Uta Reinöhl*

What are and what aren't complex nominal expressions in flexible word order languages

Abstract: This paper tackles the challenge of how to identify multi-word (or "complex") nominal expressions in flexible word order languages including certain Australian languages and Vedic Sanskrit. In these languages, a weak or absent noun/adjective distinction in conjunction with flexible word order make it often hard to distinguish between complex nominal expressions, on the one hand, and cases where the nominals in question form independent expressions, on the other hand. Based on a discourse-based understanding of what it means to form a nominal expression, this paper surveys various cases where we are **not** dealing with multi-word nominal expressions. This involves, in particular, periphery-related phenomena such as use of nominals as free topics or afterthoughts, as well as various kinds of predicative uses. In the absence of clear morpho-syntactic evidence, all kinds of linguistic evidence are relied upon, including, in particular, information structure and prosody, but also derivational morphology and lexical semantics. In this way, it becomes frequently possible to distinguish between what are and what aren't complex nominal expressions in these languages.

Keywords: noun phrase, flexible word order, non-configurationality, discourse, prosody, Australian languages, Vedic Sanskrit

1 Introduction

This paper deals with the question of how to identify complex (i.e. multi-word) nominal expressions[1] in languages that show a lack of clear formal (i.e. morpho-syntactic) characteristics of such expressions. Languages which fall into this ca-

---

1 I use the more neutral term "expression" instead of "phrase" following Himmelmann (1997) because of the long tradition of associating the latter with strict syntactic rules of adjacency and

*Corresponding author: Uta Reinöhl, Language Typology, Department of English and Linguistics, Johannes Gutenberg-Universität Mainz, Mainz 55122, Germany, E-mail: uta.reinoehl@uni-mainz.de.

tegory are so-called "free word order", "flexible word order" or "non-configura-
tional" languages (e.g. Hale 1983; Austin and Bresnan 1996) as found for
instance in Australia. As for the label, while none is ideal, I will here mostly
use the theoretically most neutral and empirically most adequate label of
"flexible word order". These languages are characterized by an unusually flex-
ible syntax, allowing not only for main constituents to appear in different orders
and for "zero" arguments of various types – phenomena shared by many
languages – but also for considerable syntagmatic flexibility inside complex
nominal expressions, even including discontinuity. Numerous flexible word
order languages are also known for showing no or only a weak noun-adjective
distinction, that is, members of a "flexible" word class of nominals that may be
entity- and/or property-denoting display the same inflectional potential (if
applicable) and have roughly the same syntactic distribution.[2] It is this language
type which is at issue here, represented not only by certain Australian languages
including Warlpiri (e.g. Hale 1983; Nash 1986 [1979]; Simpson 1991),
Nunggubuyu (Heath 1986), or Jiwarli (e.g. Austin 2001), but also by early Indo-
European languages, in particular Vedic Sanskrit (e.g. Schäufele 1991; Reinöhl
2016), among others.

The combination of considerable syntactic flexibility and a weak noun-
adjective distinction creates a situation where it can be difficult to determine
whether multiple nominals in a sentence constitute one complex nominal
expression or whether they should rather be analyzed as separate nominal
expressions. This is the case even when all elements in question bear the
same morphological marking (e.g. for case, gender and/or number) as will
become clear. By contrast, in languages such as English, morphological and
particularly syntactic distributional clues make it easier to recognize whether we
are dealing with one complex expression or several separate ones. For example,
word order alone coerces a distinction between the use of *red* as an attribute or
secondary predicate in *I'm going to paint the red door* vs. *I'm going to paint the
door red*; only in the first instance are we dealing with a multi-word or "com-
plex" nominal expression consisting of *red* and *door*.

The challenge of identifying whether nominals belong together and form a
complex expression or not is compounded by the fact that at least some flexible
word order languages employ nominal elements in a considerable range of

obligatoriness as exhibited, for instance, by English NPs. It is these characteristics precisely that
do not apply in the languages at issue in this paper.
2 The family of morphologically nominal elements may also extend to, e.g., pronominals,
numerals and quantifiers. Note that I focus in this paper only on relations between syntactic
words (and not between higher-level constructions).

syntactic functions and with great frequency. In Vedic Sanskrit, for instance, sentences often abound with morphologically nominal elements (nominals including entity- or property-denoting ones, and also pronominals, numerals, certain quantifiers, etc., which inflect for case, gender and number), which, however, by no means necessarily function as complex nominal expressions given a discourse-based understanding.

While this paper focuses on languages that lack clear morphosyntactic evidence for complex nominal expressions, it will be argued that clues from various levels of linguistic structure – ranging from derivational morphology and ordering preferences to, particularly, information-structural and prosodic information – nonetheless often make it possible to decide whether we are dealing with one or with several expressions. Central to this approach is a discourse-based understanding of what it takes to form a complex nominal expression following Himmelmann (1997), which will be discussed in more detail in Section 3. A consequence of this approach is that several construction types that have been included in some of the literature under the label of "discontinuous NPs", for instance the "split topicalizations" figuring prominently in (Fanselow and Féry 2006 (ms)), do not in fact constitute complex nominal expressions given the understanding adopted here. The importance of distinguishing between real and merely apparent complex nominal expressions along the lines proposed in this paper is emphasized by a count of construction types. In a sample of roughly 900 Vedic Sanskrit prose sentences from the *Maitrāyaṇī-Saṁhitā* (early 2nd millennium BC), each containing multiple nominals inflected for the same values, roughly two thirds of the sentences have to be excluded from consideration given a discourse-based approach.[3]

This paper both surveys the literature as well as expands on the state of the art regarding the scope of the research question, the languages considered, and the range of construction types at issue, offering a compact, qualitative survey. Firstly, the literature has so far focused primarily on discontinuous nominal expressions, which flexible word order languages are so very well-known for (e.g. Hale 1983; McGregor 1997; Schultze-Berndt and Simard 2012). However, the challenge of identifying complex nominal expressions arises not only in cases of discontinuity, but also when nominals are juxtaposed. In other words, rather than studying the difference between "real" and "apparent"

