John Benjamins Publishing Company

Parts of speech, comparative concepts and Indo-European linguistics

Luca Alfieri

University of Studies Guglielmo Marconi

The paper adopts and further elaborates on the distinction between comparative concepts (CC) and descriptive categories (DC) by proposing a partly new definition of the parts of speech (PoS), and uses that definition to provide a new analysis of PoS in Latin and RV Sanskrit. More, specifically, the paper shows that in Latin three major classes of morphemes are found (nouns, adjectives and verbs), whereas in the RV only two major classes are found (verbal roots and nouns) and the typical "adjective" is a derived stem built on a verbal root meaning a quality (i.e., roughly a nominalization). The data described are then used to contribute to the CC debate in the field of PoS, by showing its relation with historical Indo-European linguistics, by critically analysing traditional labels such as noun, adjective, verb, root, stem and lexeme, and by questioning the alleged incommensurability between CCs and DCs.

1. Introduction

It is well known – at least for a large part of the scholars interested in typology – that comparative concepts (CC) and descriptive categories (DC) are two structurally different types of classes from a logical and methodological point of view. Still, in the most radical formulations, the logical difference between the two types of categories was understood as if they represented two totally unrelated and almost incommunicable sets of classes (see the Introduction to this volume).[1] Three consequences derive from such a radical dichotomy between DCs and CCs: i) there is no principled way to pass from CCs to DCs in single languages; ii) CCs may serve as a tool for comparing languages, but they cannot be used to describe single language grammar; iii) two sets of unrelated technical terms will be developed, one for either set of classes.

---

1. See Lazard (1992, 1997, 2001a: 365, 2001b: 141), Dryer (1997, 2016), Haspelmath (2007, 2010, 2012), Croft (2000, 2001, 2005, 2016), Croft & van Lier (2013).

Points i)-iii) have already been questioned in *Linguistic Typology* 2016, where many scholars pointed out that "logically independent" does not necessarily mean "unrelated" (see Introduction). However, none of the papers in LT addressed parts of speech (PoS) research specifically. The present work aims to fill the gap. It adopts and further elaborates on the distinction between CCs and DCs by proposing a partly new comparative definition of PoS, and employs that definition to provide a new analysis of the PoS in two ancient Indo-European (IE) languages, Latin and the Sanskrit language of the Rg-Veda Saṃhitā (RV) – often termed as "Early Vedic" (Dahl 2010: 2) – the oldest literary monument of the Indo-Aryan family.

The data described in the paper will be used to make a possibly novel contribution to both (a) the study of IE linguistics and (b) the CC debate. As for the first point, the paper will show that: (a.i) in Latin 3 major classes of lexical morphemes are found (nouns, adjectives and verbs), while in the RV only 2 major classes are found (verbal roots and nouns) and the typical "adjective" is a derived stem built on a verbal root meaning a quality (roughly a nominalization or a participial-like construction);[2] (a.ii) the difference between the PoS system in Latin and in RV Sanskrit is the result of a previously neglected (or not fully understood) typological change, the change from a PoS system with only two major classes of simple lexical units, verbal roots and nouns, to a PoS system with three major classes of simple lexical units, nouns, verbs and adjectives: schematically [N (AV)] → [N, A, V];[3] (a.iii) that the minimum verbal unit of the Sanskrit language, the root (Skt. $d^hātu$-), is a unit of language analysis that is radically different from the Latin simple verb stem from a functional and categorical point of view.[4] As for the

---

2.   Simple lexical units (or *items*, as American Structuralists called them) can also be termed as *lexical morph(eme)s, lexemes* or *lexical roots*. Each term has its own problems. See Mugdan (1986, 2015), Blevins (2016: 19ff.) and Haspelmath (forth.) on the term *morph(eme)*; Aronoff (1994: 16ff.), Touratier (2009) and Mugdan (2015: 253ff.) on the term *lexeme*; Alfieri (2016) and below fn. 4 and Section 7.2 on the term *root*. In the following, the term *root* is avoided, while *lexical morph(eme)* and *lexeme* are used as synonyms, even if *lexeme* means 'simple lexical unit' only in some authors (e.g., Martinet 1980: 16), while it refers to the dictionary word, which can be simple or derived, in others (e.g., Lyons 1977: 19).

3.   On the schemas [N, A, V] and [N (AV)], see Beck (2002: 6ff.) and below Sections 5–6.

4.   As for claims a.i)-a.iii), see also Alfieri (2009, 2013, 2016, 2018, 2020); as for claim a.i)-a.ii), see also Bozzone (2016), though she projected the lexical structure [N (AV)] to Pre-Proto-IE, rather than viewing it a "synchronic" stage of RV Sanskrit. Moreover, note that the term *root* has both a synchronic and a diachronic meaning. The antecedent of a set of genetically related forms is a diachronic root (i.e., a unit of the linguist): PIE *b^her-oh_2 'to bring' > Lat. *fer-o*, Gk. φέρ-ω, Skt. $b^hár-ā-mi$, but the simple lexical morpheme used as input for productive word-formation rules is a synchronic root (i.e., a unit of the speaker): Skt. *tap-* → *tápyate* 'he burns', *tápas* 'heat'. In the following, the term *root* is used only in its synchronic meaning unless

second point, the paper will show that none of the consequences i)-iii) implicit in the idea of a radical dichotomy between CCs and DCs is fully acceptable. Specifically, it will show that: (b.i) CCs are not only a tool for comparing languages, but also a tool for clarifying some aspects of the synchronic grammar of single languages (in this case: Latin and Sanskrit), as well as the diachrony of the IE family; (b.ii) the CCs of the PoS can be used to explain the traditional terms that refer to PoS, rather than to produce a new set of terms that have no relation with the canonical labels of *noun, verb, adjective* and *root*; (b.iii) CCs and DCs can be more profitably seen as the two poles of a continuum, rather than as two totally unrelated sets of classes.

2.   The comparative concept debate in the field of the PoS

As shown in the Introduction, the participants in the CC debate shared the idea that typology is the study of the general theory of LANGUAGE (that is, "general linguistics") and that the categories needed to describe the functioning of LANGUAGE – or to compare languages, which is the same – are structurally different from the categories needed to describe the individual languages. The debate, therefore, revolved around how the categories needed to describe the functioning of LANGUAGE should be designed, how they differ from the categories needed to describe single languages and what kind of inconsistencies arise when the two types of categories are confused and a descriptive-like category is used to solve a general-typological problem.

Apart from the idea that CCs are basically function-based concepts, while DCs are primarily form-based categories, all the participants in the debate agreed that CCs can be of different types and a full list of all types of CCs employable in cross-linguistic research is not yet available. However, many scholars agreed that different CCs point to different levels of generality in the continuum that goes from LANGUAGE to single languages, and CCs can be divided into at least three types in terms of their level of generality (Croft 2016), namely "pure" CCs and "hybrid" CCs, which are further divided into "constructions" and "strategies" (or "portable" CC, in Beck's terms, 2016).[5]

---

otherwise specified (clearly, in the sense of "old-time synchrony" defined by Janda & Joseph 2003: 21).

5.   The list above was developed by Croft (2016) with the aim of being a consensus view, but it does not cover exactly the type of CCs described by Dahl (2016) and Lander & Arkadiev (2016). On the idea of a continuum between LANGUAGE and single languages, see Gil (2016).

Pure CCs are the most general type of CCs. They do not represent linguistic categories in the strict sense; rather, they are the zones of universal-cognitive space arbitrarily designed by the researcher to describe some kind of cross-linguistic variation (thus, they are not "true" or "false", nor do they "exist" outside the research for which they are designed, see Haspelmath 2007, 2010). Pure CCs are typically (though not only) designed on conceptual maps – that is, maps of universal-cognitive space – which can be purely semantic or "hybrid", if they include a discourse-pragmatic (or, more generically, syntactic) component.[6] CCs are universal in the sense of Coseriu's "universals of the theory" (2001), since all languages can be classified on how they code the zones of universal cognitive space defined on a map. Pure CCs, in other words, are found in all languages with the same form, because strictly speaking they are formless – that is, they are purely functional concepts that exclude any reference to linguistic forms in their definition. Paramount pure CCs are the notions of BASIC WORD ORDER, ALIGNMENT or the definition of the ADJECTIVE that is common in word-order typology (that is, a property concept with no reference to any formal feature). These notions do not refer to the grammar of any individual language in their completeness, but can be applied to all languages: a language can be SVO or SOV, it can be ergative or accusative, it can code property concepts in one way or in another, but no single language grammar needs the super-ordinate notions of "alignment", "basic word order" or "property concept".

Hybrid CCs aim at a lower level of generality than pure CCs and embrace some reference to linguistic form in their definition. Constructions are form-meaning pairings: any form-meaning pair used to code a specific function or a specific zone of cognitive space defined on a conceptual map is a construction. Both pure CCs and constructions are universal, though in a different sense. Pure CCs are found in all languages with the same form, since they are not linguistic categories in the strict sense. Constructions are already linguistic categories, but they are underspecified categories: they are found in all languages, since each language

---

6. On semantic maps, see Haspelmath (2014) and *Linguistic Discovery* 8.1 (2010). Discourse-based pragmatic categories (or functions) can be seen as interpersonal or communicative functions (Rijkhoff 2016), or at least as syntactic categories, if syntax is considered a universal language-external layer, as in Croft (1991: 61). The term *hybrid* is common, though it is ambiguous with reference to conceptual maps: purely conceptual maps are based only on extra-linguistic factors, while hybrid maps already include a linguistic component. However, discourse-pragmatic notions are extra-linguistic just as semantic-conceptual notions, and the same holds true for syntactic notions, if syntax is meant roughly as "discourse". The map, therefore, is really hybrid only if it has some reference to linguistic forms – that is, if it is semantic-syntactic and syntax is meant as grammatical (thus also formal) processing.

codes the space on conceptual maps in one way or another, but the specific way whereby each language codes the zone of cognitive space arbitrarily identified in the research is language-particular, either for markers that define the construction, for the overall space on the conceptual map occupied by it (that is, its function), or for the level of language structure at which the construction is fixed (e.g., the simple stem, the derived stem, the phrase, the clause, etc.). Paramount constructions are the definition of the ADJECTIVE as a quality modifier proposed by Alfieri (2014a), or the QUALITY PREDICATE studied by Wetzer (1996) and Stassen (1997). While pure CCs are found in all languages with the same form, constructions are found in all languages but have a different form in each language: all languages show a quality modifier construction or an adjectival predicate, but the most typical quality modifier is a simple stem in Latin, while it is a relative clause in Garo, and the most typical quality predicate is noun-like in Latin, but verb-like in Lao.

