear dividing line between syntax and morphology, have notable parallels in some important morphological treatments of compounding. Anderson (1992:292–319) discusses the status of compounding within his ‘a-morphous morphology’ theory. He argues that compounds are formed by ‘word structure rules’, distinct from the ‘word formation rules’ that deal with most morphology. These rules can apply either in the lexicon or the syntax. Notably, he proposes the following ‘phrase structure’ rule for English $N + N$ compounds.

- (30) $N \rightarrow N N$

³⁹ Duncan’s (2007) proposal is made with specific reference to noun incorporation; nonprojecting categories were also proposed to account for incorporation structures by Asudeh (2007).

Here, the \underline{N} represents the head. Anderson points out that this ‘word structure rule’ is similar to ordinary X’ rules, but differs in a number of ways, for example, in that the sisters of a head are lexical categories, rather than phrases. It is obvious enough how this corresponds to a syntactic phrase structure rule involving a nonprojecting category (e.g. 24). A similar analysis for English compounding is proposed by Ackema and Neeleman (2004). They consider syntax and morphology to be two distinct structure-generating components of the grammar, but components that are both separate from the lexicon (which is a repository of exceptions). In their view, syntax and morphology can come into competition. In relation to English compounding, they assume that the following two structural possibilities are in competition: the tree in 31 is the ‘syntactic’ structure, while the tree in 32 is the ‘morphological’ structure.



Their ‘syntactic’ structure corresponds to the traditional possibilities of X’ syntax, while their ‘morphological’ structure corresponds to a rule involving nonprojecting categories. Insofar as ‘morphology’, in their conception, is distinct from the lexicon and functions as an independent structure-generating system, it would involve only a minor reassignment to treat this morphological rule as part of the syntactic component, which would correspond closely to my proposal.

In the following sections I present a syntactic analysis of the most common compound types found in Classical Sanskrit, demonstrating that the use of nonprojecting categories provides a fully adequate account of their features and status.

6. DVANDVAS. As discussed above, dvandvas are essentially compounds of coordinate nouns.⁴⁰ To repeat one of the examples given above, the nouns *ratha-* ‘chariot’, *gaja-* ‘elephant’, and *aśva-* ‘horse’ can be coordinated to create the dvandva compound [*ratha- gaja- aśva-*] ‘chariot(s), elephant(s), and horse(s)’.

As discussed in the previous section, I propose that Sanskrit’s specialized compound syntax, which is distinct from the ‘regular’ syntax found outside of compounds, can be modeled using the nonprojecting categories of Toivonen (2003), assuming recursive adjunction of \hat{Y} to \hat{X} . The uninflected ‘stem’ form of words that appear in nonfinal position in a compound are of category \hat{X} , while inflected words, or words that can stand independently, outside a compound, are of category X^0 . This means that the ‘special’ syntax of compounds can be stated in phrase structure rules with reference to nonprojecting categories. The following phrase structure rule is required for dvandvas.⁴¹

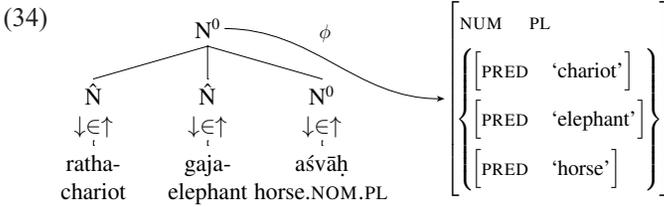
⁴⁰ Padrosa-Trias (2010) denies the existence of coordinate compounding, arguing that coordination is a purely syntactic process and that apparent coordinate compounds involve asyndetic coordination in the syntax. Although my analysis of Sanskrit dvandvas does treat them as a kind of asyndetic coordination, I do not exclude the possibility of genuine coordinate compounds, either in earlier/later stages of Sanskrit, or crosslinguistically.

⁴¹ I understand ‘adjunction’ in purely phrasal terms, and not as a combination of phrasal (c-structure) and functional (f-structure) features. This is why the functional annotations do not conform to what is expected of ‘adjunction’ under the ‘structure-function’ mapping principles proposed by Bresnan (2001:99–122) and Toivonen (2003); such principles are important generalizations, but mismatches between phrasal structure and functional relations are possible. The importance of keeping a clear distinction between phrasal structure and functional relations is discussed, in relation to coordination/subordination mismatches, by Belyaev (2015).

$$(33) N^0 \rightarrow \hat{N}^+ \quad N^0$$

$$\downarrow \in \uparrow \quad \downarrow \in \uparrow$$

The annotation $\downarrow \in \uparrow$ below each daughter node specifies that the daughters together constitute a set at functional structure (which is how coordination is represented in LFG). This will produce the structure in 34 for the compound [ratha-gaja-aśvāḥ].



The compound is therefore an N^0 , but an N^0 consisting of three words, just like *very happy man* in 28 above. The nonfinal elements in this N^0 are \hat{N} , and so they are instantiated by the stem forms of the words concerned. That is, stem forms are lexically specified as instantiating only nonprojecting category nodes, while inflected forms are lexically specified as instantiating only projecting category nodes. Since the final word in the mother N^0 is a lexical N^0 , it is inflected for number, case, and gender.⁴²

The rule in 33 is sufficient for dvandvas with an inflected final element, that is, dvandvas that are not further embedded in a compound structure and that participate in the ‘ordinary’ noncompound syntax of their clause. However, as with all compounds in Sanskrit, it is equally possible for dvandvas to appear embedded within a larger compound. When not final in a compound sequence, the final element of a dvandva will, of course, appear in stem form; that is, it must be an \hat{N} , not an N^0 . At the same time, the node dominating the dvandva must be \hat{N} , not N^0 , since all embedding within a compound sequence involves adjunction of nonprojecting categories. Therefore, alongside the phrase structure rule in 33, we must also assume the rule in 35.⁴³

$$(35) \hat{N} \rightarrow \hat{N}^+ \quad \hat{N}$$

$$\downarrow \in \uparrow \quad \downarrow \in \uparrow$$

The phrase structure rules in 33 and 35 are essentially the same, except for the parameter of projection/nonprojection. That is, the projecting/nonprojecting status of the category on the left-hand side of the rule determines the projecting/nonprojecting status of the rightmost category on the right-hand side. In LFG, categories can be parametrized in phrase structure rules for c-structural features; parametrized categories are known as ‘complex’ categories.⁴⁴ For example, $V_{[fin]}$ can be used to represent the category of finite Vs, while $V_{[inf]}$ and $V_{[ptc]}$ represent infinitive and participle Vs, respectively. Variables can be used to range over sets of exclusive features; so, $V_{[_{ftness}]}$ represents a V parametrized for finiteness: the feature $[_{ftness}]$ must be instantiated with a particular

⁴² Note that the number marking on the final element applies to the compound as a whole, and not specifically to the final element. This is best (and unproblematically) accounted for as a semantic issue.

⁴³ All of the compound types that can be treated as syntactically formed can ultimately be analyzed as right-headed, such that if the node dominating a particular compound is itself X^0 , its rightmost daughter will necessarily also be X^0 . Because of this, even if a dvandva, say, constitutes the second member of another compound—for example, a bahuvrīhi—it will inherit the parameter of projection from the compound within which it is embedded, and its final member will therefore be inflected. Bahuvrīhis are usually analyzed as nonheaded compounds, and avyayībhāvas are usually analyzed as left-headed: my analysis of these is presented below, and this analysis can be extended to the few other compound types (e.g. some minor types of tatpuruṣa) that are traditionally analyzed as left-headed.

⁴⁴ On parametrized phrase structure rules and ‘complex’ categories see, for example, Kuhn 1999, Frank & Zaenen 2002, Falk 2003, and Crouch et al. 2011.

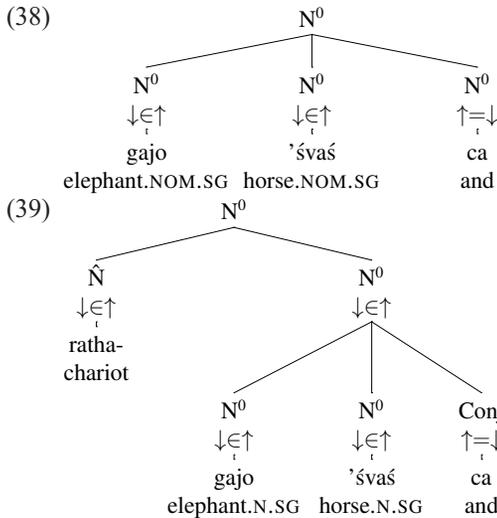
value, say [fin], [inf], or [ptc], within any given phrase structure rule. Since projection/nonprojection is a c-structure feature, it can be parametrized. What we have hitherto referred to as X^0 is essentially just X plus the feature ‘projection’; likewise, \hat{X} is X plus the feature ‘nonprojection’.

We can therefore represent X^0 and \hat{X} as $X_{[proj]}$ and $X_{[nonproj]}$ respectively. The variable that ranges over [proj] and [nonproj] we can label [_pr]. So within any phrase structure rule, $X_{[_pr]}$ must be instantiated as either X^0 (i.e. $X_{[proj]}$) or \hat{X} (i.e. $X_{[nonproj]}$) in all of its occurrences. We can therefore generalize over the rules in 33 and 35 with the rule in 36, which is precisely equivalent to either rule in 37.

$$\begin{aligned}
 (36) \quad & N_{[_pr]} \rightarrow N_{[nonproj]}^+ N_{[_pr]} \\
 & \quad \downarrow \in \uparrow \quad \downarrow \in \uparrow \\
 (37) \quad & \text{a. } N_{[_pr]} \rightarrow \hat{N}^+ N_{[_pr]} \\
 & \quad \downarrow \in \uparrow \quad \downarrow \in \uparrow \\
 & \quad \text{b. } \left\{ \begin{array}{l} N^0 \rightarrow \hat{N}^+ \quad N^0 \\ \downarrow \in \uparrow \quad \downarrow \in \uparrow \end{array} \right. \quad \left. \begin{array}{l} \hat{N} \rightarrow \hat{N}^+ \quad \hat{N} \\ \downarrow \in \uparrow \quad \downarrow \in \uparrow \end{array} \right\}
 \end{aligned}$$

Parametrizing over projection/nonprojection therefore permits the phrase structure rules for Sanskrit compounds to be expressed much more succinctly, since only one rule is required to cover both of the possible contexts in which a particular compound may appear.