---

discontinuity (Schultze-Berndt and Simard 2012), I here consider the difference between "real" and "apparent" complex nominal expressions. Secondly, I expand on the languages for which the question of how to delimit real complex nominal expressions has been discussed, going beyond the Australian continent – where I focus on discussions of data from Warlpiri, Jaminjung and Gooniyandi for which the issue has been most explicitly raised – to include Vedic Sanskrit. Other ancient Indo-European varieties are not considered here, as most of them show a somewhat more tightly organized phrasal syntax in the nominal domain structured around articles or adpositions, such as e.g., in Ancient Greek and Latin. Thirdly and as a result of including Vedic Sanskrit, some additional construction types will be discussed. While Vedic Sanskrit is similarly "non-configurational" as some Australian languages (Reinöhl 2016: 33–40), it shows certain particularities which broaden the range of phenomena considered here. In McGregor (1997) on Gooniyandi and in Schultze-Berndt and Simard (2012) on Jaminjung – the two primary studies in which authors have reflected on how to delimit (discontinuous) real complex nominal expressions from apparent ones in Australian languages – the discussion focuses on certain types of afterthoughts, part-whole structures, and secondary predicates. This set of phenomena is expanded on here to include further types of periphery-related constructions besides afterthoughts (Sections 4.1 and 4.2), further types of predicative constructions in addition to secondary predicates (Section 4.3), and a brief comment on appositions (Section 4.4). By building on previous literature in this way, the collection of phenomena brought together here is likely to cover many pitfalls for erroneously assuming a complex nominal expression in "flexible word order" languages. At the same time, this list is likely not exhaustive given the manifold grammatical functions in which nominals appear in these languages. In considering this range of construction types it is important to take note of the fact that we are not dealing with particularly unusual constructions from a cross-linguistic vantage point; the particular interest arises from the challenge of identifying them in "flexible word order" languages given the less direct mapping of function onto form.

2 A lack of coercive morphosyntactic evidence

In a language like English where grammatical functions are often linked to specific phrasal slots, it is comparatively easy to determine from morphosyntactic indications alone whether two or more elements form a complex expression or not. For example, in the sentence *I'm going to paint the red door* it seems uncontroversial that *red* and *door* belong to one syntactic unit – an NP – a

statement that can be underscored empirically by phrase structure tests that target, e.g., permutation. It seems similarly uncontroversial that *red* and *door* do not belong to one nominal expression in the sentence *I'm going to paint the door red*. Here, *door* is an argument of the verb, whereas *red* contributes another "secondary" predicate which describes the change of state undergone by the door at the culmination point of the event encoded by the main verb *to paint*. The question of whether a nominal element is an attribute or an additional predicate or has yet another function, and how that element relates to other nominals with the same inflectional marking (if applicable) – i.e. whether it is part of a complex nominal expression or not – can be difficult to answer in a language where grammatical functions are not directly associated with syntactic slots. (Note, however, that at least robust correlations between slots and functions seem to be found in most, if not all, flexible word order languages). However, while displaying two different word orders, both of the following two examples from Jaminjung are analyzed by the authors as involving complex nominal expressions, because they identify one discourse referent respectively:

(1) Jaminjung (non-Pama-Nyungan)
     mulanggirrng mali *hambag*   *nganthi-yu*
     fierce   thing be.nuisance 2SG > 3SG-say/do.PST
     '(Try and run away!) You disturbed a **dangerous thing**!' (From a scene in the Frog Story where a dog jumps up at a beehive and eventually causes it to fall down and the bees to chase him.)
     [Schultze-Berndt and Simard 2012: 1020]

(2) Jaminjung (non-Pama-Nyungan)[4]
     ^**mulanggirr** *ngantha-ma-ya*   **wirib** \
     fierce   2SG > 3SG-have-PRS   dog
     'You have a ***dangerous* dog**!'   [Schultze-Berndt and Simard 2012: 1035]

On the other hand, one and the same word order may lend itself to different functional interpretations, as famously argued in Hale (1983) for Warlpiri (see also Simpson 1991). Hale, focusing on discontinuity, discusses differences between what he calls "merged" and "unmerged" interpretations.

---

4 A focus reading on the adnominal modifier is a correlate of discontinuous expressions in Australian languages and in Vedic Sanskrit (Reinöhl 2020), as well as in languages that show a more rigid NP structure such as, e.g., Polish (Siewierska 1984).

> I take it to be a fact of Warlpiri that a sentence like [...] below is open to at least two interpretations:
>
> *Maliki-rli ø-ji yarlku-rnu wiri-ngki*
> dog-ERG PERF-1OBJ bite-PST big-ERG
> 'The/a big dog bit me.'
> 'The/a dog bit me and it was big.'
> [...] On one interpretation, the expression *wiri-ngki* (big-ERG) is taken as a modifier of *maliki-rli* (dog-ERG), forming with it a unit expression [...] This is the 'merged' interpretation. [...] On another interpretation [...], *wiri-ngki* is simply predicated of *maliki-rli*. (Hale 1983: 38)

One way of interpreting this passage is that, in the case of a "merged" interpretation, we are dealing with a single nominal expression consisting of *maliki-rli … wiri-ngki*, while in the "unmerged" interpretation, we are dealing with two separate expressions, with *wiri-ngki* serving as an additional predicate besides the main predicate. Crucially, whether two (or more) nominals are conceived of as constituting one expression or two (or more) separate ones has implications for every analysis that seeks to describe any formal or functional aspect of the expression (or the expressions) in question.

The challenge of identifying how two (or more) nominals relate to each other is not only due to the less direct mapping of functions onto slots, but is also in part due to the absent or weak noun-adjective distinction in numerous flexible word order languages (see Dixon 1980: 272–275, Nordlinger 2014: 237–238 on Australian languages, Bhat 1994 on Sanskrit, and Baker 2001a more generally on "non-configurational" languages). I illustrate this point with the Vedic Sanskrit words *svargá-* 'heaven, heavenly'. In (3) *svargá-* is used as a referring expression denoting the entity 'heaven' in a similar way to *loká-* 'world' in (4). In (5) these two nominal elements combine into one construction with *svargá-* now being used as an attributive modifier. Other than contextual evidence – *svargám … lokám* paraphrases *svàr*, another word for 'heaven', which is used in the preceding verse in the original text – no formal sign distinguishes the use of *svargá* in (3) and (5).