Finally, strategies (or portable CCs) are the most specific type of CCs. They are usually induced via generalization (that is, abstracted) from a specific category of a single language or group of languages. They embrace some reference to a specific linguistic form in their definition; thus they cannot be applied to all languages but are confined to the languages that display some formal and functional similarity to the language from which the portable CC in question was first abstracted. Paramount portable CCs are CASE or MIDDLE, that is to say categories that can be used to compare the languages in which something similar to the Latin case or the Greek middle is found, but cannot be used to compare languages in which nothing similar to such categories exists (e.g., not all languages display case inflection, perfect tense or middle voice), and also in the languages in which such categories are found, they are not defined through exactly the same function (e.g., the Greek genitive also codes the agent if joined with *hypó*, while the same function is coded by the ablative with *a* or *ab* in Latin, etc.).[7]

Each type of CC is legitimate and can lead to interesting generalizations, if used consistently. Still, not all research questions can be addressed with each type of CC with equal validity; rather some very general typological problems seem to be structurally insoluble if portable CCs or, more generally, descriptive-like CCs are employed. The PoS may easily represent a case of this type, since all languages divide the minimum lexical units into classes, although the classes defined differ,

---

7.   Clearly, the notions of MIDDLE and CASE can be defined in purely functional terms, transforming a portable CC into a construction or a pure CC: if we say that in English the ABLATIVE is coded through the preposition *from* (*from the house*), as many Generativists do, we are implicitly defining CASE as a purely functional CC. However, this definition can easily be misleading: if it is accepted, several CASE constructions can be found in Greek, some of which are coded by case endings (e.g., the GENITIVE), others by prepositional phrases (e.g., the ABLATIVE).

sometimes substantially, from one language to another.[8] In fact, if a DC or a porta-ble CC is employed to compare the PoS in two languages – no matter how similar or genetically related – the well-known paradox of the so-called *controversial cat-egory assignment* cannot be avoided (that is, the case in which on the basis of the same empirical data, the same language is claimed to show or to lack a given PoS by different researchers).[9] In the following pages this problem will be discussed, reviewing the former analysis of the PoS in Latin and Sanskrit.

3.   The PoS in Latin and in Sanskrit: State of the art

The PoS system in Latin is almost uncontroversial, at least in its most superfi-cial features. In Latin three major word classes are found: nouns, adjectives and verbs.[10] These classes are defined through well-known inflectional features: case, number and gender for the noun (*virtus*, Example 1); person, tense, mood and diathesis for the verb (*habetur*, Example 1); agreement and comparison for the adjective (*clara* and *aeterna*, Example 1):

(1)   *virtus*   *clara*
      virtue(F).NOM.SG splendid.F.NOM.SG

      aeterna=que   *habe-tur*
      lasting.F.NOM.SG=and have-PRS.3SG.PSS

      'mental excellence is a splendid and lasting possession'   (Sall., *Cat.* 1.4)

---

8.   This is, more or less, the position claimed by Dryer (1997: 116ff.), Croft (2000, 2001: 29ff., 63ff., 2005), Croft & van Lier (2012), Haspelmath (2012) and Cristofaro (2009).

9.   This problem has a long history in PoS research. See the debate on the PoS in Salish (Demirdache & Matthewson 1995; Croft 2001: 76ff.; Beck 2002, 2013), Mundari (*Linguistic Typology* 9.3, 2005), Iroquoian (Sasse 1993a vs. Mithun 2000), Tagalog (*Theoretical Linguistics* 35(1), 2009), Chamorro (*Theoretical Linguistics* 38(1–2), 2012), only to quote the most famous cases (see Stassen 1997: 31ff., Haspelmath 2007, Shachter & Shopen 2007: 18ff. and Cristofaro 2009 for further examples). A similar problem affects Dixon's adjectival theory as a whole: on the basis of similar evidence, Dixon (1982 [1977[1]]) claimed that most languages lack adjec-tives and merge quality concepts with nouns or verbs, but Dixon (2004) also said that the adjective is a universal class that is always distinguished from nouns and verbs.

10.   The categorization of stems is more specific than the categorization of the word-forms built on those stems, but it is not structurally different: the word-form *amā-mus* 'we love' and the stem *amā-* 'to love' are defined through the same features, though they show different values in each feature: no specified value for *amā-* vs. 1pl.pres.act. for *amāmus* (Ramat 1999, 2014). Therefore, in the following the labels "stem classes" and "word classes" are used as syn-onyms.

The adjective might seem to be a sub-class of the noun rather than an independent PoS, and in fact it was conceived as such up to the Speculative Grammar in the Middle Ages.[11] However, this view is misleading. In Latin adjectives and nouns show almost the same endings from a formal point of view (although adjectival stems in -*u*- and diphthong are excluded), they do not have a rigidly fixed word order in the phrase, and adjectives can be used as nouns without any overt marking, only by being settled in the appropriate syntactic slot: e.g., *antiqui dixerunt* 'the elders said'.[12] However, comparison and agreement clearly divide nouns and adjectives in Latin. Bar some exceptions (see below), comparison is excluded with nouns and, if used as modifiers or as predicates, adjectives agree, but nouns cannot: Lat. *malae puellae veniunt* 'evil(f.pl.) girls(f.pl.) come' vs. *puellae, donum diaboli, veniunt* 'girls (f.pl.), devil's gift (nt.sg.), come', but not *puellae **donae diaboli*. Therefore, adjectives are noun-like in Latin, but they differ from nouns, since only nouns are already marked for gender in the lexicon (Lat. *angui-* 'snake(ᴍ)', *virtus* 'virtue(ꜰ)'), and so cannot be used as modifiers if not overtly trans-categorized, while adjectives are gender-neutral in the lexicon (*clara-* 'splendid', *nigru-* 'black'), but have to agree, so can be used both as referents and as modifiers without being overtly trans-categorized.[13]

However, quite a few problems arise if the traditional notion of PoS is looked at more closely.[14] From a technical point of view, a PoS is an unweighted bundle of features, that is to say a bundle composed of a different number of features in which the relative value (the weight) of each feature is not fixed. Therefore, even if semantics is left aside, each PoS is defined by different features: in Latin, the adjective is defined by agreement and comparison, but also by the presence of quite a few adjectival-forming affixes (e.g., -*oso*-, which converts nouns into adjectives: *negotium* 'business' → *negotiosus* 'active') and adjectival-selecting affixes (e.g., -*tat*-,

---

11.   The idea that Latin has three PoS can be traced to the Middle Ages (rather than to the 18th century, as said by Rijkhoff & van Lier 2013: 1 fn. 1), but is not shared by Latin grammarians (see Alfieri 2014b).

12.   This construction is often defined 'substantivized adjective' in classical grammars and 'syntactic conversion' or 'zero marked trans-categorization' in typological works (Rijkhoff & van Lier 2013: 21ff.).

13.   In Aronoff's terms, the endings of the noun are *morphomic* (1994: 22, 45), since they depend on a formal feature that is independent of both syntax and semantics, while the endings of the adjective are not *morphomic* at all, since they depend on a semantic-syntactic feature (i.e., agreement). See also Corbett (2006: 126ff.) on the topic.

14.   See Robins (1964: 225ff.), Crystal (1967: 24), Matthews (1967: 155), Lyons (1979: 49ff.) and Gross (1979).

which converts adjectives into nouns: *gravis* 'heavy' → *gravitas* 'seriousness').[15] However, the various features, or more precisely, the various construction that define the adjective do not describe exactly the same group of items: agreement defines adjectives, but not-agreeing adjectives (that is, adjectives that do not mark agreement overtly) are found (*sapiens* 'wise', *audax* 'bold'), and also agreeing nouns, though rare, are possible (*lupus* 'wolf' → *lupa* 'she wolf' and *magister* 'teacher' → *magistra* in *historia […] magistra vitae* 'history is life's teacher', Cic., *De or.* II.36). In the same vein, comparison defines adjectives, but not-comparable adjectives are found (*sinister* 'left', *dexter* 'right'), just as nouns or pronouns that in specific circumstances can be compared (*amicissimus* 'very friend' in Plaut. *Most.* 340, or *ipsissimus* 'he in very person' from the pronoun *ipse* 'he in person' in Plaut. *Trin.* 988). In sum, the group of items defined by agreement do not coincide with the group of items defined by comparison and neither group coincides with the group of items defined by adjectival-selecting or adjectival-producing affixes. Nonetheless, if the various features that usually define the adjective are gathered together to form a bundle, the bundle can be used to delimit a class of items that is descriptively useful for illustrating the functioning of Latin grammar (and especially inflection), although the margins of the class are fuzzy on cases and one does not exactly know what to do when the various features that compose the bundle are in contrast.

3.1   The former (Western) classifications of Sanskrit

When Sanskrit entered the horizon of western scholars in the beginning of the 19th century, by and large it was described using the traditional DCs elaborated by Latin grammarians, partly because the similarity between the two languages seemed to allow such an operation with relatively minor problems.[16]

At present, it is broadly agreed that, bar a few marginal facts, the PoS systems in Latin and Sanskrit are almost identical, the two languages being related genetically. In both cases, three major word classes are found and these classes are defined through grossly the same inflectional features: case, number and gender

---

15.   The constraints that define each suffix can change diachronically: *-no-* attaches to verbal roots in PIE (Lat. *plē-nus* 'full' from PIE *$pelh_1$- 'to fill'), but it attaches only to nouns in Latin (*pater-nus* 'paternal' from *pater* 'father').

16.   More precisely, the native tradition of Indian grammar influenced the earliest Western grammars of Sanskrit, especially those published between Roth (1660–68) and Bopp (1827). However, during the 19th century, in German universities a Western standard of Sanskrit grammar was developed using Latin DCs and this standard is at the basis of any modern grammar of Sanskrit (Law 1993). Clearly, influxes of Sanskrit indigenous grammar are not excluded in the Western standard (especially regarding the notions of root and word-formation, see Alfieri 2013b, 2014c), but the bulk of the standard is clearly rooted in the Latin-based tradition.

for the noun; person, tense, aspect and diathesis for the verb; agreement and comparison for the adjective. Also in this case, adjectives may resemble a sub-class of nouns, since adjectives show almost the same endings as nouns (if minor differences are excluded),[17] have a free position in the phrase and can be used as nouns only by being settled in the appropriate syntactic slot (Example 2):

(2) **āmā́su**   *cid*   *da-dʰi-ṣe*   **pak-vá-m**   *antáḥ*
    raw.F.LOC.PL   indeed   RED-put-PF.2SG   cook-NM-NT.ACC.SG   inside
    '[Indra] you placed the cooked (milk) inside the raw (cows)'   (1.62.9ᶜ)

However, in this case too agreement clearly singles out nouns and adjectives, since adjectives agree, while nouns cannot, since they are gender-marked in the lexicon. Agreement mismatches are thus possible with nouns used as modifiers or as predicates, though not with adjectives (Example 3):[18]

(3) **strī́**   *hí*   **brahmā**   *ba~bʰūvi-tʰa*
    woman(F).NOM.SG   then   brahman(M).NOM.SG   PF~be-PF.2SG
    "for you, brahmin(NOM.M.SG), have turned into a woman(NOM.F.SG)!"
    (8.33.19ᵈ)

Beneath the surface, however, things get complicated. While the definition of the noun and the verb are (or at least seem to be) similar (see Section 6), the definition of the adjective differs in the two languages. Agreement divides adjectives and nouns also in Sanskrit, but most of the items subject to agreement in Sanskrit are derived adjectives built from verbal roots, rather than simple adjectives, as in Latin (see Alfieri 2016, 2018 and Example 4):

(4) *agní-s*   **tig-ména**   *śoc-í-ā*
    Agni(M)-NOM.SG   be_sharp-NM.INS.NT.SG   burn-NM.NT-INS.SG
    yā́s-a-d   *víśva-m*   *ní*   *atrí-am*
    drive-SBJ.3SG   every-M.ACC.SG   into   Atri(M)-ACC.SG
    "with a sharp flame, Agni will attack every Atri [a daemon]"   (6.16.28ᵃᵇ)

---

17.   In Sanskrit, adjectival stems in diphthong are not found, while adjectives in *-u-* are found (*tapú-* 'hot'). Moreover, in RV Sanskrit (not in the classical language) adjectives tend to show the instr.pl. ending *-ebʰis* (from the ending of the PIE consonant stems *-bʰi*), while nouns more readily show *-āis* (from the ending of the PIE pronouns *-ois*), see Lazzeroni (2008).