As noted above, dvandvas are essentially coordinate compounds. Apart from this, it is not possible to coordinate elements within a compound. So, the rule licensing ordinary, noncompound, coordination of X s must be prevented from applying within a compound. That is, the structure in 38 is perfectly admissible, but the structure in 39 is impossible, since the N^0 that undergoes coordination is ‘within’ a compound.



It would be unproblematic to prevent coordination from applying to nonfinal elements in a compound by simply stating that nonprojecting categories cannot be coordinated, except by dvandva coordination. But the restriction must apply also to the final, projecting, element of a compound, which is, at least superficially, indistinguishable from other X^0 categories that can undergo coordination. The solution is to admit a further c-structure feature that may be subject to parametrization, controlling coordinability. The phrase structure rules for compounding will then specify a feature, [nocoord], of all elements on the right-hand side of the rule, whereas the rule for coordination of X^0 is restricted to X^0 s with the feature [coord] (in the lexicon projecting words will be

underspecified for this feature). The rule for dvandva coordination in 36 above can therefore be stated with more detail as in 40, while the rule for X^0 coordination will be as in 41.⁴⁵

$$(40) N_{[_{pr}]} \rightarrow N_{[nonproj],[nocoord]}^+ N_{[_{pr}],[nocoord]}$$

$$(41) X_{[coord]} \rightarrow X_{[coord]}^+ X_{[coord]} \text{ Conj}$$

In the phrase structure rules given below, I omit the [nocoord] feature so as to keep the rules simple, but it can be assumed to be present. In both phrase structures and trees, I also retain X^0 and \hat{X} in place of (i.e. as abbreviations for) $X_{[proj]}$ and $X_{[nonproj]}$, respectively, for consistency with earlier sections of the article and with previous work on nonprojecting categories.

7. TATPURUṢAS. The analysis of tatpuruṣas is also relatively simple. The canonical type, the vibhakti-tatpuruṣa introduced above, involves either N + N or N + Adj sequences, with a ‘case’ relation inferrable between the two elements. To repeat the examples given above, [*svarga-patita-*], lit. ‘heaven-fallen’, is interpreted by inferring an ablatival relation between the elements, that is, ‘fallen from heaven’, while [*rāja-bhāryā-*], lit. ‘king-wife’, is interpreted by inferring a genitival or possessive relation between the elements, that is, ‘the king’s wife’. The case relation is not a structural but an abstract syntactic relation, and so in LFG is represented at f-structure. The particular relation between any two words is entirely contextually determined, partly on the basis of the lexical meanings of the two elements concerned, but also, where this admits of some ambiguity, on the wider clausal context. For example, [*bhū-patita-*], lit. ‘earth-fallen’, is interpreted as involving a directional (accusative case) relation between the elements, that is ‘fallen to earth’. The difference between this and [*svarga-patita-*] is entirely dependent on the relative positions of heaven and earth. However, the compound [*vṛkṣa-patita-*], lit. ‘tree-fallen’, is more ambiguous: this could mean ‘fallen from a/the tree’ or ‘fallen onto a/the tree’, depending on the context. The former is perhaps the neutral interpretation, but only because the property of having fallen from a tree is more common, and generally more salient, than the property of having fallen onto a tree. Similarly, [*grāma-gata-*] could mean either ‘having gone to the village’ or ‘having left (i.e. gone from) the village’, though again the former is the less marked interpretation. In other cases the interpretation of the compound may not be ambiguous, in broad terms, but it is possible to attribute at least two different thematic roles to the first element. So, for example, [*ātāpa-śuṣka-*], lit. ‘sun-dried’, which can be taken to mean either ‘dried in the sun’ or ‘dried by the sun’.

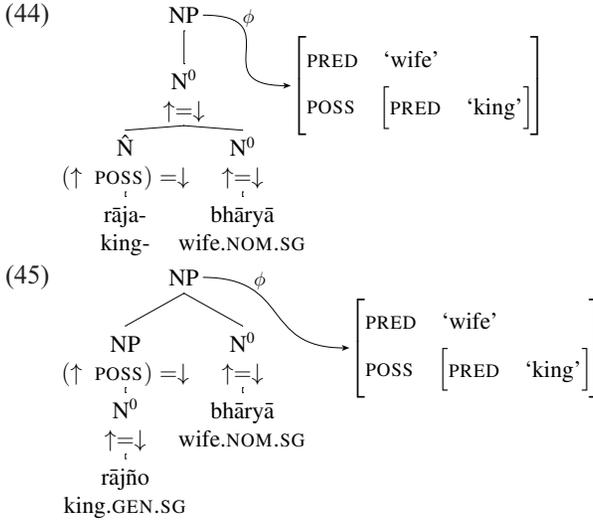
The important point is that the functional relation between two elements of a tatpuruṣa cannot be determined structurally, so the rules governing the formation of tatpuruṣas must permit selection between a variety of relations. The following rules specify the formation of N + Adj and N + N tatpuruṣas respectively; a disjunctive list specifies the range of possible functional relations between elements (the full specification has been omitted to simplify the representation).

$$(42) \text{Adj}_{[_{pr}]} \rightarrow \hat{N} \text{Adj}_{[_{pr}]}$$

$$(43) N_{[_{pr}]} \rightarrow \hat{N} N_{[_{pr}]}$$

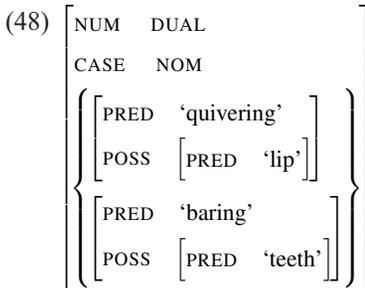
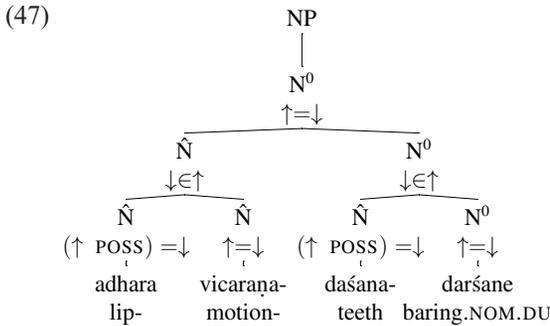
⁴⁵ The placement possibilities for coordinating conjunctions in Sanskrit are more complex than suggested by 41, but I ignore the details here since they are not relevant for the present purposes.

The resulting c-structure and f-structure for the compound [*rāja- bhāryā*] are shown in 44. For comparison, 45 shows the same structures for the equivalent noncompounded phrase, *rājño bhāryā* ‘wife of the king’; the functional structure corresponding to this is identical to that in 44.



It is now possible to exemplify how the embedding of compounds inside other compounds works. The tree in 47 shows the structure resulting from the application of both the tatpuruṣa and dvandva rules to the phrase in example 46, which is a simplified version of part of the compound in example 20. The corresponding functional structure is given in 48.

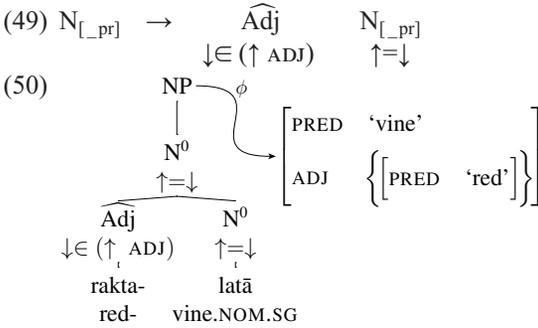
(46) [[*adhara- vicaraṇa-*]_{ip} [*daśana- darśane*]_{ip,dd}
lip- motion- teeth- baring.NOM.DU
‘quivering of the lips and baring of the teeth.’



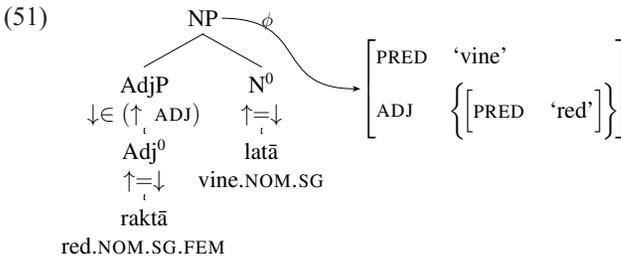
This structure results from the application of the dvandva coordination rule to an N^0 , followed by the application of the $N + N$ tatpuruṣa rule to both conjuncts. In the case of the leftmost conjunct, the mother N is nonprojecting, so the parameter of nonprojection is passed to the rightmost element of the tatpuruṣa, giving an \hat{N} that is instantiated by the stem form *vicaraṇa-*. In the case of the rightmost conjunct, the N is projecting, so the parameter passes down to the rightmost element of the tatpuruṣa (which is also thereby the rightmost element of the dvandva as a whole), and, as an N^0 , it is instantiated by the inflected word form *darśane*.

8. KARMADHĀRAYAS. As noted above, the term karmadhāraya is applied to a number of distinct compound types. The three types introduced above are all simple to account for within the proposed framework.

8.1. ADJ + N. Compounds of $Adj + N$, in which the Adj functions as an adjectival modifier of the noun, are particularly simple. The phrase structure rule in 49 is required to account for these compounds. For a compound such as [*rakta- latā-*] ‘red vine’, the c-structure and f-structure in 50 result.