(3) Vedic Sanskrit
     svargá   *u*   *tvám ápi*   *mādayāse*
     heaven.LOC.SG.M PRT 2SG   EMPH enjoy.SUBJCT.MID.2SG
     'May you enjoy yourself in **heaven**'   [RV 10.95.18]

(4)   Vedic Sanskrit
    asáu   *vái*   **lokáḥ**   *pratnám*
    DEM.NOM.SG.M   PRT   world.NOM.SG.M   former.NOM.SG.N
    '(what was) before is that (i.e. the ulterior) **world**'   [MS I 5,5(1)]

(5)   Vedic Sanskrit
    svargám   *evá* **lokám**   *eti*
    heaven.ACC.SG.M PRT world.ACC.SG.M go.PRS.3SG
    'he goes to the **heavenly world**'   [MS I 4,7(2)]

In English, knowing whether an element is an adjective or a noun restricts the grammatical functions that it might have. For example, adjectives can appear as secondary predicates, while nouns are not usually employed in that function (unless they receive special marking, that is, e.g. *she signed the papers **as the director***). This restriction of possible interpretations given a noun-adjective distinction is lacking in languages of the type discussed here. Note that this is not to say that all nominals are completely flexible as to property- or entity-denoting use in the languages discussed here. Many elements show at least a preference for either one of the usages or are even attested exclusively in only one of them (see Nash 1986 [1979]: 15 on Warlpiri; McGregor 1990: 256 on Gooniyandi). For instance, while Vedic *svargá-* 'heaven, heavenly' shows such semantic flexibility, *loká-* 'world' does not. In some cases, there may also be formal indications for a particular use, as for instance in Vedic Sanskrit where derivational morphology or gender marking may indicate a property- or entity-denoting use of a nominal.[5]

    In view of the lack of clear formal clues for whether we are dealing with one complex nominal expression or with several, some early analyses opted for the somewhat radical move of considering free nominals and pronominals to be adjuncts or secondary predicates in principle, and instead regarded pronominal clitics in second ("Wackernagel") position as the actual arguments (e.g. Jelinek 1984; Laughren 1989; Baker 2001b). These proposals and especially the best-known one, Jelinek's (1984) Pronominal Argument Hypothesis, were soon debunked, however (e.g. Austin and Bresnan 1996 drawing on Simpson 1991). Apart from several aspects of Warlpiri grammar that Jelinek's analysis could not account for, possibly the strongest argument was that not all "non-configurational" languages

---

5 Given the lack of strongly grammaticalized adnominal function words (Himmelmann 1997) such as articles or argument-marking adpositions, there is also no evidence from their obligatory repetition in the languages at issue here. Thus, when dealing with two separate nominal expressions, a function word may have to be repeated in languages such as English or German, for instance in the case of an afterthought (*Please give me **the** tomato, **the** really ripe one!, Gib mir bitte **die** Tomate da, **die** ganz reife!*).

have an auxiliary complex with pronominal clitics in second position, an example of which is Jiwarli (Austin 2001).[6] While most authors now recognize that complex nominal expressions are a feature of "non-configurational" languages in the same way as of other languages (cp., e.g. Schultze-Berndt and Simard 2012; Louagie and Verstraete 2016), the empirical challenge of identifying when we are dealing with a complex nominal expression and when we are not remains.

3 Discourse-based evidence and the role of prosodic packaging

Given the lack of definite morphosyntactic clues, it is necessary to adopt a discourse-based understanding of what it takes for two or more syntactic words to form a complex nominal expression. Himmelmann (1997: 111, 117–119), following work by John DuBois and Sandra Thompson, argues that a complex nominal expression can be identified by the component elements sharing a discourse-functional role, which may be a "tracking" use (corresponding to what is normally referred to as "referential" uses where participants are introduced or tracked) or a "non-tracking" use (e.g., where a nominal expression is used predicatively). When two or more elements share such a discourse-functional use – for instance, where they work together in introducing a new participant into the discourse – we are dealing with a complex nominal expression. If they do not share a role in discourse, e.g. when one of them introduces a participant while the other one is used predicatively, we are not dealing with one complex nominal expression, but with separate ones. Thus, Hale's (1983) "merged" use may be interpreted as corresponding to nominals sharing a discourse-functional role in contrast to an "unmerged" interpretation where they do not. From this perspective, the difference between, e.g., Vedic Sanskrit and English, is that while in the former, the sharing of a discourse function does not map in a particularly direct way onto morphological or syntagmatic traits, it is much more directly tied to formal structure in the latter. (Therefore, in English,

---

6 A crucial theoretical step in several later analyses of Australian "non-configurational" languages was to divorce phrase structure (as traditionally understood) from functional structure, an essential pillar in particular in the framework of Lexical Functional Grammar (LFG). In fact, the LFG framework was in part developed in order to accommodate syntactic evidence from "non-configurational" languages (Bresnan 2001: 5ff.). For Australian languages, LFG studies include, e.g., Simpson (1991), Austin and Bresnan (1996) or Nordlinger (1998). While the present study is not framed in LFG, the same assumption that functional structure does not necessarily equate in any direct way to phrasal structure holds here, too.

morphosyntactic clues alone can be sufficient to identify whether we are dealing with a complex nominal expression or not.)

For illustration, in the Vedic Sanskrit example (5) above, *svargá-* 'heaven' *and loká-* 'world' together establish the 'heavenly world' as a goal sharing a "non-tracking" ("orienting") use (Himmelmann 1997: 118). To give another example, in the following example from Jaminjung, context suggests that we are dealing with two elements sharing a discourse-functional use. The two highlighted elements refer to an already introduced participant as a 'dangerous thing' (an example of a "tracking" use).

(6)   Jaminjung (= ex. (1))
      [**mulanggirrng mali**] *hambag*   *nganthi-yu*
      fierce   thing   be.nuisance   2SG > 3SG-say/do.PST
      '(Try and run away!) You disturbed a **dangerous thing**!'
                                         [Schultze-Bernd and Simard 2012: 1020]

Wallace Chafe (e.g. Chafe 1994) and others have emphasized the importance of prosodic phrasing for information packaging. Schultze-Berndt and Simard (2012) argue for Jaminjung that the prosodic domain of the "intonation phrase", intermediate between prosodic word and intonation unit, is the domain within which nominals forming complex expressions are found – continuous or discontinuous (see also McGregor 1997 for a study on Gooniyandi that takes into account prosodic phrasing).[7] By contrast, an afterthought, for instance, is not found in the same intonation phrase as the previously mentioned, co-referential nominal. The Jaminjung intonation phrase is demarcated by "a slight pitch reset at the left edge and lowering of pitch and final syllable lengthening at the right edge" (Schultze-Berndt and Simard 2012: 1022).