18.   See also *pāpāḥ kaníyā gaccʰanti* 'evil(f.pl.) girls(f.pl.) come' vs. *kaníyās, dānám ásurasya, gacchanti* 'girl(f.pl.), daemon's gift(nt.sg.), come', but not ***kaníyās, dānā* (f.pl.) *ásurasya, gaccʰanti*. Note that in Sanskrit the final consonant of each word is always subject to sandhi: that is the nom.pl. ending is *-ās* both in *pāpāḥ* and *kaníyā*, but *-ās* before a voiceless stop becomes *-āḥ*, while it becomes *-ā* before a voiced stop or a vowel.

Moreover, the group of agreeing nouns (that is, the nouns that regularly join to the feminine suffix *-ī-* and agree) is far larger in Sanskrit, including not only specific names of deities and animals as in Latin (*devá-* 'God' → *devī́-* 'Goddess', *vŕ̥ka-* 'wolf' → *vr̥kī́-* 'she-wolf'), but also a large number – probably the majority – of derived nouns, such as the agent nouns in *-tar-* → *-trī́-*. See Example 5:

(5)   *úd   u   stómāso   aśvín-or   a-bud^h-rañ*
      up   now hymn(M).NOM.PL aśvín(M)-GEN.DU PST-wake_up-AOR.3PL

      jāmí   bráhmāi   uás-aś
      related.NT.NOM.PL prayer(NT)-NOM.PL dawn(F)-NOM.PL

      ca   dev-ī́-ḥ
      and   God-F-NOM.PL

      'The praise songs of the Aśvins have awakened, also our family
      formulations, and the Dawns, the Goddesses"   (7.72.3^{ab})

In the same vein, adjectival-forming or adjectival-selecting affixes are not a diagnostic criterion for distinguishing nouns and adjectives in Sanskrit. Indian native grammarians divide Sanskrit suffixes into two groups. Primary suffixes (usually called *kr̥t*) attach only to roots and build derived nominal stems, while secondary suffixes (usually called *tadd^{h}ita*) attach to simple or derived nominal stems and build secondarily derived nominal stems. Some affixes are more likely to build nouns (especially with root accent), but others more readily build adjectives (especially with suffix accent), but there is no clear-cut distinction between the two cases. E.g., the *kr̥t* suffix *-as-* often builds nouns with root accent, but it can also build adjectives with suffix accent: *máh-as-* 'greatness' and *mah-ás-* 'big' from *mah-* 'be big'. The *tadd^{h}ita* suffix *-ya-* usually builds adjectives of relation from nouns, but it also builds nouns from other nouns: *pítr-ya-* 'paternal' from *pitár-* 'father' and *vīr-yá-* 'manliness, strength' from *vīrá-* 'hero'.[19]

Comparative and superlative affixes show the same distribution as all other suffixes in Sanskrit, so they cannot be used to define adjectives. The primary comparative-superlative suffixes *-īyāṃs-* and *-iṣṭha-* attach only to verbal roots: see *náy-iṣṭ^{h}a-* 'who conducts in the best way' from *nī-* 'to conduct'; *jáv-īyāṃs-* 'faster' from *jū-* 'move fast'. See Example 6:

(6)   *yó   véd-iṣṭ^{h}o   a-vyat^h-íṣu*
      who.M.NOM.SG know-SUP.M.NOM.SG not-waver-M.LOC.PL

---

19.   This interpretation of the *kr̥t* and *tadd^{h}ita* suffixes is quite traditional in Sanskrit grammars; for discussion see Alfieri (2009: 34 fn. 62). Also in this case, diachronic changes may determine the passage of a suffix from one class to the other (see Burrow 1955: 119ff.).

áśvā-vant-aṃ   *jari-tṛ́-bʰyaḥ*   *vájaṃ*
horse-ADJ-M.NOM.PL   sing-NM-M.DAT.PL   prize(M)-ACC.SG

'[Indra] who among the unwavering is the best in finding the prize that
brings horses for the singers'   (8.2.24[ab])

The secondary comparative-superlative suffixes *-tara-* and *-tama-* attach to adjectives, to nouns, pronouns, prepositions and numerals without distinction: *tavás-tara-* 'stronger' from *tavás-* 'strong' (a secondary adjective derived from the verbal root *tū-* 'be strong'); *vīrá-tara-* 'more man', *vīrá-tama-* 'the most man' from *vīrá-* 'man'; *ka-tará-* 'which among two', *ka-tamá-* 'which among many' from *ká-* 'which'; *ut-tara-* 'higher', *ut-tama-* 'highest' from *úd-* 'up' and *śata-tama-* 'hundredth' from *śatam* '100'.[20] See Example 7:

(7)   **ámbi-tam-e**   **nádī-tam-e**
     mother(F)-SUP-VOC.SG   river(F)-SUP-VOC.SG

     dév-i-tam-e   *sárasvati*
     God-F-SUP-VOC.SG   Sarasvati(F).VOC.SG

     'Best mother, best river, best Goddess, Sarasvati'   (2.41.6[a])

Looking at these data, Sanskrit philologists usually conclude that the PoS system in Latin and in Sanskrit are *almost* identical, though in Sanskrit the adjective is not as sharply distinguished from the noun as in Latin, since only some of the criteria that are used to define the adjective in Latin give a positive answer in Sanskrit.[21] But this conclusion, possible as it may be in practice, is highly problematic from a theoretical point of view. To the extent that there is no objective criterion to weight the various features that define the "adjective" in Latin, it is logically impossible to establish what to do when two or more features are in contrast or when one of these features gives only a partly positive answer. In other words, Sanskrit philologists are perfectly able to describe Sanskrit data and their differences with respect to Latin, but they have no objective criterion to decide whether these differences are more adequate to claim that both languages show the "same" adjectival category, despite minor differences, or that Sanskrit has no adjective class at all, despite

---

20. This distribution of comparative-superlative suffixes is a conservative feature of Sanskrit: Av. *gaotəma-* 'a proper name', lit. 'big cow', Gk. βασιλεύτερος 'big king'; Av. *uštama-*, Gk. ὕστερος parallel to Skt. *uttama-*, etc. For discussion, see Lazzeroni (2005, 2013) and Alfieri (2009: 14).

21. See, e.g., Whitney (2000 [1879[1]]: 111), Delbrück (1888: 188f.), Wakernagel (1905: 1), MacDonell (1975 [1910[1]]: 178), Renou (1952: 338, 1965: 231) and Morgenroth (1977: 65).

some similarities with Latin.[22] Again, this is not a serious problem for Sanskrit philologists, who usually also know Latin, so they can use the Latin DC "adjective" as a reference and detail the differences in practice. Nonetheless, a similar strategy paves the way to misunderstandings and ambiguities when it is transferred from language-individual grammar to cross-linguistic research.

3.1.1   *Joshi (1967) and Bhat (1994, 2000)*

Sanskrit philologists often claim that the PoS systems in Latin and Sanskrit are *almost* identical, but more radical formulations can be found. As Speijer puts it (1974 [1896[1]]: 2): "Während Nomen und Pronomen im Indischen einen merkbaren Unterschied der Flexion aufweisen, werden die beiden Kategorien des Nomens, Substantiv und Adjectiv, im Flexion, Composition, Derivation fast unterschiedlos behandelt". Speijer's view is based on the same data discussed above and reported in any Sanskrit manual, but his conclusions are more radical in comparison to the common view (see fn. 21) and also compared to what he claimed elsewhere (e.g., Speijer 1998 [1886[1]]: 179), where the noun and the adjective are said to be *almost* identical, rather than *almost completely* identical. The reasons underlying Speijer's radical formulation are hard to fathom: his claim may be a shortcut used for the sake of brevity in a (relatively short) manual; it may be a tribute to Indian native grammar, which does not recognize the adjective as a relevant category;[23] or it may be the result of a complex reasoning in which Speijer evaluated the empirical data discussed so far and concluded that, despite the vulgate, Sanskrit has no adjectives, but quality nouns. Be that as it may, an objective criterion to establish whether the Sanskrit data support or discourage the employment of the category "adjective"

---

22.   One of the main inconsistencies of contemporary PoS theory is the impossibility of distinguishing major and minor classes, that is to say of establishing when a given group of items represent an independent PoS or a sub-class of a different PoS. On this point, see Sasse (1993b), Stassen (1997: 31ff.), Croft (2000, 2001: 78ff., 84, 2005), Shachter & Shopen (2007: 4) and Haspelmath (2012: 111ff.).

23.   Sanskrit grammarians recognize the existence of quality nouns (*gaṇavacana* 'indicator [*vacana*] of a quality [*gaṇa*]'), of gender-marked nouns (*liṅgavacana* 'indicator of gender [*liṅga*]'), of nouns commonly employed as appositions (*sāmāyavacana* 'indicator of a general class [*sāmāya*]') and of a class of exocentric words which are usually compounds (*anyapadārtʰa* 'exocentric, which takes its meaning [*artʰa*] from a different [*anya*] word [*pada*]'). Moreover, in the analysis of compounds and phrases, they acknowledge that two nouns can be co-referential (*samānādʰikaraṇas* 'referred to the same substrate') and either noun is the head of the substrate (*viśeṣya* 'qualified') and either is the modifier (*viśeṣyaṇa* 'qualifier'). However, they do not accept a true "adjective" class comparable to that defined by grammarians. See Cardona (1997a, 1997b) and Radicchi (1973–74) for Indian PoS theory, and Pontillo & Candotti (2011) for the terms above.

cannot be found if the Sanskrit data are described simply by transferring the Latin-based DC "adjective" to the cross-linguistic level.

As confirmation, Joshi (1967), followed by Bhat (1994, 2000), took up a part of the evidence discussed so far (namely, the ability of adjectives to head a NP and the lack of a fixed position in the NP), radicalized Speijer's view and claimed Sanskrit as a typical language "without" adjectives or with "true" noun-like adjectives (that is, adjectives that are indistinguishable from nouns). Obviously, this view can be criticized from a "splitter" point of view (or from the point of view of single language description, which is almost the same), saying that gender agreement is solidly established in Sanskrit (see Example 3) and it clearly singles out nouns and adjectives, although for the most part gender-agreeing stems are derivatives from verbal roots (see below).[24] But even this position is weak from a theoretical point of view. The point is that there is no principled way to establish whether the existence of an agreement system mainly limited to derived stems is or is not a sufficient criterion to say that Sanskrit has "adjectives", inasmuch as the category "adjective" is defined as an unweighted bundle of features, since the weight of the single features that compose the bundle cannot be determined objectively.