The noncompound equivalent of this sequence differs only in that the modifying adjective constitutes a full $AdjP$ phrase; it must therefore be adjoined to NP , and it is of course fully inflected. In f-structure terms, there is no difference between this and the compound construction.



8.2. ADJ + ADJ AND ADV + ADJ. This compound type involves the modification of an adjective by either an adjective or an adverb, for example, [*udagra- ramaṇīya-*] ‘intensely lovely’ (lit. ‘intense-lovely’), [*madhura- ukta-*] ‘sweetly spoken’ (lit. ‘sweet-spoken’), [*punar- ukta-*] ‘spoken again’ (lit. ‘again-spoken’), [*evam- bhūta-*] ‘being so’ (lit. ‘thus-being’). The rule in 52 accounts for this type.⁴⁶ As this is so similar to the preceding type (*mutatis mutandis*), I omit example c-structures and f-structures.

⁴⁶ I assume that Adj and Adv are distinct categories, following for example Payne et al. 2010, but nothing depends on this; in fact, the rules would be simpler if only a single category A were utilized.

$$(52) \text{Adj}_{[\text{pr}]} \rightarrow \left\{ \begin{array}{l} \widehat{\text{Adj}} \mid \widehat{\text{Adv}} \\ \downarrow \in (\uparrow \text{ADJ}) \quad \uparrow = \downarrow \end{array} \right\} \text{Adj}_{[\text{pr}]}$$

One subtype of Adj + Adj karmadhāraya, which is covered by the preceding rule but for which a more complex analysis might be desirable, involves a compound of which both elements are so-called ‘past participles’, as in the following example.

$$(53) [\text{snāta- anulipta-}]_{\text{kd}} \\ \text{bathed- anointed-} \\ \text{‘having bathed and (then) anointed oneself’}$$

The Indian grammatical tradition understands such compounds as implying a temporal sequence, such that the action referred to by the first element temporally precedes that referred to by the second element. This could be covered by treating the first participle as an adjunct to the second, giving a literal sense of something like ‘having anointed oneself after bathing’. The potential complication with analyzing such compounds is that the grammatical status of the ‘past participle’ is somewhat ambiguous. It is, at least formally, a verbal adjective, which in categorial terms is an adjective (hence its appearance under this heading), but in Classical Sanskrit it often functions as the equivalent of a finite past-tense verb form. In compounds such as the example given above, it is practically equivalent, at least in functional terms, to the Sanskrit converb (also called the absolutive or gerund), which cannot appear in compound phrases.⁴⁷

The correct analysis of such compounds depends not only on the categorial status of the past participle, but also on how its semantics are modeled. Altogether, it is possible for the rule in 52 above to apply to this sort of compound, but a more complex rule may provide a more nuanced analysis; for the present I leave this subtype to one side.

8.3. N + N. N + N karmadhārayas involve an identity between two nouns, of which both may be common nouns, or one may be a proper noun of some sort. Essentially, the first noun functions as a modifier of the second, restricting the set of possible referents. For example, [*rāja- ṛṣi-*], lit. ‘king-seer’ (i.e. seer who is also a king); [*amātya- devadatta-*], lit. ‘minister-Devadatta’ (‘Devadatta the Minister’); [*strī- jana-*], lit. ‘women-folk’ (‘people who are women’); [*dhvani- śabda-*], lit. ‘*dhvani*-word’ (i.e. ‘the word *dhvani*’); [*kāñcī- pura-*], lit. ‘Kāñcī-city’ (‘the city of Kāñcī’). These compounds can be modeled in an entirely parallel manner to the Adj + N compounds. The relevant phrase structure rule is given in 54; the c-structure and f-structure for [*amātya- devadatta-*] ‘Devadatta the Minister’ are shown in 55.

$$(54) \text{N}_{[\text{pr}]} \rightarrow \left\{ \begin{array}{l} \widehat{\text{N}} \\ \downarrow \in (\uparrow \text{ADJ}) \quad \uparrow = \downarrow \end{array} \right\} \text{N}_{[\text{pr}]}$$

$$(55) \begin{array}{c} \text{NP} \\ \downarrow \\ \text{N}^0 \\ \uparrow = \downarrow \\ \widehat{\text{N}} \quad \text{N}^0 \\ \downarrow \in (\uparrow \text{ADJ}) \quad \uparrow = \downarrow \\ \text{amātya-} \quad \text{devadattaḥ} \\ \text{minister-} \quad \text{Dd.NOM.SG} \end{array} \left[\begin{array}{l} \text{PRED} \quad \text{‘Devadatta’} \\ \text{ADJ} \quad \left\{ \left[\text{PRED} \quad \text{‘Minister’} \right] \right\} \end{array} \right]$$

9. BAHUVRĪHIS. The analysis of bahuvrīhi compounds is the most problematic of all the compound types discussed here. As noted above, bahuvrīhis are often described as

⁴⁷ At least not in the sorts of free and productive compounds discussed here.

‘exocentric’, since their reference is to an entity external to the compound, but it is not immediately obvious whether this should be understood as a syntactic or a semantic property, or both. Scalise and Guevara (2006) discuss exocentric compounding in a typological perspective, and show that exocentricity in compounding may manifest itself in a variety of ways, syntactic and semantic.

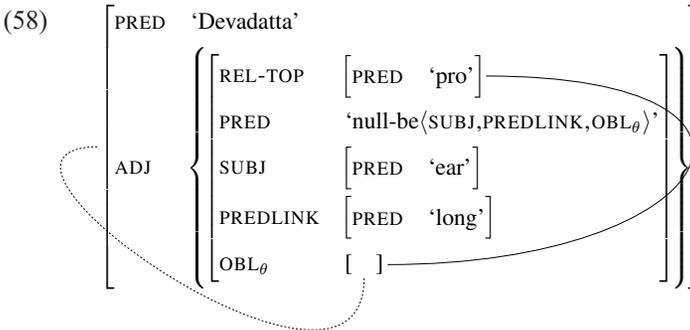
In functional terms, a bahuvrīhi can be analyzed as expressing an embedded predication to which the external head bears some contextually determined relation. So in the phrase in 56, the embedded predication is ‘ears (are) long’, and in this case, as in many bahuvrīhis, the relation of the external head is one of possession.

- (56) [dīrgha- karṇo] devadattaḥ
 long- ear.NOM.SG.M Devadatta
 ‘Devadatta, whose ears are long/long-eared Devadatta’

Other relations are possible, depending (just as with tatpuruṣas) on the lexical meanings of the words involved and the wider context. In particular, when a past participle is used as the first member of a bahuvrīhi, the external head may bear an argument role in relation to the event referred to by the participle (usually an instrument, if the participle is interpreted passively).

- (57) [jñāta- sarvasvo] devadattaḥ
 known- entirety.NOM.SG.M Devadatta
 ‘Devadatta, who knows everything/by whom everything is known’

In functional terms, then, a phrase such as [dīrgha-karṇo] devadattaḥ will be modeled in parallel manner to a phrase involving a relative clause. The f-structure representation for such a phrase will be as in 58.



The embedded predication is represented using a ‘null-be’ predicate that selects for a SUBJECT argument (the element predicated of), a PREDLINK argument (the predicated element), and an oblique argument OBL_θ.⁴⁸ The oblique argument can be instantiated to a variety of thematic roles, including recipient/possessor.

In phrase-structural terms, a number of different possibilities could be suggested for modeling the internal structure of bahuvrīhi compounds. If the exocentricity of bahuvrīhis were attributed not only to the semantics and functional structure, but also to the phrase structure, then bahuvrīhis could be modeled by means of a specialized exocentric phrase structure category, which we might call ‘B’.⁴⁹

⁴⁸ The assumption of a null copular and the use of the PREDLINK argument is one way of formalizing verbless predications in LFG; for discussions of this and the alternative possibilities, see Rosén 1996, Butt et al. 1999, Dalrymple et al. 2004, Falk 2004, Nordlinger & Sadler 2007, Attia 2008, Sulger 2009, Dione 2012, Laczko 2012, and Lowe 2013.

⁴⁹ LFG admits a number of exocentric phrasal categories. The most widely accepted is S, the exocentric clausal node, commonly assumed for nonconfigurational languages; the ‘expression node’ E (Aissen 1992) is

$$(59) B \rightarrow \widehat{\text{Adj}} \{\widehat{\text{N}} \mid \text{N}^0\}$$

Besides the fact that ‘B’ would be a somewhat ad hoc solution, however, there is a formal problem: when the final element of a bahuvrīhi is also the final element of the compound (i.e. when the bahuvrīhi is not embedded in a larger compound, or forms the last element of a larger compound), the final element is of course inflected, meaning that it must be a projecting N^0 (since nonprojecting categories in Sanskrit are uninflected under the proposals made here). But an N^0 directly dominated by ‘B’ does not, in fact, project a phrase, so could only be classed as an N^0 by ignoring the key defining feature of that category. The only alternative would be $\widehat{\text{N}}$. But admitting inflected $\widehat{\text{N}}$ s would undermine the otherwise clear distinction between $\widehat{\text{N}}$ and N^0 .⁵⁰ The same problem affects another possible solution, namely an Adj^0 dominating an $\widehat{\text{Adj}}$ and an $\text{N}^0/\widehat{\text{N}}$.

$$(60) \text{Adj}_{[\text{pr}]} \rightarrow \widehat{\text{Adj}} \text{N}_{[\text{pr}]}$$

Again, according to this rule, when the mother is Adj^0 , the rightmost daughter must be N^0 , but an N^0 that does not project a phrase. One way to overcome this problem would be to assume that bahuvrīhis are in fact Ns that can head an NP. However, this would run counter to the usual assumption that these compounds are fundamentally adjectival structures, an assumption for which there is good evidence. A rule such as the following would work, but the claim that bahuvrīhis are essentially nominal structures would be hard to sustain.

$$(61) \text{N}_{[\text{pr}]} \rightarrow \widehat{\text{Adj}} \text{N}_{[\text{pr}]}$$

Gillon (2007) discusses the status of bahuvrīhis in Classical Sanskrit and provides a number of arguments for treating them as fundamentally adjectival. First, the final element, when inflected, shows full adjectival agreement (even though the final element is invariably a noun, categorially). Furthermore, the possibilities for secondary derivation from bahuvrīhis parallel those for adjectives; that is, the same affixes that can be used to derive nouns, adverbs, and so on from adjectives are also found attached to bahuvrīhis. These facts strongly suggest that bahuvrīhis should be treated as adjectival structures; in the formalism proposed here, then, the node directly dominating a bahuvrīhi compound must be of category Adj .