Focusing on discontinuous nominal expression, one type of syntactic pattern found in some Australian languages (see Louagie and Verstraete 2016: 49–54) has the following structure. One nominal stands initially, followed by elements such as a verbal lexeme, TAM morphemes and/or pronominal clitics in second position, and the other nominal coming in third position, as illustrated by some examples in this paper including example (7) from Jaminjung. Schultze-Berndt and Simard (2012) indicate with the back-slash that the discontinuous

---

7 Heath (1986) argues for Nunggubuyu that prosodic phrasing is not helpful in identifying domains because "individual NP-segments are typically of three to six syllables each and each normally has its own independent intonation pattern" (Heath 1986: 377). It is a question beyond the scope of this study whether there are other "non-configurational" languages besides Nunggubuyu where we may not be able to rely on evidence from prosodic phrasing.

nominal expression is found inside one intonation phrase. The authors outline that the 'dog' is a previously introduced referent, which is then taken up again as a '*dangerous* dog'. There is no obvious syntactic, information-structural, or prosodic evidence that would suggest a different interpretation (e.g., where one of the elements would be a secondary predicate or afterthought).

(7)   Jaminjung (= ex. (2))
       ^***mulanggirr*** *ngantha-ma-ya*   ***wirib*** \
       fierce   2SG > 3SG-have-PRS dog
       'You have a ***dangerous*** **dog**!'   [Schultze-Berndt and Simard 2012: 1035]

In the case of Vedic Sanskrit we are dealing with a language for which there is no recorded evidence (disregarding modern-day recordings of recitations, despite the common assumption of their considerable historical faithfulness due to the long and uninterrupted history of oral transmission). However, there are both prosodic and syntactic clues pointing to the external boundaries of a core "domain" that resembles the one in which (continuous or discontinuous) complex nominal expressions are found in Jaminjung, Gooniyandi and possibly other Australian languages. The left clausal edge in Vedic can in many cases be inferred as being the single syntactic word preceding "Wackernagel" clitics, which are pronominals and/or discourse particles (Hock 1982, 1996; Hale 1996). At the right clausal edge (or close to it), we typically find the finite verb which is "unaccented" in main clauses (i.e. has the low pitch in the Vedic pitch-accent system).[8] If discontinuous, a commonly found pattern is that the nominal expression brackets the Wackernagel position, or the Wackernagel position plus an additional lexeme, as in (8). Here, *ubháyīr … devátā* together introduce a referent with a contrastive reading on the modifier, i.e. '*both* kinds of gods', similar to the Jaminjung example (7) above. This sentence stands paragraph-initially and the two nominals introduce the new topic, which underscores that we are indeed dealing with one complex expression.

---

8 The phenomenon of the de-accented finite verb in the matrix clause has been interpreted as a reification of the cross-linguistic phenomenon that pitch contours tend to fall over the stretch of declarative statements, given SOV order and relative before correlative clause order in Vedic, cp. Delbrück (1878: 76–78); Klein (1992). De-accentuation can be gathered from the lack of an acute accent on a vowel in the verb form in question; in (8), the enclitic finite verb form *śaṁsante* is by convention written as one word with the preceding local adverb *ā́* 'to, at', a lexicalized combination.

(8)   Vedic
      ubháyīr   vā́   agnihotríṇi   **devátā**
      both_kinds.NOM.PL.F PRT Agni_sacrificer.LOC.SG.M deity.NOM.PL.F
      ā́śaṃsante   yā́bhyaś   ca   juhóti   yā́bhyaś   ca   ná
      hope.3PL   REL.DAT.PL.M and   sacrifice.3SG REL.DAT.PL.M and   not
      '**Both** kinds of gods hope for the Agni-sacrificer, (those) for whom
      (he) sacrifices and (those) for whom (he) (does) not.'   [MS I 5,11 (2)]

If a nominal precedes the core clausal domain (i.e. stands in the left periphery
as, e.g., a free topic) or follows it (i.e. stands in the right periphery as, e.g., an
afterthought), it does not normally share a discourse role with a nominal inside
the core domain. The structural differences between the patterns found in some
Australian languages (as well as among them) and Vedic notwithstanding – e.g.
regarding what kinds of elements occupy second position – the evidence sug-
gests that there are close parallels with regard to domain structures across these
unrelated, flexible word order languages (see Reinöhl 2020 for a detailed com-
parative study). Domain boundaries thus established can be of crucial help in
distinguishing between which elements are likely contenders for together form-
ing complex nominal expressions and which ones are not.

4 When nominals do not belong together

The following sub-sections discuss what seems like the major cases where
several nominals, even if bearing the same morphological marking, do not in
fact share a discourse-functional use. This includes cases where nominals are
distributed over different domains (in particular, one nominal being placed in a
domain that contains the clausal core while the other one[9] is placed in the right
or left periphery) as well as cases where context suggests that the nominals in
question do not share a discourse-based role even when occurring in the same
domain.[10] I also add some brief remarks on appositional structures. It is worth
emphasizing that all constructions discussed here appear also in languages
where functions are more closely tied to specific morphosyntactic characteristics
and that there are long and rich research traditions on each of them, which,

---

9 For simplicity's sake, I restrict myself to cases involving only two syntactic words.
10 I focus on cases where at least one of the nominals occurs in argument function. However,
other combinations are also in principle possible, but seem less frequent. For instance, a
secondary predicate and a nominal main predicate could be mistaken as forming a complex
expression.

however, will not be particularly elaborated on here. The point is thus not that we are dealing with unusual construction types, but that it is possible to identify them in the languages at issue here also, and, based on this, to be more discerning in admitting "real" complex nominal expressions, rather than include all sorts of instances only because two or more nominals share the same morphological marking.

4.1 Right periphery

Right periphery-related phenomena include right dislocations and after-thoughts. In the case of "right dislocations" (e.g. Lambrecht 1994), a topical argument may or may not be realized pronominally inside the clausal core, while a lexical element specifies the referent in the right periphery. Right-dislocated nominals are typically separated from the preceding clause by a pause, and they are characterized by specific prosodic contours (Kalbertodt et al. 2015); think of a realization with a flat intonation contour of, e.g. *I saw him, Peter*. In the same way as in more configurationally organized languages, such evidence for domain boundaries is relevant for the identification of right periphery-related constructions in the languages at issue here. The prosodic break is marked by a slash by McGregor in the following Gooniyandi example.