3.1.2   *Works following Dixon's approach (2004)*

As a confirmation, a few years after Bhat's publication, Dixon (2004) took up his previous work on adjectival typology (Dixon 1982 [1979[1]]) and proposed a different definition of the "adjective". He accepted the idea that PoS are simple lexemes and that simple lexeme classes can be defined also as unweighted bundle of features, but enlarged the bundle so as to include any possible difference between simple lexical units meaning qualities, objects and actions in any language. In other words, Dixon tried to pass from the mere transfer of a Latin-based DC on the cross-linguistic level, as the Sanskrit philologists usually did, to the building of a portable CC based on an open-ended list of possible diagnostic features (e.g., occurrence as verbal or non-verbal predicates, occurrence of head or modifier of a NP, different possibility in the predicate slot, different possibility in transitivity, comparative constructions or in forming adverbs, see Dixon 2004: 14ff.). However, following a partly similar claim in Stassen (1997: 359ff.), Alfieri (2009)

---

24. The terms "splitters" and "lumpers" have been current in PoS research since Croft (2001: 63ff.). Splitters are the scholars who tend to use all the available criteria in all languages to define the PoS, with the results of having a different group of items for each criterion (that is, for each construction), splitting each PoS into a myriad of sub-classes. Lumpers are those who tend to ignore arbitrarily (i.e., without an explicit and cross-linguistically constant criterion) some of the criteria that can be used to distinguish two or more classes of lexemes, and so lump these classes into a single PoS.

applied Dixon's criteria to RV Sanskrit and claimed that Sanskrit displays a small class of noun-like adjectives (that is, of primary lexical units meaning qualities divided from primary lexical units meaning objects by means of agreement), but it also shows a large number of verb-like adjectives, that is to say lexical units meaning roughly a quality that can be predicated verbally, as in Example 8–11:[25]

(8) *yā-bʰiḥ*   *sómo*   ***mód-a-te***
     who(F).INS.PL soma(M).NOM.SG enjoy-PRS1-3SG.MD

     harṣ-a-te   *ca*
     be_excited-PRS1-3SG.MD and

     'thanks to which [sc. the Waters] Soma is delighted and becomes excited'
                                                                    (10.30.5ᵃ)

(9) *ayáṃ*   *ha*   *túbʰyaṃ*
     3SG.M.NOM indeed you.DAT.SG

     váruṇo   ***hṛ-ṇî-te***
     Varuṇa(M).NOM.SG be_angry-PRS9-3SG.MD

     'Varuṇa now is angry with you'   (7.86.3ᵈ)

(10) *jāyā*   ***tap-ya-te***   *kitavá-sya*
     wife(F).NOM.SG be/make_hot-PRS4-3SG.MD gambler(M)-GEN.SG
     'the wife of the gambler is grieved [lit. 'is hot, burns (with pain)']'
                                                                    (10.34.10ᵃ)

(11) *ná* *sváp-nâya*   ***spṛh-aya-nti***
     not sleep-NM.M.DAT.SG be_eager-PRS10-3PL
     'They [sc. the Gods] are not eager for sleep'   (8.2.18ᵇ)

Also in this case, Alfieri's conclusion is neither true nor false, it simply is the product of a specific definition of the "adjective". However, this definition – like that employed by Bhat – is based on an unweighted bundle of features and the

---

25.   Sanskrit roots range from stative-unaccusative meanings (*śubʰ-* 'be beautiful', *tṛṣ-* 'be thirsty', *śī-* 'to lie'), to unergative meanings (*i-* 'to go', *bʰā-* 'to shine'), to true transitive-causative meanings (*han-* 'kill', *bʰid-* 'split'), not to mention roots that are compatible with a wide array of meanings that go from the stative to the transitive (*svād-* 'be, become, or make tasty', *bṛh-* 'be, become or make big, thick or strong', *tap-* 'become hot, heat'). Therefore, only a limited number of roots are listed in dictionaries with an exclusive quality meaning, but the absence of a quality meaning among the translational equivalents of a Sanskrit root in a dictionary does not preclude such a root being able to code a quality if inflected in the appropriate form. Sanskrit philologists acknowledge the high frequency of verb-like adjectival predicates indirectly. Gren-Eklund showed that nominal sentences with adjectival meanings are rarer in Sanskrit than they are in the modern European languages (1978: 34): but noun-like adjectival predicates are rare precisely because adjectival predicates are often coded verbally.

classifications based on unweighted bundles of features arrive at family resemblances, but forbid any consistent typology. In other words, if the "adjective" is defined through an open-ended list of features whose relative weight cannot be determined, one can identify one "adjectival" category for each of the (possibly infinite) features included in the list, but there is no principled way to establish which of the classes so identified is a major PoS, which is a minor class, how many adjective classes can be accepted in a language, and so on.

In sum, the debate on the Sanskrit adjective class is a paramount case of inconsistent category assignment. On the basis of the same empirical data, Sanskrit has been classified as having an adjectival category almost identical to that in Latin, as a language with quality nouns but no adjectives (or with noun-like adjectives) and as a language with verb-like adjectives. The radical difference between the results and the absence of any objective criterion to prefer one result over the others confirm that the method whereby these results are obtained is simply inconsistent. Both DCs and portable CCs are bundles of unweighted features and when the base of comparison – the *tertium comparationis* – is itself a bundle of unweighted features there is no logical method to establish what to do when the various features that compose the bundle are in contrast. In other words, if the Latin-based DC "adjective" is employed to classify Sanskrit adjectival typology, or if a portable CC abstracted from the Latin-based category "adjective" is employed for the same task, the only basis of comparison is the alleged existence of the Latin-based category "adjective", although the existence of this category has never been demonstrated but has rather been presupposed on the basis of misleading Eurocentric questions such as: "what are the adjectives of language X?" or "does language X have adjectives?". It is as if, in a juridical study, instead of asking "how does the Italian constitution work?" or "what are the most salient roles in the Italian constitution?" we asked "who is the king of Italy?" and went on to discuss whether the Prime Minister or the President (who have distinct roles in Italy) is the best candidate for being the king of Italy, ignoring the possibility that Italy might not have a king and that our typology of social roles should not include the role "king" as a primitive.[26] Therefore, the only possible way to overcome the paradox of inconsistent category assignment is to develop a cross-linguistically consistent method to descend from the study of LANGUAGE to the functioning of single languages in the field of PoS. This is what we will try to do in the next section.

---

26. See Dryer (1997: 116–119), Croft (2000, 2001: 29ff., 63ff., 2005), Croft & van Lier (2012), Haspelmath (2007, 2010, 2012) and Cristofaro (2009) for a similar conclusion.

4. A relatively new PoS theory

The PoS theory below is the result of a partly original elaboration based mainly on two previous sources, namely Hengeveld's PoS theory (1992: 47ff.) and Croft's definition of PoS (2001: 29ff., 63ff.). More precisely, it is a re-definition of the "Amsterdam Typology", and specifically of the lexical inventories [N, A, V] and [N (AV)], in the light of the criticism raised by the advocates of a radical distinction between CCs and DCs (see fn. 26).[27] Its goal is to show that, in contrast to the claims of the supporters of a radical division between DCs and CCs, a principled way to pass from CCs to DCs in the field of PoS can be obtained, at least as for IE languages.

The idea on which the theory is based is simple. Lexemes cannot be defined, if not starting from the constructions that define them, and constructions cannot be compared, if not on the basis of their function defined on a conceptual map. Therefore, if we develop a conceptual map on which the pure CCs of PoS are defined; a cross-linguistically valid method to classify the single language constructions on a conceptual map; and a cross-linguistically valid method to extract the lexemes from the constructions on the map, then a cross-linguistically consistent method to descend from the study of LANGUAGE to the description of the PoS in single languages is obtained. In other words, we are proposing to divide the continuum that goes from LANGUAGE to individual languages into three layers, which roughly correspond to the three types of CCs defined in Section 2 and to the three meanings in which the traditional notion of PoS can be understood.

4.1 PoS-concepts

The first and highest layer of analysis in the continuum is that of pure CCs. At this level, PoS are not single language DCs, but language-external universal concepts

---

27. For an overview on contemporary PoS theory, see Plank (1997), Anward, Moravcsik & Stassen (1997), Evans (2000), Comrie & Vogel (2000), Sasse (2001), Rijkhoff (2007) and Bisang (2013). On the Amsterdam Typology, see Anward (2000), van Lier (2009), Ansaldo et al. (2010), Hengeveld (2013), Alfieri (2013a). For a criticism, see Croft (2001: 65ff.) and Cristofaro (2009: 453ff.). The criticisms are serious, but they did not prevent the achievement of important results (Bisang 2013: 291ff.), such as the definition of different lexical inventories across languages (Hengeveld 1992: 69ff.; Beck 2002), the implicational hierarchy of PoS: verbs > nouns > adjectives (Hengeveld 1992: 68ff.), the discovery of correlations between PoS systems and word order (Hengeveld, Rijkhoff & Siewierska 2004), and the notion of flexibility (Rijkhoff & van Lier 2013, van Lier 2009, 2017).

defined on a conceptual map. The map can be purely semantic or "hybrid".[28] In this case, as in Croft (2001), the map is "hybrid", since it combines two proto-typically correlated universal parameters, a semantic concept (Object, Quality and Action) and a discourse-pragmatic function (Referent, Modifier and Predicate). See Table 1 (Croft 2001: 92):

Table 1. PoS as pure comparative concepts

	Referent	**Modifier**	**Predicate**
Object	**Object Referent**	Object Modifier	Object Predicate
Quality	Quality Referent	**Quality Modifier**	Quality Predicate
Action	Action Referent	Action Modifier	**Action Predicate**

Each slot in the table represents a zone of conceptual space defined in terms of semantic notions and discourse functions. All the slots are relevant for establishing the PoS system of a language, but three slots have a special status. The Object Referent is the NOUN, the Action Predicate is the VERB and the Quality Modifier is the ADJECTIVE: these categories are not the formal categories of any language, they are the zones of conceptual-functional space that are the most typical intersection between a discourse function and a semantic notion (thus, they are termed as "unmarked correlations" by Croft 2001: 89). More specifically, they are the zones of conceptual-universal space arbitrarily identified by the researcher whose coding in single language is the subject of the typology.

4.2 PoS-constructions

The second and more particular layer in the continuum is that of "hybrid" CCs or constructions in Croft's terms (2016). PoS-constructions are the single-language constructions that code the various slots in Table 1. PoS-constructions are found in all languages, since each language in some way codes each slot in the table, but the specific features that define the construction that codes each slot in each

---

28. Both types of maps (thus, both types of definitions of PoS) have been proposed in the literature. PoS are basically semantic notions in Thompson (1988), Dixon (2004) and Haspelmath (2012: 122–4), discourse-based, pragmatic categories in Hengeveld (1992: 51ff.) and Hopper & Thompson (1984), and discourse-based categories further specified in terms of their prototypical semantics in Croft (2001: 87).

language are language-particular.[29] Nonetheless, constructions can be compared in terms of their function, that is, on the space that they occupy on the map.

Clearly, comparison cannot be immediate. In each language, several constructions can code the same slot, having overlapping functions, and several slots can be coded through a single construction. In Latin, the Quality Predicate can be coded through a verb-like or a noun-like construction: *aquae tepent* vs. *aquae tepidae (sunt)* 'the waters are warm'. The Quality Predicate slot, therefore, is split between two constructions with similar (though not identical) meanings, the latter of which is more frequent than the former. On the other hand, in Latin the Verb construction typically codes only the Action Predicate slot, but in the so-called omnipredicative languages (e.g., Classical Nahuatl, Launey 1994) a single construction codes the Action, the Quality and the Object Predicate slot, taking over three different slots (see Stassen 1997: 29 for the notion of takeover). However, for the sake of argument we establish that a single construction can take over several slots, but each slot has to be linked only to a single construction in each language.