Gillon (2007) argues that bahuvrīhis (at least the type considered here) are best analyzed as derived from $\text{Adj} + \text{N}$ karmadhāraya compounds by addition of a phonetically null possessive suffix. This proposal is similar to the analysis of English bahuvrīhis proposed by Kiparsky (1982), which likewise involves a null head. That is, a karmadhāraya like [*dīrgha- karṇa-*] ‘long ear’ is converted into a bahuvrīhi [*dīrgha- karṇa-*] ‘long-eared’ by null affixation. Gillon argues that support for this analysis comes from the common use of the *-ka-* suffix on bahuvrīhis (cf. §3.5), which he treats as a nonnull alternative to the null suffix.⁵¹

Within the lexicalist and syntactic analysis of compounding pursued here, it is not possible to assume an affix that attaches to syntactically formed compounds without

often assumed as an exocentric superclausal node for expressing the relation between dislocated elements and their clauses; the CCL ‘clausally scoped clitic cluster’ (Bögel et al. 2010, Lowe, 2011) is utilized for headless clitic clusters, and is therefore an exocentric phrasal node. ‘B’ would be another exocentric phrasal node.

⁵⁰ A further problem with assuming an exocentric category B is that one could not use parametrization for projection/nonprojection as a way of determining whether the rightmost daughter of B was $\widehat{\text{N}}$ or N^0 in any given instance (since an exocentric category cannot be parametrized for projection/nonprojection, by definition).

⁵¹ Gillon (2007) further argues that such suffixation patterns are paralleled in English, where *-ed* in *long-eared*, *long-legged*, and so on can be treated as an adjectival bahuvrīhi-forming suffix, alternating with a null bahuvrīhi suffix when such compounds are used as nouns (e.g. *red-head*).

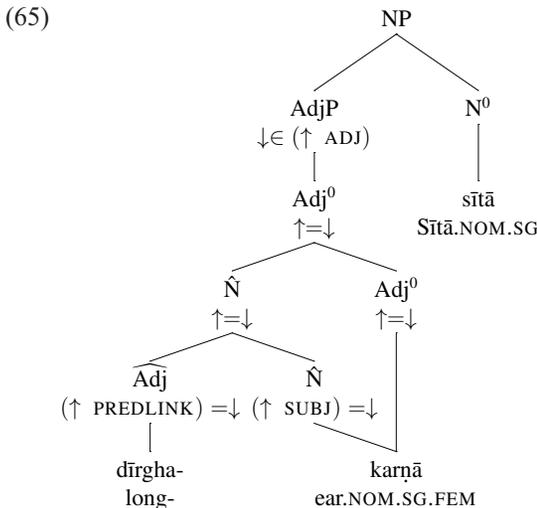
nouns, and to resolve the difficulties in the analysis of bahuvrīhi compounds, I propose to analyze them using the theory of LEXICAL SHARING.⁵⁴

A number of authors in LFG, beginning with Wescoat (2002, 2005, 2007, 2009), admit the possibility that certain lexical items can instantiate two nodes in the phrase structure.⁵⁵ Examples discussed by Wescoat include pronoun-auxiliary contractions in English, for example, *I'll*, *you've*, *he'd*, and preposition-determiner contractions in French and German, for example, French *au*, *du*. These forms are ‘portmanteau words’: they display a number of properties that require them to be treated as single lexical elements; for example, they cannot be split up and their phonological form cannot be derived by regular phonological processes of contraction from two distinct elements. But at the same time, they display some features of two-word sequences, for example, patterning paradigmatically with unambiguous two-word sequences.

The possibility of lexical sharing permits a very neat analysis of bahuvrīhi compounding, and in particular the fact that nouns constituting the final element in a bahuvrīhi can show full adjectival agreement. I propose that forms of nouns inflected as adjectives are stored in the lexicon with the specification that they must instantiate two nodes in the phrase structure, \hat{N} and Adj^0 . There also exist stem forms of nouns that can instantiate \hat{N} and $\widehat{\text{Adj}}$, when a bahuvrīhi is embedded within another compound.

For example, the phrase in example 64 shows a bahuvrīhi agreeing in number, case, and gender with the noun it modifies. The final element of the bahuvrīhi is a noun that is inherently masculine and so, except when used in a bahuvrīhi agreeing with (or referring to) a nonmasculine noun, cannot have feminine or neuter forms. The tree in 65 shows the structure for this example.

- (64) [dīrgha- karṇā]_{bv} sītā
 long- ear.NOM.SG.F Sītā.NOM.SG
 ‘long-eared Sita’



⁵⁴ Strictly speaking, the ‘constrained lexical sharing’ of Lowe 2015b, but the specifics are unimportant here.

⁵⁵ Besides Wescoat, see also Broadwell 2007, 2008, Alsina 2010, Belyaev 2014, and Lowe 2015b. Lexical sharing represents an instantiation in LFG terms of ‘multidominance’ in syntactic representation; for the equivalent in other syntactic theories, see, for example, Citko & Gračanin-Yuksek 2013:5ff., with references, for minimalism; Williams 2003 for representation theory; and Ramchand 2008, Svenonius 2011 for ‘spanning’ in nanosyntax. Within LFG it is similar to, but distinct from, treatments of ‘mixed categories’, such as the English gerund that displays features of both noun and verb, such as by Bresnan (1997), Malouf (2000), Mugane (2003), and Bresnan and Mugane (2006).

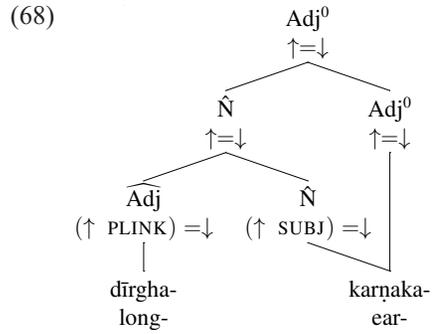
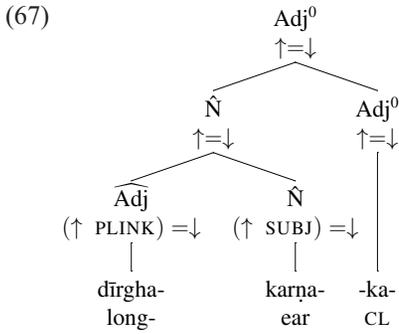
Under this analysis, the final ‘noun’ in a bahuvrīhi simultaneously instantiates both the Adj head of the compound sequence and the \hat{N} that supplies the subject for the embedded predication. It is necessary to ensure both that forms such as *karṇā* can only appear with a sequence of \hat{N} and $\text{Adj}^0/\widehat{\text{Adj}}$ that is produced by the bahuvrīhi rules in 62 and 63 above and, conversely, that the rules in 62 and 63 can only be instantiated by forms such as *karṇā*, and not by any sequence of N and Adj. In fact, both requirements fall out easily given the rules proposed. For the latter requirement no additional constraint is required: the annotation under the \hat{N} in 63 supplies a null copular PRED, which prevents the instantiation of Adj by any word that supplies its own PRED (as all adjectives will), since in that case there would be two distinct specifications of the PRED value (a ‘PRED clash’), which is not permitted. The former requirement is achieved by a specification in the lexical entry requiring that the f-structure projected from the Adj^0 contains an attribute SUBJ. The lexical entry also supplies no PRED value for the f-structure projected from the Adj^0 , and this in combination with the requirement for a SUBJ means that it cannot be associated with an Adj^0 that is not part of a bahuvrīhi structure. This is because the Adj^0 in a bahuvrīhi already has its PRED specified in the phrase structure rules, and this PRED subcategorizes for a SUBJ. But any other Adj^0 would not have a PRED value independently specified, so could not subcategorize for a SUBJ, meaning that the resulting f-structure would fail the requirement for COHERENCE (which disallows governable grammatical functions that are not subcategorized for in the PRED value of the relevant f-structure). I therefore propose lexical entries of the type in 66 for nouns that can appear as the final element in a bahuvrīhi.⁵⁶ The important specification is the first under the Adj^0 , which enforces the appearance of a SUBJ feature in the f-structure projected from the Adj^0 .

(66)	karṇā: \hat{N}	Adj^0
	(\uparrow PRED) = ‘ear’	(\uparrow SUBJ)
		(\uparrow GEND) = FEM
		(\uparrow NUM) = SG
		(\uparrow CASE) = NOM

According to this lexical entry, the adjectival agreement properties of a noun appearing in a bahuvrīhi are associated, at least in functional terms, with the Adj node that it instantiates, while the lexical meaning is associated, appropriately, with the N node that it instantiates.

In the case of bahuvrīhis showing *-ka-* suffixation, there are two possible analyses. For a compound such as *dirgha-karṇa-ka-*, discussed in §3.5, it would be possible either to treat the element *-ka-* as a clitic, filling the Adj^0 node in the c-structure for the bahuvrīhi (with appropriate functional constraints to prevent its use in other contexts), or else to treat the sequence *karṇa-ka-* as a single lexical element (i.e. taking *-ka-* as an affix) with the same specifications as in 66 above. That is, the possible c-structures for the bahuvrīhi *dirgha-karṇa-ka-* are as follows.