(9) Gooniyandi
     goobardiya /   *ngila /*   *joonyoo /*   **niyaji-ya**   *yard*   *ngaragbinmi /*
     Goobardiya   east   Joonyoo   this-LOC   yard   they:made:it
     mayari-ya /
     house-LOC
     'At Goobardiya and east at Joonyoo, **there** they made yards, **at the homesteads**.'   [McGregor 1997: 101]

In the Vedic example (10), showing a right dislocation, the clause-initial verb 'to ignite' activates 'the fire' as part of its semantic frame. Being activated, the latter is realized pronominally inside the clause, and the lexical information is added only later in the right periphery.[11]

---

11 The normal verb-final order is reversed in this example.

(10) Vedic Sanskrit
     *indhe   ha   vā **etad**   adhvaryuḥ*
     indhe   ha   vā etad   adhvaryuḥ
     ignite.3SG PRT PRT DEM.ACC.SG.N Adhvaryu.NOM.SG.M
     *idhmenā**gniṃ***
     idhmena agniṃ
     fuel.INS.SG.N fire.ACC.SG.M
     '(He) lights **it**, the Adhvaryu, with a kindling, **that fire**.'
                         [ŚaB 1.3.5.1, adapted from Schäufele 1991: 44]

Afterthoughts involve new information that is introduced in the right periphery, specifying a referent that is lexically referred to in the core clause (see Schultze-Berndt and Simard 2012: 1025–1028 on Jaminjung). Compare the following examples from Jaminjung (11), Gooniyandi (12) and Vedic Sanskrit (13).[12]

(11) Jaminjung
     buyud = *biyang jabul* = *ni*   *burr-angu* = *rrgu* = *rndi* \
     sand = SEQ   shovel = ERG/INS 3PL > 3SG-get/handle.PST = 1SG.OBL = EVID
     buj-mawu   ***buyud*** \
     bush-dweller sand
     'They got **sand** for me with a shovel, **the bush kind of sand**.'
                         [Schultze-Berndt and Simard 2012: 1027]

(12) Gooniyandi
     *nganyi-ngga **jawangari** nyaglooni* / ***nyamani***
     I-ERG   kangaroo I:speared:it big
     'I speared a **kangaroo**, **a big one**.'   [McGregor 1997: 101]

(13) Vedic Sanskrit
     bṛhadúkṣaḥ   ***marútaḥ***   *viśvávedasaḥ*
     richly_raining.NOM.PL.M Marut.NOM.PL.M owning_all_treasure.NOM.PL.M
     právepayanti párvatān   *ádābhyāḥ*
     shake.3PL   mountain.ACC.PL.M invulnerable.NOM.PL.M
     'The **richly-raining Maruts**, **owning all treasure**, shake the mountains, **the invulnerable ones**.'
                        [RV 3.26.4, adapted from Schäufele 1991: 58]

---

12 McGregor (1997) includes "afterthoughts" as types of discontinuity, for a critique of which see Schultze-Berndt and Simard (2012: 1028). Also see Schultze-Berndt and Simard (2012: 1026) for references to literature on afterthoughts in Australian languages.

In addition to or alternatively to information-structural conditions, it remains a question for future research whether or to what extent phonological and/or morphological heaviness effects may also be implicated in triggering the realization of nominals in the right periphery in the languages discussed here (see Delbrück 1878: 39 for some remarks regarding Vedic Sanskrit as well as remarks in Reinöhl 2020).

4.2 Left periphery

There are at least two types of cases in which a nominal placed in the left periphery may be mistaken as forming a complex expression with an element inside the clausal core, both of which involve phenomena that have been subsumed under the label of "topicalization". Firstly, "split topicalization" constructions have been argued to be a type of discontinuous NP structure by some authors. Thus, Fanselow and Féry (2006 (ms)) prominently discuss "split topicalization" constructions in their cross-linguistic overview of "discontinuous NPs", i.e. constructions such as German *Schlösser hat Peter schon viele besichtigt,* '(As for) castles (,) Peter has visited many' (Fanselow and Féry 2006 (ms): 1). The authors themselves, however, in accordance with the literature, state at several points in their paper that such split topicalizations involve separate NPs, with separate roles in discourse (e.g. Fanselow and Féry 2006 (ms): 19–20 and elsewhere; see also Nolda 2008; Schultze-Berndt and Simard 2012: 1047).[13] In information-structural terms, the first element is analyzed by some authors as introducing a topic expression and the second one an expression in focus (cp. Schultze-Berndt and Simard 2012: 1047–1048 for discussion).

Given the prominent role accorded to split topicalizations, the claims in Fanselow and Féry (2006 (ms)) regarding the number of languages allegedly allowing for discontinuous NPs may not be taken at face value if a discourse-based understanding of what it takes to form a complex nominal expression is adopted.[14] In fact, it is subject to future research whether or to what existence split topicalization play a major role in "non-configurational" languages (in the

---

13 By contrast, the sub-type of "cohesive" NP identified by Fanselow and Féry (2006 (ms)) seems to abide by traditional criteria of what may constitute a complex nominal expression.
14 A formal indication that we are dealing with a split topicalization construction can be a lack of morphological agreement (see Schultze-Berndt and Simard 2012: 1047 discussing cases such as German *Zeitungen liest er nur eine – die* TAZ, '(As for) newspapers (,) he reads only one – the TAZ').

strict, Australianist sense) at all. Schultze-Berndt and Simard (2012: 1048) state for instance that there is no such construction in their Jaminjung corpus.

Another type where elements may be mistaken as forming a complex expression involves free topics in the left periphery. Compare the following example from Jaminjung, for which Schultze-Berndt and Simard (2012: 1048) contemplate an analysis as a topic-comment structure based on information-structural and prosodic evidence, pending more detailed research. Despite surface continuity, we would not be dealing with a complex nominal expression in such cases.