The method whereby each slot is linked to a single construction is typicality, which is measured objectively as text frequency. If two constructions code the same slot in Table 1, a corpus of text is gathered, the relative frequency of each construction is measured and only the more frequent construction is projected on the map. Thanks to this method, the most typical (i.e., most frequent) construction that codes say, the NOUN slot (i.e., the Object Referent) in language X is considered *the* Noun construction of that language and it is the only construction mapped onto the NOUN slot, even if rigorously speaking it is not *only the* Noun construction that codes the NOUN slot in that language. Thanks to this method, the problem of distinguishing major word classes and minor word-classes or sub-classes can be overcome in a principled way, since minor classes and sub-classes are those that are defined by less frequent constructions or by constructions that code only a part of a slot (see fn. 22).[30]

Text frequency supplies an objective method to anchor a single construction to each slot in Table 1. Still, the single-language constructions so identified do not

---

29. PoS-constructions are termed differently depending on the approach of the research and of the level of language analysis at which the construction is fixed. They can be labelled as *token classes* (Broschart 1997; Vogel 2000), *phrasal categories* (Gil 2016; Mosel 2017) or, in the inflectional IE languages, simply *word classes* (Haspelmath 1996). In the following, we will confine ourselves to using the labels *construction classes* or *word classes*.

30. Clearly, minor constructions shall not be ignored, but they are looked at secondarily, only after the major word classes are defined thanks to the major constructions.

necessarily code exactly the same space in Table 1, therefore they do not share exactly the same function. The Latin Verb (i.e., the Action Predicate) construction codes a single slot in Table 1 (the Action Predicate), but the Classical Nahuatl Verb construction typically codes all three slots in the Predicate row (the Action, Quality and Object Predicate). However, comparability is always partial, or it is identity (Moravcsik 2016). If two constructions code the same slot in Table 1, they are comparable in relation to that slot, although the overall space that they code in the table is different in absolute terms. In fact, the ultimate purpose of typologies based on pure CCs is to discover how a given zone of conceptual space is coded across languages, not to discover how the typological space coded by a given single-language construction is coded across languages, which is the ultimate objective of the studies based on portable CCs.

4.3 PoS-lexemes

The third and most particular layer in the continuum is the level of portable CCs or "strategies" in Croft's term (2016). At this layer, PoS are the classes of simple lexical units (or *lexemes*) that enter the constructions.

Also, PoS-lexemes are universal from a certain point of view, but their universality differs both from that of PoS-concepts and of PoS-constructions. If PoS-concepts are the same in all languages and PoS-constructions are found in all languages with the same function but different forms, PoS-lexemes are found in all languages, but they have different forms and also different functions from language to language. A given class of lexemes can enter only one PoS-construction or may enter more, as the case may be: even if all languages define classes of lexemes, only some languages define a specific class of lexemes that enters only the Adjective (i.e., Quality Modifier) construction, as it is the case in Latin and in all languages with "true adjectives" (or with lexical inventory of type [N, A, V]); in other languages a single class of lexemes is the input form on which say, both Adjective and Verb constructions are built, as is the case in languages with verb-like adjectives (or with lexical inventory of type [N (AV)]).[31]

Once a single construction is anchored to each slot in Table 1 by means of text frequency, a principled way to extract lexemes from constructions must be defined, so as to stop splitting lexeme classes *ad infinitum*. Also, in this case, the

---

31. PoS-constructions and PoS-lexemes differ also in single language grammar (Haspelmath 1996). The IE *nomina actionis* (Skt. *vard^h-ana-* 'growing') and the Semitic *maṣdar* (Ar. *dars* 'studying') are nouns in terms of their word or construction class, but they are productively derived from verbal roots, so they fall into the class of verbal roots in terms of their lexeme class (Skt. *vard^h-* 'to grow', Ar. *d.r.s.* 'to study').

method is simple and is based on pure distributional reasoning. Only the constructions projected on the table are used to define lexeme classes. That is to say, we do not require that the lexemes falling into a single class be totally identical from a distributional point of view, which is clearly impossible (see the references in fn. 8, 9 and 14 for discussion). We confine ourselves to saying that the constructions mapped in Table 1 are the only distributional environments employed to define lexeme classes and that items which show the same behaviour with respect to these constructions are equated for the purposes of our typology (that is, they are grouped into one class). Also, in this case, the units compared are similar rather than identical; but if the similarity is defined in an objective manner, the comparison is consistent and any methodological opportunism (as Croft puts it, 2001: 70ff.) is avoided. In this way, the major lexeme classes in each language are objectively defined on their distributional privileges relative to the major constructions, avoiding heterogeneous and non-hierarchical criteria.

Obviously, we may discuss how many constructions are necessary to define the PoS system of a language. A different PoS theory emerges if we consider only the unmarked correlations or all the constructions in Table 1: the lower the number of slots (and thus of constructions) considered, the greater the possibility of lumping two or more lexeme classes into one; conversely, the higher the number of slots, the greater the possibility of splitting one lexeme class into two. However, in either case, the number of lexeme classes to be accepted in each language depends on the distribution of lexical units in the constructions projected on the table, and the number of constructions on the table is determined in an objective manner, through text frequency. The PoS theory proposed so far, therefore, is a "lumping" theory in Croft's terms, but it supplies a consistent method to define different lexical inventories across languages.

In sum, PoS-concepts, PoS-constructions and PoS-lexemes represent three types of CCs, each of which relates to a different layer of generality in the descent that goes from LANGUAGE to single languages, but scholars usually fail to distinguish them clearly.[32] The three meanings of the notion of PoS are summed up in the following table (Table 2):

---

32. Dryer (1997), Croft (2001) and Haspelmath (2012) accept PoS-concepts, but do not clearly divide PoS-constructions and PoS-lexemes, and consider both only as DCs. Hopper and Thompson (1984) focus on PoS-constructions, but disregard PoS-lexemes, while Dixon (2004: 2) defines PoS as lexemes but considers lexemes as primitive notions, although lexemes are the result of the work of speakers (or linguists) who extract them from the constructions in which they appear. Finally, Hengeveld defines PoS both as constructions (1992: 51) and as lexemes (1992: 61), but does not clearly distinguish between the two plans.

Table 2. The three meanings of the traditional notion of PoS

Type of CC	Level of analysis	Type of classes	Graphic symbolization
Pure	Concepts	PoS-concepts	NOUN, VERB, ADJECTIVE…
Hybrid	Discourse	PoS-constructions	Noun, Verb, Adjective…
Portable	Lexicon	PoS-lexemes	noun, verb, adjective…

In the following, the PoS theory sketched above is used to compare the PoS system of Latin and RV Sanskrit. For reasons of space, only unmarked correlations are considered. However, Alfieri (forth.) has shown that the same result is also obtained if the analysis is enlarged to all the slots in Table 1.

5. The Latin PoS system

The PoS system in Latin is well known. The distinction between PoS concepts, constructions and lexemes is thus expected to confirm what is already common knowledge, rather than bring in substantial novelties.

The most typical Latin construction that codes the NOUN (Object Referent) slot in Table 1 is a "noun" in the most canonic sense: a word-form marked by case, number and gender, or simply: […]-Case.[33] The […]-Case construction can be filled by different types of stems: a simple noun stem (*milit-* in *miles* 'soldier'), a derived verb stem (*amant-* 'lover' from *amo* 'I love') and a simple or derived adjective stem (*album* 'list' from *albus* 'white'), etc. However, there is little doubt that simple noun stems are the most typical (i.e., the most frequent) fillers of the […]-Case constructions. The most typical Latin Noun construction, therefore, is [noun]-Case (Example 1):

> (12) **arm-a**   **viru-m**=*que*   *can-o*
> weapon(NT)-ACC.PL man(M)-ACC.SG=and sing-PRS.1SG
> 'I sing the weapons and the man'   (Verg., *Aen*. I.1)

Similarly, the most typical construction that codes the ADJECTIVE (Quality Modifier) slot in Table 1 is the Latin Adjective construction. This construction is a word-form marked by agreement (and case, gender, number), or simply: […]-Agr. The […]-Agr construction can be filled by a simple adjective stem (*magn-* in *magnus*

---

33. The label […]-Case is a summary label that includes all the inflectional features defined in Section 3 (for Latin) and Section 4 (for Sanskrit), and the same holds for the labels […]-Agr and […]-Pers.

'big'), a participle (*notus* 'known' from *nosco* 'know') or a noun joined to an adjectival suffix (*gloriosus* 'proud' from *gloria* 'glory'). However, the most typical filler of the Adjective construction is a simple adjective stem, so the most typical Adjective construction in Latin is [adjective]-Agr (Example 13):

(13) *tibi*   *ne*   **tener-as**   *glacie-s*   *sec-e-t*
      you.DAT   not   soft-F.ACC.PL   ice(F)-NOM.SG   cut-SBJ-3SG

      asper-a   *planta-s*
      rough-F.NOM.SG   palm(F)-ACC.PL

      'Ah, might the jagged ice not cut your tender feet'   (Verg., *Ec.* X.49)

In the same way, the most typical construction that codes the VERB (Action Predicate) slot in Table 1 is the Latin Verb construction, which is a word-form marked by person, tense, mood and diathesis: schematically […]-Pers. The […]-Pers construction can be filled by a verb stem (*cad-* in *cado* 'fall'), an adjective stem joined to a verbalizing affix (*albesco* 'become white' from *albus* 'white') a noun stem joined to an empty verbalizing affix (*maculare* 'to stain' from *macula* 'stain'), etc. However, the most typical filler of this construction is a simple verb stem and the most typical Latin Verb construction is [verb]-Pers (Example 14):

(14) *miser*   *Catull-e*   **desina-s**   *inepti-re*
      miserable.M.VOC.SG   Catullus(M)-VOC.SG   cease-PRS.2SG   be_a_fool-INF

      et   *quod*   **vide-s**   *peri-sse*
      and   that.NT.ACC.SG   see-PRS.2SG   be_lost-INF.PST

      perd-itum   *duc-as*
      lose-PTC.NT.ACC.SG   consider-SBJ.PRS.2PL

      'Miserable Catullus, cease to be a fool, and that which you see to have been
      lost may you consider lost'   (Cat., *Car.* VIII.1)

Obviously, those exemplified above are the most frequent constructions that code the unmarked correlation defined in Section 3, but not the only constructions possible for those functions. For instance, in the first 40 chapters of Sallust's *De coniuratione Catilinae*, 393 Adjective constructions (i.e., Quality Modifiers) are found: 334 (85.0%) are (prefixed) simple adjectives (e.g., *bonus* 'good', *obscurus* 'obscure'), 30 (7.6%) are (prefixed) deverbal adjectives (*adulescens* 'young', *invisus* 'hateful'), and 29 (7.4%) are (prefixed) denominal adjectives (*urbanus* 'urban', *egregius* 'illustrious'). If all the constructions found in the corpus are gathered, the following table is obtained (Table 3):[34]

---

34.   The appositions (e.g., *consul* 'consul', *senator* 'senator', *eques* 'knight', etc.) are not included, since they are considered as non-typical Object Modifiers.