⁵⁶ As noted, the feminine form of *karṇa-* used as an example here does not exist as an independent form, but is used only in bahuvrīhi compounds. If the same noun is used in a bahuvrīhi that appears in the masculine, the form used will be identical to a form that could exist independently. Rather than assume two homophonous versions of every case form of every noun, one used only in bahuvrīhis and one elsewhere, it is possible to assume a single lexical entry, with the two possible c-structure instantiations (i.e. N^0 or $\hat{N} \text{Adj}^0$) and associated f-descriptions treated as alternatives. The same will apply to the two possible instantiations of stem forms (as \hat{N} or $\hat{N} \widehat{\text{Adj}}$). In fact there will be (at least) three alternative instantiations for stem forms, since $\hat{N} \widehat{\text{Adv}}$ is required as a possibility for avyayībhāvas (see next section).

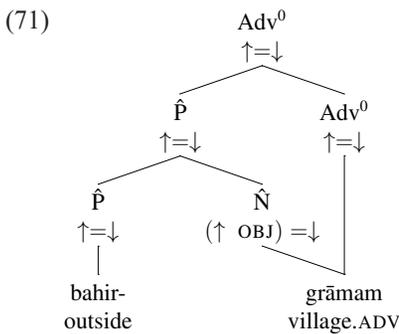


Which analysis is correct depends purely on a decision as to the status of *-ka-*, something that is beyond the scope of the present work (cf. §3.5).

10. AVYAYĪBHĀVAS. As discussed above, *avyayībhāva* compounds are essentially adverbs formed of a compounded prepositional structure, for example, [*bahir- grāmam*], lit. ‘outside-village’. According to the Indian grammatical tradition, such compounds are left-headed, since the head of a prepositional phrase is clearly the preposition. However, these compounds are functionally very close to adverbs, and morphologically, too, show adverbial marking on the second element that cannot be accounted for purely in terms of a prepositional structure. This suggests an analysis along parallel lines to the analysis of *bahuvrīhis*, discussed in the previous section. That is, I propose the following phrase structure rules for *avyayībhāva* compounds.

- (69) $Adv_{[_{pr}]} \rightarrow \hat{P} \quad Adv_{[_{pr}]}$
 $\quad \quad \quad \uparrow=\downarrow \quad \quad \quad \uparrow=\downarrow$
 $\quad \quad \quad \hat{N} \in CAT(\uparrow OBJ)$
- (70) $\hat{P} \rightarrow \hat{P} \quad \hat{N}$
 $\quad \quad \quad \uparrow=\downarrow \quad (\uparrow OBJ)=\downarrow$

These rules parallel the rules for *bahuvrīhis* (62–63 above) closely. An *avyayībhāva* compound is of category *Adv*, dominating a \hat{P} and an *Adv*. This \hat{P} is constrained by the annotation involving the *CAT* predicate to be expanded by the rule in 70, that is, as $\hat{P} \hat{N}$. Adverbial forms of nouns, such as *grāmam* ‘village.ADV’, will then instantiate both the \hat{N} and *Adv* nodes, as follows.



This structure will depend on a lexical entry for *grāmam* parallel to that for *karṇā* provided in 66 above.

- (72)

$grāmam: \hat{N} \quad Adv^0$ $(\uparrow PRED) = \text{‘village’} \quad (\uparrow OBJ)$
--

11. DIACHRONY. The constructions and phenomena discussed in this article are specifically constructions and phenomena of the Classical Sanskrit period. At a slightly earlier stage, in the Vedic Sanskrit language (c. 1500–600 BC), some of the same patterns are seen, but in general the evidence regarding the morphosyntactic status of compounding points more toward a morphological status. For example, in the earliest texts, the *R̥gveda* and *Atharvaveda*, the productive recursivity of compounding is highly restricted: no compounds of more than three members, and only a very few compounds of three members, are attested (e.g. Macdonell 1916:§185). Compounds involving pronouns (i.e. that display inbound anaphora) are attested, for example, *R̥gvedic tāt-apas-* ‘whose work is that’ (*bahuvrīhi*), but they are not common, and evidence regarding the anaphoric possibilities into and out of compounds, and for *asamartha* compounding, is more limited.

Most of the compound types attested in Sanskrit have clear parallels in cognate languages, and thus can be projected back to the earliest reconstructable ancestor of Sanskrit, Proto-Indo-European.⁵⁷ In the other old Indo-European languages, compounds are, as in Vedic Sanskrit, largely restricted to two members, and appear much more clearly to be formed via morphological processes, displaying anaphoric islandhood much more consistently, for example. The status of compounding as a process of word formation in old Indo-European languages is generally taken for granted (e.g. Clackson 2002:163). As for the parent language, it is again widely assumed (usually without discussion) that compounds in Proto-Indo-European were words, that is, that compound formation was a morphological/lexical process (e.g. Meier-Brügger 2010:427–30). However, it is also widely held that the compounding processes reconstructable to Proto-Indo-European had a syntactic origin—that is, that they arose via univertation and lexicalization of originally syntactically juxtaposed sequences (e.g. Brugmann 1889:3, Schindler 1997, Clackson 2002). There is clear evidence for this in what are often called ‘*unechte Komposita*’, compounds that preserve an original case form in the prior member. For example, Vedic Sanskrit *dāmpati-* ‘master’ and Ancient Greek *despôtēs* ‘master’ both continue Proto-Indo-European **dems-poti-*, lit. ‘master of the house’, in which the first element **dems-* reflects a genitive singular form of the word **dem-* ‘house’ (though neither the *dam-* of the Sanskrit form nor the *des-* of the Greek would have been synchronically analyzable as a genitive singular). Compound patterns in which the nonfinal element is not inflected (i.e. appears in what we have called the ‘stem’ form, for Classical Sanskrit) are assumed to have originated by precisely the same processes of univertation, but at a stage of Pre-Proto-Indo-European preceding the development of case inflection, for example, before it became obligatory to use the genitive (and other cases) to express syntactic dependencies between nouns (e.g. Brugmann 1889:23ff., but see also e.g. Schindler 1997, Dunkel 1999).

In both cases, what we are presumably dealing with is the listing and subsequent lexicalization of specific syntactic sequences, for example, due to high frequency of a particular collocation, and subsequently the reanalysis of the still discernable syntactic structure underlying the collocation as a morphological structure, which could then be treated as productive and extended to permit the creation of compounds directly in the morphology. Both of these processes can be understood in terms of grammaticalization: the univertation and lexicalization of specific syntactic sequences involves a change from full word to subsyntactic unit for both members of the resulting compound, while

⁵⁷ For a brief overview of compounding in Indo-European see Kastovsky 2009.

the reanalysis of an originally syntactic process as a morphological process can be understood as a grammaticalization in terms of the morphologization of a syntactic construction (see e.g. Haspelmath 2004:26, and Vincent & Börjars 2010:284–85). It is parallel, for example, to the univerbation of preverb-verb complexes in the history of many languages, including Sanskrit, Ancient Greek, and so on: while it originates in the lexicalization of specific preverb-verb sequences, the lexicalization is generalized to the construction as a whole.

Classical Sanskrit dvandva compounding differs from the other types of compounds discussed here in that its development from a phrasal syntactic construction can be observed within the historical period (as discussed in detail by Kiparsky 2010). The so-called ‘devatā-dvandvas’ of Vedic involve pairs of conventionally associated divine or human referents, for example, ‘Mitra and Varuṇa’ (gods), ‘Heaven and Earth’ (divine personifications), ‘mother and father’. In some instances, these dvandvas clearly involve two syntactically independent words that need not, for example, appear adjacent to one another; the only evidence of their close association is that both nouns must appear in the dual, although each noun itself refers to only a single individual. For example, *indrā ... vāruṇā* (Indra.DU ... Varuṇa.DU) ‘Indra(SG) and Varuṇa(SG)’ (*R̥gveda* 4.41.3; two words intervene). This represents the earliest stage in the development of these dvandvas; at this stage, both forms appear in the appropriate case, for example, *mitrāyora vāruṇayora* (Mitra.GEN.DU Varuṇa.GEN.DU) ‘of Mitra and Varuṇa’ (*R̥gveda* 7.66.1). In some instances, however, the first member of the pair appears in an invariant form (which reflects the nominative/accusative dual), and in this case the two words are more consistently adjacent, for example, [*indrā- vāruṇayoh*] (Indra.DU-Varuṇa.GEN.DU) ‘of Indra and Varuṇa’ (*R̥gveda* 1.17.1). This stage represents the beginnings of univerbation. Subsequent developments involve obligatory adjacency, loss of dual marking on the first member, and loss of accent on the first member, resulting in the same dvandva construction seen in Classical Sanskrit and discussed above, which fully parallels the other major compound types in having its first member appear in uninflected ‘stem’ form. The developments seen in the evolution of devatā-dvandvas in Vedic—that is, the lexicalization and univerbation of common collocations, and the subsequent reanalysis of an originally syntactic structure (asyndetic coordination) as a morphological one—are undoubtedly very similar to those that must have occurred in the evolution of Proto-Indo-European compounding patterns. Kiparsky (2010:320–21) explicitly refers to this process as a process of grammaticalization.

Although the precise morphosyntactic status of compounds in Vedic Sanskrit remains to be established, as noted above it is widely accepted that compounding processes in Proto-Indo-European, including those that underlie most of the major compound types of Classical Sanskrit, were fundamentally morphological processes. Since I have argued that in Classical Sanskrit these same compounding processes are SYNTACTIC processes, this implies a change at some point in the history or prehistory of Sanskrit whereby morphological processes were reanalyzed as syntactic processes. Insofar as these morphological processes themselves originated as syntactic processes in (Pre-)Proto-Indo-European, this can be seen a reversion, or at least a reanalysis in the opposite direction from that assumed for earlier periods. And insofar as it is possible to treat the morphologization of syntactic patterns as a grammaticalization, the reanalysis of morphological processes as syntactic can be considered an example of DEGRAMMATICALIZATION.⁵⁸

⁵⁸ On the existence, or not, of degrammaticalization as the reverse process to grammaticalization, see for example Ramat 1992, Geurts 2000, Campbell & Janda 2001, van der Auwera 2002, Haspelmath 2004, and Norde 2009, 2010.