(14)   Jaminjung
       *[jarlig = biya]*ₜₒₚ   *[bardawurru]*ꜰₒᴄ   *gani-ma-ya* \
       child = SEQ   many   3SG > 3SG-have-PRS
       '(**As for) children**, she has **many** ....'
                                       [Schultze-Berndt and Simard 2012: 1048]

The next two examples from Vedic, both found in the same paragraph, are structured in an exactly parallel fashion. Despite appearances if only example (15) is taken into consideration, the topic in the left periphery does not form a complex nominal expression with the demonstrative that follows the Wackernagel position. This analysis is suggested when the lack of gender agreement in example (16) is taken into account.[15] It becomes clear that the gender of the demonstrative *eṣò* is determined in both examples by *valmī́kaḥ* 'termite mound', which functions as the subject of the relative clause, being co-referential with the referent of the demonstrative heading the correlative clause. *Ráso* 'sap' and *ū́rg* 'invigoration' precede the respective complex clauses as free topics.

(15)   Vedic Sanskrit
       ráso   *vā́*   **eṣò**   *'syā́*   *údaiṣad*
       sap.NOM.SG.M   PRT   DEM.NOM.SG.M   DEM.GEN.SG.F   rise.IMPF.3SG
       yád   *valmī́kaḥ*
       REL.NOM.SG.M   termite_mound.NOM.SG.M
       '**Sap** – **that** (which) arises from this one [i.e. the earth] (, that is) what
       the termite mound (is).'   [MS I 6,3(4)]

---

15 For the larger context and an interpretation of this passage see Amano (2009: 213–214).

(16) Vedic Sanskrit

ū́rg	*vā́*	***eṣò***	*'syā́*	*údaiṣad*
invigoration.NOM.SG.F	PRT	DEM.NOM.SG.M	DEM.GEN.SG.F	rise.IMPF.3SG

yád	*valmī́kaḥ*
REL.NOM.SG.M	termite_mound.NOM.SG.M

'**Invigoration** – **that** (which) arises from this one [i.e. the earth] (, that is) what the termite mound (is).'   [MS I 6,3(4)]

In Section 3, it was argued that the initial position directly preceding the Wackernagel slot is normally part of the core clause in Vedic Sanskrit. The previous two examples show that this is not an absolute rule, suggesting that a comprehensive analysis that takes into account not only, e.g., word order, but also evidence from other levels of linguistic structure, is required in languages of the type discussed here.[16]

4.3 Various types of nominal predicates

Nominal predicates of various kinds are frequently a reason for finding several nominals with the same morphological marking in a sentence. In previous studies, secondary predicates in particular have been identified as being easily mistaken for attributive modifiers in different Australian languages (McGregor 1997: 102–103; Schultze-Berndt and Simard 2012: 1028–1030; for a study focusing on secondary predicates see Simpson 2006 on Warlpiri). As already discussed in Section 2, secondary predicates do not however form a nominal expression together with a nominal in argument (or adjunct) function, but constitute an additional predication beside the main predicate. The secondary predicate adds information regarding a referent in argument (or adjunct) role which holds in the time frame set by the main predicate or at its culmination point (see Himmelmann and Schultze-Berndt 2006). Thus, in the examples below, fruits are ripe when falling down[17], and the cows and

---

16 It is subject to investigation how frequent this construction type is in Vedic. In similar, but more easily discernible, instances, the case marking of the free topic does not match the one of the clause-internal pronominal (if present) or the verbal agreement (i.e. non-nominative cases do not normally control agreement), or the free topic precedes the whole clause including the lexeme hosting Wackernagel elements (see Schäufele 1991: 66–67 for examples).

17 Schultze-Berndt and Simard (2012: 1029) outline how an attributive reading, which would suggest the introduction of a new topic 'ripe fruit', does not seem plausible as the falling of fruit had already been discussed. For suggestions regarding prosodic characteristics of secondary predicates in Australian languages see references in Schultze-Berndt and Simard (2012: 1030).

wishes (actually and metaphorically) walk away in a state of not having been milked.[18]

(17) Jaminjung
      jarlag = _biyang_   _jag_   _ga-rdba-ny_   **_mangarra_** \
      good = SEQ   go.down   3SG-fall-PST   plant.food
      'The **fruit** fell down **ripe**.'   [cf. Schultze-Berndt and Simard 2012: 1030]

(18) Vedic Sanskrit
      _tā́_   _yáthā_   **_dhenávó_**   **_'dugdhā_**
      DEM.NOM.PL.F   as   COW.NOM.PL.F   un_milked.PPP.NOM.PL.F
      apakrā́manty   _evám_   _asmād_   **_āśíṣó_**
      walk_away.PRS.3PL   that_way   DEM.ABL.SG.M   wish.NOM.PL.F
      'dugdhā   _ápakrāmanti_
      un_milked.PPP.NOM.PL.F   walk_away.PRS.3PL
      'As the cows walk away un-milked, that way wishes walk away from him
      un-milked.'   [MS I 4,5(6)]

In Vedic, predicates of main and subordinate clauses as well as inside complex converb constructions are also frequently morphologically nominal. If a sentence lacks a finite verb, but possesses two nominal forms, it is easy to identify one as the subject and the other as the predicate as in the next example (with nominal matrix predicates typically in the clause-initial position).

(19) Vedic Sanskrit
      duranuvédo   _vā́_   _amútra_   **_yajñáḥ_**
      hard_to_find.NOM.SG.M   PRT   beyond.LOC.SG   sacrifice.NOM.SG.M
      'The **sacrifice** is **hard to find** in the beyond.'   [MS I 4,8(2)]

When dealing with more than two nominals, however, it is not always straight-forward how to group them, which can be seen when comparing (20) with (21). In (20), the two forms _svargó … lokáḥ_ together form the nominal predicate as a complex expression (bracketing the topic-introducing particle _vái_ in Wackernagel position). While the clause in (21) seems to be superficially struc-tured in a similar way, the initial element _vā́g_ alone acts as the nominal

---

18 That the two mentions of 'un-milked' present secondary predicates rather than attributes is underscored by the forms _yáthā_ 'as' and _evám_ 'so, that way', which target the state of the referents during the events expressed by the finite verb forms.

predicate, while *gāyatrī́ śatā́kṣarā* form a complex expression. This interpretation can only be gathered from world knowledge and context, for which see the passages in Amano (2009) (see also Amano 2009: 47–50 for a discussion of word order variation in this particular prose text).