Table 3. Latin Adjective constructions table (the number before the schema of the construction refers to frequency)

Adjectives	Number	Percent
1. [adjective]-Agr	**334**	**83.7%**
1. [adjective]-Agr	307	78.1%
2. Pre-[adjective]-Agr	27	6.9%
2. [...]$_V$-Agr	**30**	**7.5%**
4. [verb-NM]-Agr	16	4.1%
5. Pre-[verb-NM]-Agr	14	3.6%
2. [...]$_N$-Agr	**29**	**7.4%**
3. [noun-ADJ]-Agr	25	6.6%
7. Pre-[noun-ADJ]-Agr	4	1.0%
Total	**393**	**100%**

If only the most frequent constructions that code the Noun, the Verb and the Adjective function in Latin are mapped in Table 1, Table 4 is obtained:

Table 4. Latin constructions table (only unmarked correlations are reported)

	Referent	Modifier	Predicate
Object	**[noun]-Case**	----	----
Quality	----	**[adjective]-Agr**	----
Action	----	----	**[verb]-Pers**

The constructions [...]-Case, [...]-Agr and [...]-Pers define three classes of simple lexemes: nouns, adjectives and verbs. Each class has a specific and exclusive distribution: the nouns enter only the Noun construction without further measures being taken (as Hengeveld says, 1992: 58), the verbs enter only the Verb construction without further measures and the adjectives typically enter only the Adjective construction, but they can also enter the Noun construction without further measures.[35] Nouns and adjectives, therefore, represent two different classes of lexemes, the nouns being marked as [+ gender] and the adjectives as [+ agreement], but their difference is neutralized in all the slots on the Referent row, where the feature [+ agreement]

---

35. Hengeveld's wording refers to the absence of trans-categorization devices, be they coded overtly through an affix, or covertly through syntactic conversion (see also Croft 1991: 58, 2001: 66).

is not pertinent (see fn. 12). The threefold division of the lexicon is mirrored in derivation. Each Latin suffix attaches to a single class of lexemes and produces items belonging to a single class of derived stems, as seen in the case of adjectival-forming and adjectival-selecting affixes (see Section 3). Finally, syntax mirrors the threefold division established on the lower levels and defines three classes of constructions with exclusive inflectional features. If Table 4 is projected orthogonally dividing the lexeme and the word layer, the PoS system in Latin is obtained (Table 5):[36]

Table 5. Latin PoS table (only the primary categorization is reported; the arrows refer to grammatical processing)

	NOUN	ADJECTIVE	VERB
Phrase			
Derived stem			
Simple stem	[noun]-Case	[adjective]-Agr	[verb]-Pers
	↑	↑	↑
Lexeme	noun	adjective	verb

Traditional Latin grammars, which conflate all levels of language structure in a single word layer, conclude – quite simplistically – that Latin has three different PoS, although adjectives are similar to nouns. And modern scholars conclude that the PoS are defined by different criteria that do not overlap totally (Lyons 1979: 42), since the same threefold division found at the level of the lexicon holds also in derivation and in syntax, although it is manifested through different features at each level.

6. The RV Sanskrit PoS system

The same constructions that are found in Latin are also found in Vedic. In both cases, the most typical Noun construction is a word marked by case and the most typical filler of the [...]-Case construction is a simple noun (Example 15):[37]

---

36. Items can be re-categorized several times in IE languages (Simone 2007; Ježek & Ramat 2009) and they often increase (and stiffen) their categorization passing from a lower to an upper level (Lehmann 2008): *purificatio* 'purification' is an adjective in its lexeme class (*purus* 'pure'), a verb in its derived (or rather compound) stem1 class (*purifico* 'purify') and a noun in its derived stem2 or word class (*purificatio*). However, only the primary (or lexical) categorization and the final (or discourse) categorization of the most typical constructions will be discussed in this case.

37. If not otherwise specified, the translations follow those supplied by Brereton and Jamison (2014).

(15) *pác-ya-te*   *yáva-ḥ*
      cook-PRS4-3SG.MD grain(M)-NOM.SG
      'the grain ripens'   (1.135.8[d])

However, in the RV derived nouns are so frequent that one may doubt whether the most typical Noun construction is indeed [noun]-Case. But a frequency count made on a small sample of 20 RV hymns (the first 20 listed in fn. 41) shows that the simple noun construction outnumbers the derived noun construction by about 60% to 40%.[38] Derived nouns, therefore, may be more frequent in the RV than in Latin, but in both cases the most typical Noun is [noun]-Case.

A slightly more complex situation is found with the VERB. In the RV the most typical Verb construction is a word-form marked by person, as in Latin (Example 16):

(16) **táp-a-nti**   *śátru-ṃ*   *svàr*
      make_hot-PRS1-3PL rival(M)-ACC.SG sun(M).NOM.SG

      ṇá   *bʰūmā*
      as   earth(NT).ACC.PL

      '[the Gods] scorch the rival, like the sun [scorches] the worlds'   (7.34.19[a])

However, two interpretations of Skt. *tápanti* are possible. The simplest segmentation is [root-AFF]-Pers. The affix can be a discontinuous morph if the template-and-pattern morphology is accepted: *t.p- + -á.a- + -nti → tápanti*, but it can also be a suffix, as the Indian grammarians said: *tap- + -a- + -nti → tápanti*.[39] In either case, while in Latin the most typical Verb construction is a simple verbal lexeme marked by person, in Sanskrit it is a 'verbal' lexeme marked by an affix and person.[40]

---

38.   More specifically, 677 Noun constructions are found in the sample: 399 (58.9%) are simple nouns (i.e., [noun]-Case, e.g., *mātár-* 'mother'), 223 (32.9%) derived nouns built from verbal roots (i.e., [root-NM]-Case, e.g., *pṛthivī́-* 'earth', lit. 'the wide' from *prath-* 'to stretch, extend'), 45 (6.6%) are derived nouns built from primary nouns (i.e., [noun-ADJ]-Case, e.g., *rathín-* 'charioteer' from *rátha-* 'chariot') and 10 (1.4%) are miscellaneous construction types. On the whole, noun-based constructions total 445 (65.7%), while root-based constructions number 235 (34.3%).

39.   The template-and-pattern morphology was proposed by Saussure (1878), implicitly, Meillet (1903: 116) and Benveniste (1962: 147ff.) for the IE family. For a discussion on the topic, see Alfieri (2016: 132, fn. 10; 157ff.).

40.   If Aronoff's analysis of Latin verb inflection (1994: 33ff., 39ff., 45ff.) was applied also to RV Sanskrit, the vowel *-a-* should be considered a stem vowel and the typical Verb construction in the RV should appear as [verb]-Pers. But in the RV 10 classes of presents are found, similar to the Semitic *binyanim*: 6 classes are formed through affixes with or without ablaut; 2 classes with the *-a-* but no ablaut (class I, that of *tapati*, which is the most frequent, and class

As in Latin, in Vedic the most typical Adjective is a word marked by agreement: […]-Agr. But Brugmann already knew that primary adjectives are scanty in the Veda (1904: 329) and many of the words usually classified as adjectives are, in fact, nominalizations built on verbal roots of stative or nearly stative meanings: *tap-ú-* 'hot' from *tap-* 'make or become hot, heat', *raṇ-vá-* 'pleasant' from *raṇ-* 'be pleasant, delight', *bṛh-ánt-* 'high, big, lofty' from *bṛh-* 'make big, strong'. It may thus be asked whether the simple adjective stem is really the most typical filler of Adjective construction in Vedic as it is in Latin.

To answer this question, a sample of 51 hymns of the RV was collected, all the Adjective constructions in the sample were gathered and a frequency count was made.[41] In the sample 892 Adjective constructions are found. As in Latin, all constructions are marked by agreement: […]-Agr. However, below syntactic level, these constructions can be divided into five broadly different types. The most frequent filler of the […]-Agr construction is a nominalized root, that is a root joined to one of the primary suffixes called *kṛt* in traditional Indian grammars: [root]-NM-Agr.[42] This construction is attested in 425 cases out of 892 (47.6% of the sample). The nominalizer is a standard *kṛt* suffix in Example 17 and a participial suffix in Example 18:

(17) *kṛṣṇā́dʰvā*　　　　　　　　*táp-ū*　　　　　　　*raṇ-vá-ś*
　　　 of_black_path.M.NOM.SG　 be_hot-NM.M.NOM.SG　 rejoice-NM-M.NOM.SG

　　　 ci~ket-a　　　　 *dyaú-r*　　　 *iva*
　　　 PF~observe-3SG　 sky(M)-NOM.SG　 as

　　　 smáya-mān-o　　　　　*nábʰo-bʰiḥ*
　　　 smile-PTC-M.NOM.SG　　 cloud(NT)-INS.PL

　　　 '[Agni], having a black road, red-hot, he appears bringing delight, smiling like heaven with its clouds'　　　　　　　　　　　　　　　　　(2.4.6^cd)

VI); 1 class with reduplication, which plays the same role as an affix from the structural point of view; 1 class with ablaut but no suffixes. Most Sanskrit roots build several present stems with different meanings: *tapati* 'burns' vs. *tápyate* 'becomes hot'. Aronoff's analysis is thus not plausible if applied to Sanskrit.

41. The sample includes the following hymns. Book 1: 1, 35, 61, 85, 135, 154, 160; book 2: 2, 4, 12, 24, 33, 35; book 3: 7, 49, 59; book 4: 49, 50, 51; book 5: 36, 83; book 6: 5, 16, 47, 54; book 7: 49, 55, 61, 63, 70, 71, 86, 103; book 8: 2, 4, 29, 33, 48; book 9: 1, 2; book 10: 14, 15, 30, 34, 87, 90, 127, 129, 130, 135, 168. The appositions (e.g., *deva…agne* 'God Agni', 6.16.12^c) are not included, since they are considered as non-typical Object Modifiers.

42. *Kṛt* suffixes are glossed as NM since they usually build derived nouns, but they can also build derived adjectives, which are often considered as a special type of agent noun built on roots that do not have an action-centred meaning (MacDonell 1975: 113 on the *-as-* suffix).