That is, the development that must be assumed for Classical Sanskrit, whereby inherited morphological processes of compound formation were reanalyzed as syntactic processes, corresponds to the reverse of the more widely attested grammaticalization process involving the morphologization of syntactic processes. Given the strong tendency for unidirectionality in grammaticalization, this development is therefore notable, though its further diachronic and typological implications remain to be investigated.

12. CONCLUSION. In this article, I have discussed a number of properties displayed by Classical Sanskrit ‘compounds’ that suggest they must be treated as syntactically formed, and I have presented a formal analysis of the most common and productive compound types within a strictly lexicalist syntactic theory, which does not merely account for the data but also models the intermediate status of compounding as somewhat closer to a lexical phenomenon than other syntactic processes. There are a number of interesting implications that come out of this study, in particular from a historical perspective, as discussed in the preceding section.

In more theoretical terms, the use of nonprojecting categories to model the distinct syntax of Sanskrit compounds permits a formalization of the somewhat intermediate status of compounding between full phrasal syntax and morphology. Nonprojecting categories therefore provide a way of understanding and modeling diachronic processes such as grammaticalization and univerbation, for which the possibility of intermediate status or ambiguous analysis is a necessary prerequisite. Moreover, they do this within a strictly lexicalist syntactic theory, for which the gradient between phrase, word, and morpheme is in principle absolute (and for that reason highly problematic). No syntactic theory, however lexicalist, can deny, or ignore, the reality of diachronic processes in which phrases become words, or words parts of words, or vice versa; the analysis proposed here, for one very specific example of such processes, provides the beginnings, at least, of a demonstration that no syntactic theory, however lexicalist, NEED deny or ignore them.

REFERENCES

- ACKEMA, PETER, and AD NEELEMAN. 2004. *Beyond morphology: Interface conditions on word formation*. Oxford: Oxford University Press.
- AISSEN, JUDITH L. 1992. Topic and focus in Mayan. *Language* 68.1.43–80.
- ALSINA, ALEX. 2010. The Catalan definite article as lexical sharing. *Proceedings of the LFG '10 Conference*, 5–25. Online: <http://csli-publications.stanford.edu/LFG/15/papers/lfg10alsina.pdf>.
- ANDERSON, STEPHEN R. 1992. *A-morphous morphology*. Cambridge: Cambridge University Press.
- ARNOLD, DOUG, and LOUISA SADLER. 2013. Displaced dependent constructions. *Proceedings of the LFG '13 Conference*, 48–68. Online: <http://csli-publications.stanford.edu/LFG/18/papers/lfg13arnoldsadler.pdf>.
- ASUDEH, ASH. 2007. Some notes on pseudo noun incorporation in Niuean. Ottawa: Carleton University, MS. Online: <http://users.ox.ac.uk/~cpgl0036/pdf/niuean.pdf>.
- ATTIA, MOHAMMED. 2008. A unified analysis of copula constructions in LFG. *Proceedings of the LFG '08 Conference*, 89–108. Online: <http://csli-publications.stanford.edu/LFG/13/papers/lfg08attia.pdf>.
- BAKER, BRETT, and RACHEL NORDLINGER. 2008. Noun-adjective compounds in Gunwinyguan languages. *Proceedings of the LFG '08 Conference*, 109–28. Online: <http://csli-publications.stanford.edu/LFG/13/papers/lfg08bakernordlinger.pdf>.
- BAKER, MARK C. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.
- BEESEY, KENNETH R., and LAURI KARTTUNEN. 2003. *Finite-state morphology*. Stanford, CA: CSLI Publications.

- BELYAEV, OLEG I. 2014. Осетинский как язык с двухпадежной системой: Групповая флексия и другие парадоксы падежного маркирования. *Вопросы Языкознания* 6.74–108.
- BELYAEV, OLEG I. 2015. Systematic mismatches: Coordination and subordination at three levels of grammar. *Journal of Linguistics* 51.2.267–326.
- BISETTO, ANTONIETTA, and SERGIO SCALISE. 1999. Compounding: Morphology and/or syntax? *Boundaries of morphology and syntax*, ed. by Lunella Mereu, 39–56. Amsterdam: John Benjamins.
- BISETTO, ANTONIETTA, and SERGIO SCALISE. 2005. Classification of compounds. *Lingue e linguaggio* 2.319–32.
- BÖER, KATJA; SVEN KOTOWSKI; and HOLDEN HÄRTL. 2011. Nominal composition and the demarcation between morphology and syntax: Grammatical, variational, and cognitive factors. *Proceedings of Anglistentag 2011*, Freiburg, ed. by Monika Fludernik and Benjamin Kohlmann, 63–74. Trier: Wissenschaftlicher Verlag Trier.
- BÖGEL, TINA; MIRIAM BUTT; RONALD M. KAPLAN; TRACY HOLLOWAY KING; and JOHN T. MAXWELL III. 2010. Second position and the prosody-syntax interface. *Proceedings of the LFG '10 Conference*, 106–26. Online: <http://csli-publications.stanford.edu/LFG/15/papers/lfg10boegeletal.pdf>.
- BOOIJ, GEERT E. 2005. Compounding and derivation: Evidence for construction morphology. *Morphology and its demarcations*, ed. by Wolfgang U. Dressler, Franz Rainer, Dieter Kastovsky, and Oskar Pfeiffer, 109–32. Amsterdam: John Benjamins.
- BOOIJ, GEERT E. 2009. Lexical integrity as a formal universal: A constructionist view. *Universals of language today*, ed. by Sergio Scalise, Elisabetta Magni, and Antonietta Bisetto, 83–100. Dordrecht: Springer.
- BRESNAN, JOAN. 1997. Mixed categories as head sharing constructions. *Proceedings of the LFG '97 Conference*. Online: <http://csli-publications.stanford.edu/LFG/LFG2-1997/lfg97bresnan.pdf>.
- BRESNAN, JOAN. 2001. *Lexical-functional syntax*. Oxford: Blackwell.
- BRESNAN, JOAN, and SAM A. MCHOMBO. 1995. The lexical integrity principle: Evidence from Bantu. *Natural Language and Linguistic Theory* 13.2.181–254.
- BRESNAN, JOAN, and JOHN MUGANE. 2006. Agentive nominalizations in Gikũyũ and the theory of mixed categories. *Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan*, ed. by Miriam Butt, Mary Dalrymple, and Tracy Holloway King, 201–34. Stanford, CA: CSLI Publications.
- BROADWELL, GEORGE AARON. 2007. Lexical sharing and non-projecting words: The syntax of Zapotec adjectives. *Proceedings of the LFG '07 Conference*, 87–106. Online: <http://csli-publications.stanford.edu/LFG/12/papers/lfg07broadwell.pdf>.
- BROADWELL, GEORGE AARON. 2008. Turkish suspended affixation is lexical sharing. *Proceedings of the LFG '08 Conference*, 198–213. Online: <http://csli-publications.stanford.edu/LFG/13/papers/lfg08broadwell.pdf>.
- BRUGMANN, KARL. 1889. *Grundriß der vergleichenden Grammatik der indogermanischen Sprachen, Band 2.1: Wortbildungslehre (Stammbildungs- und Flexionslehre)*. Strassburg: Trübner.
- BUTT, MIRIAM; TRACY HOLLOWAY KING; MARÍA-EUGENIA NIÑO; and FRÉDÉRIQUE SEGOND. 1999. *A grammar writer's cookbook*. Stanford, CA: CSLI Publications.
- CAMPBELL, LYLE, and RICHARD JANDA. 2001. Introduction: Conceptions of grammaticalization and their problems. *Language Sciences* 23.2–3.93–112.
- CITKO, BARBARA, and MARTINA GRAČANIN-YUKSEK. 2013. Towards a new typology of coordinated *wh*-questions. *Journal of Linguistics* 49.1.1–32.
- CLACKSON, JAMES. 2002. Composition in Indo-European languages. *Transactions of the Philological Society* 100.2.163–67.
- COULSON, MICHAEL. 1992. *Sanskrit: An introduction to the classical language*. 2nd edn. London: Teach Yourself Books.
- CROUCH, DICK; MARY DALRYMPLE; RONALD M. KAPLAN; TRACY HOLLOWAY KING; JOHN T. MAXWELL III; and PAULA NEWMAN. 2011. *XLE documentation*. Palo Alto: Palo Alto Research Center. Online: http://www2.parc.com/isl/groups/nltxle/doc/xle_toc.html.
- DALRYMPLE, MARY. 2001. *Lexical functional grammar*. San Diego: Academic Press.
- DALRYMPLE, MARY; HELGE DYVIK; and TRACY HOLLOWAY KING. 2004. Copular complements: Closed or open? *Proceedings of the LFG '04 Conference*, 188–98. Online: <http://csli-publications.stanford.edu/LFG/9/lfg04ddk.pdf>.