(20) Vedic Sanskrit
    svargó   *vái*   **lokáḥ**   *pratnám*
    heaven.NOM.SG.M   PRT   world.NOM.SG.M   previous.NOM.SG.N
    '(the) previous (world) is the **heavenly world**'
                                              [MS I 5,5(2), cp. Amano 2009: 181]

(21) Vedic Sanskrit
    vā́g   *vái*   **gāyatrī́**   **śatā́kṣarā**
    voice.NOM.SG.F PRT   Gāyatri.NOM.SG.F   hundred_syllable.NOM.SG.F
    'the **hundred-syllabic Gayatri** [i.e. a type of metre] is the voice'
                                              [MS I 4,13(3), cp. Amano 2009: 172]

In complex converb constructions ('genitivus absolutus' or 'locativus absolutus'), illustrated in (22), a participle pairs up with another nominal, both elements being marked as either locative or genitive. The participle typically encodes a resultative or sometimes an ongoing state, and the other nominal an argument of it, with the event or state expressed by the participle having happened prior to or happening at the same time as the event or state encoded in the matrix clause. Inside the converb construction, the two morphologically nominal forms relate to each other as subject and predicate, and thus cannot together form a complex nominal expression given the discourse-based understanding adopted here.

(22) Vedic Sanskrit
    agnér   *vā́i*   **sṛṣṭásya**   *téjā*
    Agni.GEN.SG.M PRT release.PPP.GEN.SG.M radiance.NOM.SG.N
    údadīpyata
    shine_forth.IMPF.3SG.MID
    '**When Agni** [i.e. the fire god] **had been created**, (his) radiance shone forth'
                                                      [MS I 6,5(3)]

4.4 Loose appositions

While narrow appositions such as *President Obama* are normally considered integral complex expressions, loose appositions such as *Obama, the President*

are commonly considered two separate nominal expressions of equal syntactic rank. The differences are reflected in prosodic differences as well as structural ones – for instance regarding the possibility to elide a loose apposition without change in meaning of the referring expression. While the distinction seems a comparatively straightforward one in English, there are several challenges to identifying loose appositions in the languages at issue here. Firstly, the semantic types that are found in the realm of "appositions" seems a lot broader in at least certain Australian languages in comparison to European languages (Sadler and Nordlinger 2010) and it is not straightforward how these might match onto complex or separate nominal expressions. Possibly, advances in the study of prosodic correlates could inform more in-depth analyses (see Sadler and Nordlinger 2010: 422–423 on Australian languages). Formally, straightforward clues such as having to do with the use of function words – used only once with scope only over the second NP in the loose apposition *Obama, the President* – are lacking (see Section 2). As for word order, this topic still awaits detailed cross-linguistic investigation. However, a preliminary survey of Vedic Sanskrit prose suggests that there is a similar reversal of order between close and loose appositions as in English (cp. also Bauer 2017). Only a very small and seman-tically quite generic set of social role terms including the terms for 'king' and 'god' precede proper nouns (e.g. *rā́jā váruṇaḥ* 'King Varuna, e.g. RV 1.24.7), whereas the post-nominal slot is occupied by much more varied terms (e.g. *víṣṇur gopā́ḥ* 'Vishnu, the guardian, e.g. RV 1.22.18).

5 Conclusion

This paper discusses the question of how to identify multi-word nominal expres-sions in flexible word order languages, where neither inflectional nor syntactic distributional clues offer coercive clues as to whether two or more nominal words belong together as a unit or not. The data discussed here stems, on the one hand, from a handful of languages from Australia for which the question of how to identify complex nominal expressions has been raised, namely Warlpiri, Jaminjung, and Gooniyandi. In addition, data from Vedic Sanskrit is included, which shows the same lack of coercive morpho-syntactic clues for whether two or more nominals form a complex expression. Paired with an absent or weak noun-adjective distinction, the delimitation of "real" complex nominal expres-sions from "apparent" ones can be difficult if not impossible if relying purely on morpho-syntactic evidence. However, when prosodic, information-structural and contextual information is taken into account, besides word order defaults and

derivational morphology (where applicable), it is frequently possible to identify whether or not two or more nominals form a complex expression in the sense of sharing a discourse-functional role or not. This discourse-based understanding of what it means to form a complex nominal expression is what unites the languages discussed here with languages such as English where there is a stronger morphosyntactic differentiation between types of nominal expressions.

This paper surveys previous literature and expands the discussion with respect to languages considered and regarding the types of phenomena discussed. Particularly, the focus is not exclusively on discontinuity as in previous studies, but also on juxtaposition, as the question of how to identify complex nominal expressions arises in any case in the languages under consideration. Among the phenomena discussed are nominals placed in the right or left periphery, various types of nominal predicates, and loose appositions, all of which may be mistaken as forming a complex expression with a nominal in argument or adjunct role. To exclude such cases from consideration in analyzing complex nominal expression is an essential prerequisite for a detailed analysis of the grammatical system in languages of the type discussed in this paper.

Abbreviations

ABL	ablative
ACC	accusative
DAT	dative
DEM	demonstrative
EMPH	emphatic marker
ERG	ergative
EVID	evidentiality marker
F	feminine
GEN	genitive
IMPF	imperfect
INS	instrumental
LOC	locative

M	masculine
MID	middle
N	neuter
NOM	nominative
OBJ	object marker
OBL	oblique
PERF	perfective
PL	plural
PPP	past passive participle
PRS	present tense
PRT	discourse particle
PST	past tense
REL	relative pronoun
SEQ	sequential marker
SG	singular
SUBJCT	subjunctive

References

Amano, Kyoko. 2009. *Maitrāyaṇī-Saṁhitā I-II. Übersetzung der Prosapartien mit Kommentar zur Lexik und Syntax der älteren vedischen Prosa*. Bremen: Hempen Verlag.

Austin, Peter. 2001. Word order in a free word order language: The case of Jiwarli. In Jane Simpson, David Nash, Mary Laughren, Peter Austin & Barry Alpher (eds.), *Forty years on: Ken Hale and Australian languages*, 205–223. Canberra: Pacific Linguistics.

Austin, Peter & Joan Bresnan. 1996. Non-configurationality in Australian Aboriginal languages. *Natural Language & Linguistic Theory* 14(2). 215–268.

Baker, Mark C. 2001a. Configurationality and polysynthesis. In Martin Haspelmath & Ekkehard König (eds.), *Language typology and language universals*, 1433–1441. Berlin & New York: De Gruyter.