(18) *prá*   *nā́ka-m*   **rṣ-vá-ṃ**
      away   vault_of_heaven(M)-ACC.SG elevate-NM-M.ACC.SG

      nu~nund-e   **bṛh-ánt-aṃ**
      PF~push-3SG   make_big-PTC-M.ACC.SG

      '[Varua] pushed forth the vault of heaven to be high and lofty'   (7.86.1ᶜ)

The second most frequent construction is a possessive compound, called the *bahuvrīhi* type by Indian grammarians. This construction is schematized as [...]ₙ-[...]ₙ-Agr and is found in 184 cases out of 892 (20.6%). Below its very general schema, however, different constructions are found. In 118 cases (13.2%), the second member of the compound (its "head", inasmuch as such a notion applies to an exocentric compound) is a derived noun taken from a verbal root: [...]ₙ-[root-NM]ₙ-Agr (Example 19), but in 52 cases (5.8%) it is a primary noun: [...]-[noun]-Agr (Example 20):

(19) *ví*   *suparó*   *antárikṣāṇi*   *a-kʰya-d*
      away   eagle(M).NOM.SG   midspace(NT).ACC.PL   PST-watch-AOR.3SG

      gabʰīrá-vepā   *ásura-ḥ*   *su-nītʰá-ḥ*
      deep-inspiration.M.NOM.PL[43]   lord(M)-NOM.PL   good-guidance-M.NOM.PL

      'the eagle has surveyed the midspaces – the lord possessing profound
      inspiration who gives good guidance'   (I.35.7ᵃ)

(20) **híraṇya-pāṇi-ḥ**   *savitā*   *ví-carṣaṇi-r*
      gold-hand-M.NOM.SG   Savitar(M).NOM.SG PRE-boundary-M.NOM.SG[44]

      ubʰé   *dyā́vā-pṛthiv-í*   *antár*   *īyate*
      both.F.ACC.DU heaven-earth-F.ACC.DU between   go.3SG.MD

      'Golden-palmed Savitar, whose boundaries are distant, shuttles between
      both, both heaven and earth'   (1.35.9ᵃᵇ)

The third most frequent construction coding the Adjective function is a noun joined to one of the prefixes *su-*, *dus-*, *nis-*, *sa-*, *a-*. This construction is schematized as Pre-[...]ₙ-Agr and is found in 133 cases out of 892 (14.9%).[45] In 103 cases

---

43. The word *vépas-* 'inspiration' is a regular action noun in *-as-* taken from the root *vip-* 'tremble'.

44. The adjective *vícarṣaṇi-* is of unclear meaning, although it is usually traced to *karṣ-* 'to drag, plough' with the suffix *-ani-* (Thieme 1967: 236ff.).

45. From the Indian point of view, this construction is a compound. However, unlike the nouns that build compounds, prefixes cannot stand alone in a sentence (a partial exception being represented by *su-*, which can also be found as an independent adverb, although it is much more frequent if prefixed to a noun).

(11.5%) the noun filling the construction is a derived noun taken from a verbal root: Pre-[root-NM]-Agr (Example 21), whereas in 23 cases (2.6%) it is a simple noun: Pre-[noun]-Agr (Example 22):

(21)   *huv-é*   *vaḥ*   **su-dyót-mān-am**
       invoke-1SG.MD you.PL.DAT good-brighten-NM-M.ACC.SG

       su-vr̥k-tí-m   *viś-ā́m*   *agní-m*
       well-twist-NM-M.ACC.SG clan(F)-GEN.PL Agni(M)-ACC.SG

       átitʰi-m   **su-pray-ás-am**
       guest(M)-ACC.SG good-please-NM-M.ACC.SG

       'I call for you the one of good brilliance, on Agni, the guest of the clans, who receives well-woven [hymns], who receives very pleasurable offerings'
                                                (2.4.1[ab])

(22)   *babʰrú-r*   *éko*   *víṣuṇa-ḥ*
       brown-M.NOM.SG one.M.NOM.SG changing-M.NOM.SG

       sūnáro   *yúvā*
       well.spirit.M.NOM.SG youth(M).NOM.SG

       'brown, this one [the Soma] is changeable, a spirited youth'   (8.29.1[a])

The fourth most typical Adjective construction is any (simple or derived) noun stem joined to one of the secondary suffixes that are termed as *taddʰita* suffixes by Indian grammarians: [..]$_N$-ADJ-Agr.[46] This construction is found in 94 cases out of 892 (10.5%). In 54 cases (6.1%) the nominal stem filling the construction is a simple noun: [noun]-ADJ-Agr (Example 23), whereas in 40 cases (4.4%) it is a derived noun built on a verbal root: [root-NM]-ADJ-Agr (Example 24):

(23)   *iṣ-iréṇa*   *te*   *mánas-ā*   *su-tá-sya*
       incite-NM.NT.INS.SG you.GEN.M.SG mind(NT)-INS.SG press-PTC-M.GEN.SG

       bʰakṣ-īmáhi   *pítr-iya-syeva*   *rāy-áḥ*
       partake-PRS.OPT.1SG father-ADJ-M.GEN.SG.as wealth(M)-GEN.SG

       'With a vigorous mind we would take a share of you when pressed, as of ancestral wealth'   (8.48.7[ab])

(24)   **sah-ā́-vā**   *pr̥tsú*   *tar-áṇi-r*   *ná*
       prevail-NM-ADJ.M.NOM.SG battle(F).LOC.PL pass-NM-M.NOM.SG as

       árvā   *vi-ā-naś-ī́*
       steed(M).NOM.SG PRE-PRE-traverse-NM.M.NOM.SG

---

46. *Taddʰita* suffixes are glossed as ADJ (adjectivalizer), since their typical function is that of building relational adjectives from nouns, although they can also build diminutives or other types of nouns.

ródasī *meh-ánā-vān*
world_half(NT)-ACC.DU urinate-NM-ADJ.M.NOM.SG

'[Indra] victorious in battles like an overtaking steed, traversing the two world-halves, streaming abundance' (3.49.3[ab])

The fifth most typical filler of the Adjective construction is a simple adjective stem marked by agreement, which is exactly the same construction as in Latin. The construction [adjective]-Agr is found in 56 cases (6.3%), see Example 25:

(25) *āmād-aḥ* *kṣvíṅkās* *tám*
raw_meat.eat-NM.F.NOM.PL spirit(F).NOM.PL 3SG.ACC

ad-antv *énī-ḥ*
eat-IPT.3PL colourful-F.NOM.PL

'let the *Kṣvíṅkā*-spirits, eaters of raw meat, of variegated colour, eat him [sc. the sorcerer]' (10.87.7[d])

From a merely factual point of view, almost the same constructions that are found in the RV are found also in Latin: the deverbal adjective (*fervens* 'hot'), the compound adjective (*frugi-ferens* 'fruitful, which brings harvest'), the prefixed adjective (*obscurus* 'obscure'), and so on. However, the statistical distribution of these constructions is different in the two languages. If all the constructions found in the RV are gathered, the following table is obtained (Table 6):[47]

Table 6. RV Adjective construction table (the number before the schema of the construction refers to frequency)

Adjectives	Number	Percent
1. [...-NM]$_A$-Agr	**425**	**47.6%**
1. [root-NM]$_A$-Agr	425	47.6%
2. [...]$_N$-[...]$_N$-Agr	**184**	**20.6%**
2. [...]$_N$-[root-NM]$_N$-Agr	118	13.2%
6. [...]$_N$-[noun]$_N$-Agr	52	5.8%

(*Continued*)

---

47. Constructions with a frequency below 1% are not exemplified. Comparative and superlative suffixes are treated differently depending on the construction in which they appear: the suffix *-tama-* is glossed as NM if attached to a root and as ADJ if attached to a derived stem. Moreover, in order not to multiply the construction types, comparative and superlative suffixes are disregarded if attached to simple adjectives or to adjectives already attached to *tadd^hita* suffixes: this means that *návīyasā* 'newer' (6.16.21[a]) is included in the [adjective]-Agr pattern, but *vīrávattamam* 'the richest in heroes' (1.1.3[c]) is counted as an instance of the [noun]-ADJ-Agr pattern.

Table 6. (*Continued*)

Adjectives	Number	Percent
9. [...]$_N$-[root]$_N$-Agr	7	0.8%
10. [...]$_N$-[noun-ADJ]$_N$-Agr	6	0.7%
14. [...]$_N$-[root-NM-ADJ]$_N$-Agr	1	0.1%
3. Pre-[...]$_N$-Agr	**133**	**14.9%**
3. Pre-[root-NM]$_N$-Agr	103	11.5%
8. Pre-[noun]$_N$-Agr	23	2.6%
11. Pre-[...]-[root-NM]$_N$-Agr	3	0.3%
12. Pre-[root]-Agr	3	0.3%
13. Pre-[...]-[root]-Agr	1	0.1%
4. [...]$_N$-ADJ-Agr	**94**	**10.5%**
5. [noun]$_N$-ADJ-Agr	54	6.1%
7. [root-NM]$_N$-ADJ-Agr	40	4.5%
5. [...]$_A$-Agr	**56**	**6.3%**
4. [adjective]$_A$-Agr	56	6.3%
Total	**892**	**100%**

If the constructions are grouped under the class of the lexical items they are constructed upon, it comes out that in the RV the most typical lexeme class used to build the Adjective construction is not the class of simple adjectives, as in Latin, but rather the class of verbal roots and, specifically, verbal roots of quality meaning. See Table 7:

Table 7.   RV Adjective constructions table (version 2)

Adjectives	Number	Percent
1. root	**701**	**78.4%**
1. [root-NM]$_A$-	425	47.6%
1. [root-NM]$_A$-Agr	425	47.6%
2. [root-NM]$_N$-	265	33.6%
2. [...]$_N$-[root-NM]$_N$-Agr	118	13.2%
3. Pre-[root-NM]$_N$-Agr	103	11.5%
7. [root-NM]$_N$-ADJ-Agr	40	4.5%
13. [...]$_N$-[root-NM-ADJ]$_N$-Agr	1	0.1%
11. Pre-[...]-[root-NM]$_N$-Agr	3	0.3%

(*Continued*)

Table 7. (*Continued*)

Adjectives	Number	Percent
3. [root]-	11	9.8%
9. […]_N-[root]_N-Agr	7	0.8%
12. Pre-[root]-Agr	3	0.3%
14. Pre-[…]-[root]-Agr	1	0.1%
2. noun	**134**	**15.2%**
5. [noun]-ADJ-Agr	53	6.1%
6. […]_N-[noun]-Agr	52	5.8%
8. Pre-[noun]_N-Agr	23	2.6%
10. […]_N-[noun-ADJ]_N-Agr	6	0.7%
3. adjective	**56**	**6.3%**
4. [adjective]-Agr	56	6.3%
Total	**892**	**100%**

If these constructions are mapped on the PoS table in Table 1, the following construction table is obtained (Table 8):

Table 8. RV construction table (the root-and-pattern analysis of the verb is accepted)

	Referent	Modifier	Predicate
Object	**[noun]-Case**	----	----
Quality	----	**[root-NM]-Agr**	----
Action	----	----	**[root-AFF]-Pers**

As distinct from Latin, constructions […]-Case, […]-Agr and […]-Pers define only two major classes of lexemes in the RV, since the root, joined to different affixes, is the most typical filler both of the […]-Agr and of the […]-Pers constructions.

Saying that only two classes of lexemes are found in the RV is an oversimplification. A class of primary adjectives is found, but the class is notably small. In Grassmann's dictionary (1976[5]), bar function words, 1007 primary lexemes are listed: 565 roots (56%); 410 primary nouns (40%); 38 primary adjectives (4%), as shown in Table 9:[48]

---

48.   Strictly speaking, Grassmann lists 34 primary adjectives, to which 4 further cases have been added by Alfieri (2016). The list is the following: *aghá-* 'bad', *ánūna-* 'complete', *árbʰa-* 'little', *ásita-* 'black', *āmá-* 'raw', *āhanás-* 'swollen', *āśú-* 'fast', *írya-* 'active', *udumbalá-* 'reddish',

Table 9. RV lexicon (1 = roots, 2 = nouns, 3 = adjectives)
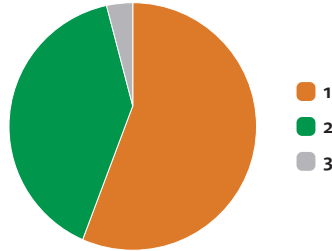
In the RV, therefore, a small class of adjectives is found. But languages with small classes of adjectives (usually 10–20 members, see Dixon 2004: 10) are often merged with languages "without adjectives", since in both cases the most typical Adjective construction is not coded by a dedicated class of adjectival lexemes stored in the lexicon. Instead, it is coded by a complex construction built by the noun-like or the verb-like lexemes that refer to quality meanings.