- DALRYMPLE, MARY, and TRACY HOLLOWAY KING. 2013. Nested and crossed dependencies and the existence of traces. *From quirky case to representing space: Papers in honor of Annie Zaenen*, ed. by Tracy Holloway King and Valeria de Paiva, 139–51. Stanford, CA: CSLI Publications.
- DALRYMPLE, MARY, and LOUISE MYCOCK. 2011. The prosody-semantics interface. *Proceedings of the LFG '11 Conference*, 173–93. Online: <http://csli-publications.stanford.edu/LFG/16/papers/lfg11dalrymplemycock.pdf>.
- DELBRÜCK, BERTHOLD. 1878. *Die altindische Wortfolge aus dem Çatapathabrāhmaṇa*. Halle: Verlag der Buchhandlung des Waisenhauses.
- DI SCIULLO, ANNA-MARIA, and EDWIN WILLIAMS. 1987. *On the definition of word*. Cambridge, MA: MIT Press.
- DIONE, CHEIKH BAMBA. 2012. An LFG approach to Wolof cleft constructions. *Proceedings of the LFG '12 Conference*, 157–76. Online: <http://csli-publications.stanford.edu/LFG/17/papers/lfg12dione.pdf>.
- DUNCAN, LACHLAN. 2007. Analytic noun incorporation in Chuj and K'ichee' Mayan. *Proceedings of the LFG '07 Conference*, 163–83. Online: <http://csli-publications.stanford.edu/LFG/12/papers/lfg07duncan.pdf>.
- DUNKEL, GEORGE E. 1999. On the origins of nominal composition in Indo-European. *Compositions indogermanicae (in memoriam Jochem Schindler)*, ed. by Heiner Eichner and Hans Christian Luschützky, 47–68. Praha: Enigma.
- ERSCHLER, DAVID. 2012. Suspended affixation and the structure of syntax-morphology interface. *Studia Linguistica Hungarica* 59.153–75.
- FALK, YEHUDA N. 2001. *Lexical-functional grammar: An introduction to parallel constraint-based syntax*. Stanford, CA: CSLI Publications.
- FALK, YEHUDA N. 2003. The English auxiliary system revisited. *Proceedings of the LFG '03 Conference*, 184–204. Online: <http://csli-publications.stanford.edu/LFG/8/lfg03falk.pdf>.
- FALK, YEHUDA N. 2004. The Hebrew present-tense copula as a mixed category. *Proceedings of the LFG '04 Conference*, 226–46. Online: <http://csli-publications.stanford.edu/LFG/9/lfg04falk.pdf>.
- FRANK, ANNETTE, and ANNIE ZAENEN. 2002. Tense in LFG: Syntax and morphology. *How we say when it happens: Contributions to the theory of temporal reference in natural language*, ed. by Hans Kamp and Uwe Reyle, 17–52. Tübingen: Niemeyer.
- GEURTS, BART. 2000. Explaining grammaticalization (the standard way). *Linguistics* 38.4. 781–88.
- GIEGERICH, HEINZ J. 2004. Compound or phrase? English noun-plus-noun constructions and the stress criterion. *English Language and Linguistics* 8.1.1–24.
- GIEGERICH, HEINZ J. 2005. Associative adjectives in English and the lexicon-syntax interface. *Journal of Linguistics* 41.571–91.
- GIEGERICH, HEINZ J. 2009. The English compound stress myth. *Word Structure* 2.1–17.
- GILLON, BRENDAN S. 1991. Sanskrit word formation and context free rules. *Toronto Working Papers in Linguistics* 11.15–45.
- GILLON, BRENDAN S. 1994. Bhartṛhari's solution to the problem of *asamartha* compounds. *Bhartṛhari, philosopher and grammarian: Proceedings of the First International Conference on Bhartṛhari (University of Poona, January 6–8, 1992)*, ed. by Saroja Bhate and Johannes Bronkhorst, 117–33. Delhi: Motilal Banarsidass.
- GILLON, BRENDAN S. 1995. The autonomy of word formation: Evidence from Classical Sanskrit. *Indian Linguistics* 56.15–52.
- GILLON, BRENDAN S. 2007. Exocentric (bahuvrīhi) compounds in Classical Sanskrit. *Proceedings of the First International Sanskrit Computational Linguistics Symposium*, Paris, October 29–31, 2007, ed. by Gérard Huet and Amba Kulkarni, 1–12. Rocquencourt: Inria. Online: <http://sanskrit.inria.fr/Symposium/Proceedings.pdf>.
- GILLON, BRENDAN S. 2009. Tagging Classical Sanskrit compounds. *Sanskrit computational linguistics*, ed. by Gérard Huet, Amba Kulkarni, and Peter Scharf, 98–105. Berlin: Springer.
- GILLON, BRENDAN S., and BENJAMIN SHAER. 2005. Classical Sanskrit, 'wild trees', and the properties of free word order languages. *Universal grammar in the reconstruction of ancient languages*, ed. by Katalin É. Kiss, 457–93. Berlin: Mouton de Gruyter.
- GONDA, JAN. 1952. *Remarques sur la place du verbe dans la phrase active et moyenne en langue sanscrite*. Utrecht: Oosthoek.

- HARLEY, HEIDI. 2009. Compounding in distributed morphology. In Lieber & Štekauer 2009b, 129–44.
- HARRIS, ALICE C. 2006. Revisiting anaphoric islands. *Language* 82.1.114–30.
- HASPELMATH, MARTIN. 2004. On directionality in language change with particular reference to grammaticalization. *Up and down the cline: The nature of grammaticalization*, ed. by Olga Fischer, Muriel Norde, and Harry Perridon, 17–44. Amsterdam: John Benjamins.
- HASPELMATH, MARTIN. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45.1.31–80.
- JACKENDOFF, RAY. 1977. *X̄ syntax: A study of phrase structure*. Cambridge, MA: MIT Press.
- JOSHI, S. D. (ed.) 1974. *Patañjali's Vyākaraṇa-mahābhāṣya: Bahuvrīhidvādhānika (P. 2.2.23–2.2.38)*. Introduction by S. D. Joshi. Text, translation, and notes by J. A. F. Roodbergen. Poona: University of Poona.
- KAPLAN, RONALD M., and JOAN BRESNAN. 1982. Lexical-functional grammar: A formal system for grammatical representation. *The mental representation of grammatical relations*, ed. by Joan Bresnan, 173–281. Cambridge, MA: MIT Press.
- KAPLAN, RONALD M., and MARTIN KAY. 1994. Regular models of phonological rule systems. *Computational Linguistics* 20.3.331–78.
- KAPLAN, RONALD M., and JOHN T. MAXWELL III. 1996. LFG grammar writer's workbench. Technical report. Palo Alto: Xerox Palo Alto Research Center.
- KASTOVSKY, DIETER. 2009. Diachronic perspectives. In Lieber & Štekauer 2009b, 323–40.
- KIPARSKY, PAUL. 1982. Lexical phonology and morphology. *Linguistics in the morning calm*, ed. by In-Seok Yang, 3–91. Seoul: Hanshin.
- KIPARSKY, PAUL. 2010. Dvandvas, blocking, and the associative: The bumpy ride from phrase to word. *Language* 86.2.302–31.
- KUHN, JONAS. 1999. Towards a simple architecture for the structure-function mapping. *Proceedings of the LFG '99 Conference*. Online: <http://csli-publications.stanford.edu/LFG/LFG4-1999/lfg99kuhn.pdf>.
- KUMAR, ANIL; V. SHEEBASUDHEER; and AMBA KULKARNI. 2009. Sanskrit compound paraphrase generator. *Proceedings of ICON-2009: 7th International Conference on Natural Language Processing*, ed. by Dipti Misra Sharma, Vasudeva Varma, and Rajeev Sangal, 229–38. Hyderabad: Macmillan.
- LACZKÓ, TIBOR. 2012. On the (un)bearable lightness of being an LFG style copula in Hungarian. *Proceedings of the LFG '12 Conference*, 341–61. Online: <http://csli-publications.stanford.edu/LFG/17/papers/lfg12laczko.pdf>.
- LAPOINTE, STEVEN G. 1980. *The theory of grammatical agreement*. Amherst: University of Massachusetts, Amherst dissertation.
- LEE, LESLIE, and FARRELL ACKERMAN. 2011. Mandarin resultative compounds: A family of lexical constructions. *Proceedings of the LFG '11 Conference*, 320–38. Online: <http://csli-publications.stanford.edu/LFG/16/papers/lfg11leeackerman.pdf>.
- LIEBER, ROCHELLE, and PAVOL ŠTEKAUER. 2009a. Introduction: Status and definition of compounding. In Lieber & Štekauer 2009b, 3–18.
- LIEBER, ROCHELLE, and PAVOL ŠTEKAUER (eds.) 2009b. *The Oxford handbook of compounding*. Oxford: Oxford University Press.
- LOWE, JOHN J. 2011. Ṛgvedic clitics and 'prosodic movement'. *Proceedings of the LFG '11 Conference*, 360–80. Online: <http://csli-publications.stanford.edu/LFG/16/papers/lfg11lowe.pdf>.
- LOWE, JOHN J. 2013. (De)selecting arguments for transitive and predicated nominals. *Proceedings of the LFG '13 Conference*, 398–418. Online: <http://csli-publications.stanford.edu/LFG/18/papers/lfg13lowe.pdf>.
- LOWE, JOHN J. 2014. Accented clitics in the Ṛgveda. *Transactions of the Philological Society* 112.1.5–43.
- LOWE, JOHN J. 2015a. *Participles in Rigvedic Sanskrit: The syntax and semantics of adjectival verb forms*. Oxford: Oxford University Press.
- LOWE, JOHN J. 2015b. English possessive 's': Clitic and affix. *Natural Language and Linguistic Theory*, to appear.
- LOWE, JOHN J. 2016. Clitics: Separating syntax and prosody. *Journal of Linguistics, FirstView article*. DOI: <http://dx.doi.org/10.1017/S002222671500002X>.