Baker, Mark C. 2001b. The natures of nonconfigurationality. In Mark Baltin & Chris Collins (eds.), *The handbook of syntactic theory*, 407–438. Malden, MA: Blackwell.

Bauer, Brigitte L. M. 2017. *Nominal apposition in Indo-European. Its forms and functions & its evolution in Latin-Romance*. Berlin & Boston: De Gruyter Mouton.

Bhat, Darbhe N. S. 1994. *The adjectival category. Criteria for differentiation and identification*. Amsterdam & Philadelphia: John Benjamins.

Bresnan, Joan. 2001. *Lexical-functional syntax*. Malden, MA: Blackwell.

Chafe, Wallace. 1994. *Discourse, consciousness, and time*. Chicago: Chicago University Press.

Delbrück, Bertold. 1878. *Die altindische Wortfolge aus dem Çatapathabrāhmaṇa*. Halle: Verlag der Buchhandlung des Waisenhauses.

Dixon, R. M. W. 1980. *The languages of Australia*. Cambridge: Cambridge University Press.

Fanselow, Gisbert & Caroline Féry. 2006. *Prosodic and morphosyntactic aspects of discontinuous noun phrases: A comparative perspective*. (unpublished manuscript) University of Potsdam. http://user.uni- frankfurt.de/~cfery/publications/Prosodic_and _morphosyntactic_aspects_of_disc ontinuous_NPs.pdf (last accessed 7/9/18)

Hale, Kenneth. 1983. Warlpiri and the grammar of non-configurational languages. *Natural Language & Linguistic Theory* 1(1). 5–47.

Hale, Mark. 1996. Deriving Wackernagel's Law: Prosodic and syntactic factors determining clitic placement in the language of the Rigveda. In Aaron L. Halpern & Arnold M. Zwicky (eds.), *Approaching second. Second position clitics and related phenomena*, 165–197. Stanford: CSLI.

Heath, Jeffrey. 1986. Syntactic and lexical aspects of nonconfigurationality in Nunggubuyu Australia. *Natural Language & Linguistic Theory* 4(3). 375–408.

Himmelmann, Nikolaus P. 1997. *Deiktikon, Artikel, Nominalphrase: Zur Emergenz syntaktischer Struktur*. Tübingen: Niemeyer.

Himmelmann, Nikolaus P. & Eva Schultze-Berndt (eds.). 2006. *Secondary predication and adverbial modification – the typology of depictives*. Oxford: Oxford University Press.

Hock, Hans Henrich. 1982. Clitic verbs in PIE or discourse-based verb fronting? Sanskrit sá hovāca gárgyaḥ and congeners in Avestan and Homeric Greek. *Studies in the Linguistic Sciences* 12(2). 1–38.

Hock, Hans Henrich. 1996. Who's on first? Toward a prosodic account of P2 clitics. In Aaron L. Halpern & Arnold M. Zwicky (eds.), *Approaching second. Second position clitics and related phenomena*, 199–270. Stanford: CSLI.

Jelinek, Eloise. 1984. Empty categories, case, and configurationality. *Natural Language & Linguistic Theory* 2(1). 39–76.

Kalbertodt, Janina, Beatrice Primus & Petra B. Schumacher. 2015. Punctuation, prosody & discourse: Afterthought vs. right dislocation. *Frontiers in Psychology* 6 (art. 1803). 1–12.

Klein, Jared S. 1992. *On verbal accentuation in the Rigveda*. New Haven: American Oriental Society.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Laughren, Mary. 1989. The configurationality parameter and Warlpiri. In Laszlo K. Maracz & Peter Muysken (eds.), *Configurationality: The typology of asymmetries*, 319–353. Dordrecht: Foris.

Louagie, Dana & Jean-Christophe Verstraete. 2016. Noun phrase constituency in Australian languages: A typological study. *Linguistic Typology* 20(1). 25–80.

McGregor, William B. 1990. *A functional grammar of Gooniyandi*. Amsterdam & Philadelphia: John Benjamins.

McGregor, William B. 1997. Functions of noun phrase discontinuity in Gooniyandi. *Functions of Language* 4. 83–114.

Nash, David. 1986 [1979]. *Topics in Warlpiri grammar*. New York & London: Garland.

Nolda, Andreas. 2008. Topic integration: Syntax and semantics of German 'split topicalization'. In Robin Sackmann (ed.), *Explorations in integrational linguistics: Four essays on German, French, and Guaraní*, 191–220. Amsterdam & Philadelphia: John Benjamins.

Nordlinger, Rachel. 1998. *Constructive case. Evidence from Australian languages*. Stanford: CSLI.

Nordlinger, Rachel. 2014. Constituency and grammatical relations in Australian languages. In Harold Koch & Rachel Nordlinger (eds.), *The languages and linguistics of Australia*, 215–262. Berlin & Boston: De Gruyter Mouton.

Reinöhl, Uta. 2016. *Grammaticalization and the rise of configurationality in Indo-Aryan*. Oxford: Oxford University Press.

Reinöhl, Uta. 2020. Continuous and discontinuous nominal expressions in flexible (or "free") word order languages – Patterns and correlates. *Linguistic Typology* 24(1).

Sadler, Louisa & Rachel Nordlinger. 2010. Nominal juxtaposition in Australian languages: An LFG analysis. *Journal of Linguistics* 46. 415–452.

Schäufele, Steven William. 1991. *Free word-order syntax: The challenge from Vedic Sanskrit to contemporary formal syntactic theory*. Urbana-Champaign: University of Illinois dissertation.

Schultze-Berndt, Eva & Candide Simard. 2012. Constraints on noun phrase discontinuity in an Australian language: The role of prosody and information structure. *Linguistics* 50(5). 1015–1058.

Siewierska, Anna. 1984. Phrasal discontinuity in Polish. *Australian Journal of Linguistics* 4. 57–71.

Simpson, Jane. 1991. *Warlpiri morpho-syntax: A lexicalist approach*. Kluwer: Dordrecht.

Simpson, Jane. 2006. Depictives in English and Warlpiri. In Nikolaus Himmelmann & Eva Schultze-Berndt (eds.), 2006. *Secondary predication and adverbial modification – the typology of depictives*. Oxford: Oxford University Press.