- MACDONELL, ARTHUR ANTHONY. 1916. *A Vedic grammar for students*. Oxford: Oxford University Press.
- MALOUF, ROBERT. 1999. West Greenlandic noun incorporation in a monohierarchical theory of grammar. *Lexical and constructional aspects of linguistic explanation*, ed. by Gert Webelhuth, Jean-Pierre Koenig, and Andreas Kathol, 47–62. Stanford, CA: CSLI Publications.
- MALOUF, ROBERT. 2000. *Mixed categories in the hierarchical lexicon*. Stanford, CA: CSLI Publications.
- MEIER-BRÜGGER, MICHAEL. 2010. *Indogermanische Sprachwissenschaft*. 9th edn. Berlin: De Gruyter.
- MITHUN, MARIANNE. 1984. The evolution of noun incorporation. *Language* 60.4.847–93.
- MUGANE, JOHN M. 2003. Hybrid constructions in Gikūyū: Agentive nominalizations and infinitive-gerund constructions. *Nominals: Inside and out*, ed. by Miriam Butt and Tracy Holloway King, 235–65. Stanford, CA: CSLI Publications.
- NORDE, MURIEL. 2009. *Degrammaticalization*. Oxford: Oxford University Press.
- NORDE, MURIEL. 2010. Degrammaticalization: Three common controversies. *Grammaticalization: Current views and issues*, ed. by Katerina Stathi, Elke Gehweiler, and Ekkehard König, 123–51. Amsterdam: John Benjamins.
- NORDLINGER, RACHEL, and LOUISA SADLER. 2007. Verbless clauses: Revealing the structure within. *Architectures, rules and preferences: Variations on themes by Joan Bresnan*, ed. by Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling, and Chris Manning, 139–60. Stanford, CA: CSLI Publications.
- OLSEN, SUSAN. 2000. Composition. *Morphologie: Ein internationales Handbuch zur Flexion und Wortbildung, 1. Halbband/Morphology: An international handbook on inflection and word-formation, vol. 1*, ed. by Geert Booij, Christian Lehmann, and Joachim Mugdan, 897–915. Berlin: Mouton de Gruyter.
- ØRSNES, BJARNE. 1996. Prominence relations and a-structure—Evidence from Danish synthetic compounding. *Proceedings of the LFG '96 Conference*. Online: <http://csli-publications.stanford.edu/LFG/LFG1-1996/lfg96oersnes.pdf>.
- PADROSA-TRIAS, SUSANNA. 2010. Are there coordinate compounds? *Morphology and diachrony: Online proceedings of the Mediterranean Morphology Meeting* 7.98–111. Online: <http://www3.lingue.unibo.it/mmm2/wp-content/uploads/2013/98-111-Padrosa-Trias.pdf>.
- PAYNE, JOHN, and RODNEY HUDDLESTON. 2002. Nouns and noun phrases. *The Cambridge grammar of the English language*, ed. by Rodney Huddleston and Geoffrey K. Pullum, 323–523. Cambridge: Cambridge University Press.
- PAYNE, JOHN; RODNEY HUDDLESTON; and GEOFFREY K. PULLUM. 2010. The distribution and category status of adjectives and adverbs. *Word Structure* 3.1.31–81.
- POLLOCK, SHELDON I. 2001. The death of Sanskrit. *Comparative Studies in History and Society* 43.2.392–426.
- POLLOCK, SHELDON I. 2006. *The language of the gods in the world of men*. Berkeley: University of California Press.
- POSER, WILLIAM J. 1992. Blocking of phrasal constructions by lexical items. *Lexical matters*, ed. by Ivan A. Sag and Anna Szabolcsi, 111–30. Stanford, CA: CSLI Publications.
- POSTAL, PAUL. 1969. Anaphoric islands. *Chicago Linguistic Society* 5.205–39.
- RAMAT, PAOLO. 1992. Thoughts on degrammaticalization. *Linguistics* 30.549–60.
- RAMCHAND, GILLIAN CATRIONA. 2008. *Verb meaning and the lexicon: A first-phase syntax*. Cambridge: Cambridge University Press.
- ROSÉN, VICTORIA. 1996. The LFG architecture and ‘verbless’ syntactic constructions. *Proceedings of the LFG '96 Conference*. Online: <http://csli-publications.stanford.edu/LFG/LFG1-1996/lfg96rosen.pdf>.
- SADLER, LOUISA, and DOUG ARNOLD. 1994. Prenominal adjectives and the phrasal/lexical distinction. *Journal of Linguistics* 30.187–226.
- SADLER, LOUISA, and ANDREW SPENCER. 2001. Syntax as an exponent of morphological features. *Yearbook of Morphology* 2000.71–96.
- SADOCK, JERROLD M. 1980. Noun incorporation in Greenlandic: A case of syntactic word formation. *Language* 56.2.300–319.
- SADOCK, JERROLD M. 1986. Some notes on noun incorporation. *Language* 62.1.19–31.

- SCALISE, SERGIO, and EMILIANO GUEVARA. 2006. Exocentric compounding in a typological framework. *Lingue e linguaggio* 5.2.185–206.
- SCHINDLER, JOCHEM. 1997. Zur internen Syntax der indogermanischen Nominalkomposita. *Berthold Delbrück y la sintaxis indoeuropea hoy*, ed. by Emilio Crespo and José Luis García Ramón, 537–40. Madrid: Ediciones de la Universidad Autónoma de Madrid.
- SELKIRK, ELISABETH O. 1982. *The syntax of words*. Cambridge, MA: MIT Press.
- SNYDER, WILLIAM. 2001. On the nature of syntactic variation: Evidence from complex predicates and complex word-formation. *Language* 77.2.324–42.
- SPENCER, ANDREW. 2003. A realizational approach to case. *Proceedings of the LFG '03 Conference*, 387–401. Online: <http://csli-publications.stanford.edu/LFG/8/lfg03spencer.pdf>.
- SPENCER, ANDREW. 2005. Case in Hindi. *Proceedings of the LFG '05 Conference*, 429–46. Online: <http://csli-publications.stanford.edu/LFG/10/lfg05spencer.pdf>.
- SPENCER, ANDREW. 2006. Syntactic vs. morphological case: Implications for morphosyntax. *Case, valency and transitivity*, ed. by Leonid Kulikov, Andrej Malchukov, and Peter de Swart, 3–22. Amsterdam: John Benjamins.
- SPENCER, ANDREW. 2013. *Lexical relatedness: A paradigm-based approach*. Oxford: Oxford University Press.
- SPEYER, JACOB SAMUEL. 1896. *Vedische und Sanskrit-Syntax*. Strassburg: K. J. Trübner.
- STUMP, GREGORY T. 2001. *Inflectional morphology: A theory of paradigm structure*. Cambridge: Cambridge University Press.
- SULGER, SEBASTIAN. 2009. Irish clefting and information-structure. *Proceedings of the LFG '09 Conference*, 562–82. Online: <http://csli-publications.stanford.edu/LFG/14/papers/lfg09sulger.pdf>.
- SURYAKANTA (ed.) 1970. *Varadāmbikā-pariṇaya-campūh of Tirumalāmbā, with English translation, notes & introduction*. Varanasi: Chowkhamba Sanskrit Series Office.
- SVENONIUS, PETER. 2011. *Spanning*. Tromsø: University of Tromsø, MS.
- TEN HACKEN, PIUS. 1994. *Defining morphology: A principled approach to determining the boundaries of compounding, derivation, and inflection*. Hildesheim: Olms.
- TOIVONEN, IDA. 2003. *Non-projecting words: A case study of Swedish verbal particles*. Dordrecht: Kluwer.
- TUBB, GARY A., and EMERY R. BOOSE. 2007. *Scholastic Sanskrit: A manual for students*. New York: The American Institute of Buddhist Studies at Columbia University in the City of New York.
- VAN DER AUWERA, JOHAN. 2002. More thoughts on degrammaticalization. *New reflections on grammaticalization*, ed. by Ilse Wischer and Gabriele Diewald, 19–29. Amsterdam: John Benjamins.
- VINCENT, NIGEL, and KERSTI BÖRJARS. 2010. Grammaticalization and models of language. *Gradience, gradualness and grammaticalization*, ed. by Elizabeth Closs Traugott and Graeme Trousdale, 279–99. Amsterdam: John Benjamins.
- WARD, GREGORY; RICHARD SPROAT; and GAIL MCKOON. 1991. A pragmatic analysis of so-called anaphoric islands. *Language* 67.3.439–74.
- WESCOAT, MICHAEL THOMAS. 2002. *On lexical sharing*. Stanford, CA: Stanford University dissertation.
- WESCOAT, MICHAEL THOMAS. 2005. English nonsyllabic auxiliary contractions: An analysis in LFG with lexical sharing. *Proceedings of the LFG '05 Conference*, 468–86. Online: <http://csli-publications.stanford.edu/LFG/10/lfg05wescoat.pdf>.
- WESCOAT, MICHAEL THOMAS. 2007. Preposition-determiner contractions: An analysis in optimality-theoretic lexical-functional grammar with lexical sharing. *Proceedings of the LFG '07 Conference*, 439–59. Online: <http://csli-publications.stanford.edu/LFG/12/papers/lfg07wescoat.pdf>.
- WESCOAT, MICHAEL THOMAS. 2009. Udi person markers and lexical integrity. *Proceedings of the LFG '09 Conference*, 604–22. Online: <http://csli-publications.stanford.edu/LFG/14/papers/lfg09wescoat.pdf>.
- WHITNEY, WILLIAM DWIGHT. 1896. *A Sanskrit grammar: Including both the Classical language, and the older dialects, of Veda and Brahmana*. 3rd edn. Leipzig: Breitkopf & Härtel.
- WILLIAMS, EDWIN. 2003. *Representation theory*. Cambridge, MA: MIT Press.

ZWICKY, ARNOLD M., and GEOFFREY K. PULLUM. 1983. Cliticization vs. inflection: English *n't*. *Language* 59.3.502–13.

Centre for Linguistics & Philology
University of Oxford
[john.lowe@ling-phil.ox.ac.uk]

[Received 14 January 2015;
revision invited 7 April 2015;
revision received 24 April 2015;
accepted 28 April 2015]