Moreover, simple adjectives are not only scanty in number in the RV. If those continuing a PIE adjective are excluded (*āmá-, nagná-, mádʰu-*, etc.), Vedic adjectives are etymologically derived from verbal roots and underwent a lexicalization process at some intermediate stage between PIE, Indo-Iranian and the RV. Skt. *gurú-* 'heavy' is etymologically traceable to **gr̥̄-* 'be tired, onerous' (ᴠɪᴀ 403), from PIE **gʷerh₂-* (see Lat. *gravis* 'heavy', Gk. βαρύς 'id.'); the root **gr̥̄-* is found in a handful of derivatives such as *gárīyas-* 'heavier', *gariṣṭʰa-* 'heaviest', *grávan-* 'a (heavy) stone for pressing the soma' (ᴇᴡᴀia 490), but the word-formation rules needed to build them are not synchronically productive in the RV. Therefore, *gurú-* cannot be productively (that is, synchronically) derived from **gr̥̄-*, but the relation between the two forms is etymologically (that is, diachronically) clear.[49]

To sum up, in the RV a different organization is found on the lexeme and on the construction or word layer. On the lexeme layer, only two major word classes

---

éni- and *éta-* 'dappled, rushing', *kalmalīkín* 'brown', *kalyá-* 'lovely, beautiful', *kr̥ṣṇá-* 'black', *gurú-* 'heavy', *tílvila-* 'rich', *tīvrá-* 'sharp', *dīná-* 'weak', *dīrgʰá-* 'long', *nagná-* 'naked', *náva-* 'new', *nīla-*° 'dark', *palitá-* 'grey', *purú-* 'many', *pūrva-* 'former', *babʰrú-* 'brown', *bradʰná-* 'pale red', *mádʰu-* 'sweet', *mádʰya-* 'middle', *yaśás-* 'glorious', *pāpá-* 'bad', *pŕ̥śni-* 'spotted', *róhita-* 'red', *śabála* 'dappled', *sána-* 'old', *stʰūrá-* 'dense, thick'. These adjectives do not fall exactly in Dixon's list of basic adjectival meanings.

49.   For an analysis of all of the RV primary adjectives, see Alfieri (2016: 152–4). The theoretical basis of that analysis is laid down by the group of Natural Morphology (Mayerthaler 1981; Dressler 1987; Panagl 1987), which investigated how semantic and formal opacity interact to determine the autonomous storage of formerly derived items. On the topic, see also Bertram, Schreuder & Baayen (2000).

are found – primary nouns and roots – in addition to a small set of primary adjectives that, as a rule, are historically derived from verbal roots. Derivation mirrors the twofold division of the lexicon: primary or *kṛt* suffixes attach only to roots and build derived nominal stems, whereas and secondary or *taddʰita* suffixes attach to simple or derived nominal stems and build secondarily derived nominal stems (see Section 3). Some (primary and secondary) derived nominals are more readily used as nouns, some others are more readily used as adjectives, but many of them can be either nouns or adjectives, depending on the context. At the levels of syntax, the sum of simple adjectives and derived adjectives determines the birth of a third major class of constructions, the Adjective. The three classes of construction are defined through almost the same inflectional features that are found in Latin, although some of the criteria that distinguish the Adjective and the Noun construction in Latin do not hold in Sanskrit. Therefore, while in Latin the 'adjective' is a class of simple lexemes, a class of derived stems and a class of words, in the RV it is a class of words, but is not or is only limitedly a class of simple lexemes and a class of derived stems (Table 10):

Table 10. RV PoS table (the arrow signalling the processing of derived nouns is dotted, since it does not represent the most common strategy for coding the Noun)

	NOUN	ADJECTIVE	VERB
Phrase			
Derived stem	[root-NM]-Case	[root-NM]-Agr	[root-AFF]-Pers
Simple stem	noun-Case		
Lexeme	noun	root	

In sum, the difference between the PoS systems in Latin and in RV Sanskrit does not lie in the features that define the classes of words, it rather lies in three facts: i) the number of major classes of lexemes (three major classes in Latin, only two in the RV); ii) the level at which the most typical Adjective construction is grammaticalized, that is to say fixed (the simple stem in Latin, the derived stem in the RV); iii) the categoriality (that is, the function) of the verbal lexeme (a monocategorial unit that typically enters only the Verb construction in addition to some Noun and Adjective constructions in Latin, a pre-categorial or poly-categorial unit that typically codes the Verb and the Adjective constructions in addition to quite a few Noun constructions in RV Sanskrit).[50] Traditional Sanskrit grammars, which tend

---

50.   On the notion of pre-categoriality, see Bisang (2008) and below Section 7.

to conflate all the levels of language structures into a single word layer, just as Latin grammars do, usually disregard these differences (or describe them only indirectly and in a very limited manner) by saying – again, quite simplistically – that the PoS system in Latin and in RV Sanskrit are almost identical, although in Sanskrit the adjective is not as clearly distinguished from the noun as it is in Latin (see Section 3).

7.   Discussion and conclusion

PoS theory has always swung between the extreme poles of language universalism and language particularism – the two positions of the pendulum described by Bossong (1992, see Introduction). Scholars working in Generative Grammar consider the noun-verb-adjective distinction as universal.[51] Consistently with this view, they developed a morphological theory based on that universality, that is, they consider that lexical morphemes are by nature equivalent to stems and to word-forms, that word-forms (thus also stems and lexemes) are universally divided into the three traditional classes of nouns, verbs and adjectives and that each affix necessarily attaches to a single class of words (noun, verb or adjective) and produces derived words necessarily cast in one of these classes (noun, verb or adjective).[52] Therefore, in their view the term *root* by nature refers to the same type of linguistic unit to which the label 'simple verb stem' is also referred, and the PoS system in Latin is almost identical to that in Sanskrit – a conclusion that is not very different from that reached by traditional Sanskrit philologists, who in practice describe the Sanskrit PoS system through the lens of Latin-based grammar.

On the other hand, many other scholars working in typology think that linguistic forms can be considered as "lacking categoriality completely

---

51.   This idea can be traced to Chomsky (1970), and has become the standard in Generative Grammar (see, e.g., Haegeman 1994: 36ff.), but has never been tested empirically, bar Baker's work (2003: 11ff.). However, Baker defines the noun, the verb and the adjective as unweighted bundles of features, so his definition of PoS has the same problems as Dixon's definition of the adjective (see Section 3.1.2).

52.   Reference is made to the *Lexicalist Hypothesis*, that is the idea that input-forms for word-formation rules are always words, divided into the universal classes of noun, verb and adjective (Chomsky 1970; Halle 1973: 10; Aronoff 1976: 46), and to the *Unitary Base Hypothesis*, that is the idea that input forms and output forms of word-formation rules must be words – nouns, verbs or adjectives (Aronoff 1974: 47–8; Scalise 1984: 137). Both ideas are at the basis of any version of Generative morphology (see Bauer 2003: 9; Lieber 2010: 3; Aronoff 2012: 29–30), but they have spread also among different frameworks, such as word-based morphology (Matthews 1972: 163 fn. 3, 1974: 40; Booij 2007: 28, 321) and typology (Lehmann 2008).

unless nounhood or verbhood is forced on them by their discourse functions"
(Hopper & Thompson 1984: 747, but see also Maranz 1997) or that "languages
differ without limits and in unpredictable ways" (Joos 1957: 96) in their lexeme
class division. In this latter view, each language defines its own classes, so that the
PoS system in Sanskrit can be different from that in Latin, but a consistent typol-
ogy of lexical inventories across languages cannot be obtained (or at least has not
yet been obtained).

The data discussed so far show that both positions are excessive. Languages
do not always define the same classes, nor do they avoid defining any class: the
number of lexeme classes defined in each single language is a cross-linguistic vari-
able, but a PoS theory that orders such variability can be developed, at least for
the IE languages. If the continuum that goes from LANGUAGE to the IE languages
is divided into three layers (PoS-concepts, PoS-constructions, PoS-lexemes), an
objective method to anchor PoS-constructions to PoS-concepts and an objective
method to extract PoS-lexemes from PoS constructions are found, at least two
different PoS systems can be defined: type [N, A, V], which is found in Latin and,
more generally, in Hengeveld's *specialized* languages (1992: 47ff.), and type [N
(AV)], which is found in all of Dixon's verb-adjectival languages (2004) and in
the non-specialized languages defined by Hengeveld (1992: 74ff.), be they *flexible*
such as Lao or *rigid* such as RV Sanskrit.[53] The PoS system in Sanskrit, therefore,
is different from that in Latin and the root is a different unit from the simple verb
stem. Both are simple verbal lexemes, but only the Latin verbal lexeme is also a
stem (thus, a word-form from which inflections have been subtracted), while the
root is the verbal lexeme of a language in which only two major classes of lexemes
are found and adjectives are coded verbally. It is a precategorial unit, or a lexeme
that assumes stem- or word-form only when it attaches to the derivational mor-
phemes (affixes and ablaut) that specify its categorial status as a noun, a verb or
an adjective.[54]

---

53. In both flexible and rigid languages one or more lexical classes are 'lacking' (if compared
with the specialized type [N, A, V]). The difference between the two types lies in the strategy
used to code the function typically associated with the 'lacking' lexeme class: flexible lan-
guages use a single class of lexemes in different functions without any formal change, while in
rigid languages a lexeme class is lacking and its function is coded through a periphrasis built
on one of the extant lexeme classes (Hengeveld 1992: 65ff.). However, if flexible languages
represent a coherent group, whose discovery has led to important generalizations (see fn. 27),
rigid languages do not seem to share much among one another, bar their being non-special-
ized and non-flexible (Alfieri 2013).

54. The formal coding of linguistic functions is arbitrary in principle (Sapir 1921: 59).
However, the template and pattern morphology (fn. 39) has an iconic function in IE (and in

If this view is accepted, the universality of the noun-verb-adjective distinction implicitly assumed by Sanskrit scholars and explicitly claimed by Generativists must be abandoned, alongside all of its corollaries, starting from the substantial identity of the PoS systems in Latin and Sanskrit, the structural equivalence between lexemes, stems and word-forms (apart from inflections), and more specifically the functional equivalence between the Sanskrit root and the Latin simple verb stem, on which Generative morphology is based. At the same time, if the method whereby Latin and Sanskrit have been compared is consistent, there is no reason to deny that it can be applied to other languages as well, so as to develop a consistent typology of the lexeme class inventories across languages, just as proposed in the Amsterdam Typology.

Clearly, these conclusions are not right nor wrong absolutely, they are simply consistent with the PoS theory proposed. Nor is the PoS theory right or wrong absolutely, it is arbitrary; one of many possible PoS theories. However, this theory and the results that descend from it can be (simplistically) considered "right" inasmuch as it is useful, that is, inasmuch it leads to relevant insights in language description and comparison. The insights that can be arrived at in Sanskrit descriptive grammar have been described in Section 6. In the following, therefore, the utility of the PoS theory proposed in historical and general-typological linguistics will be discussed.

7.1   Historical IE linguistics

Given the genetic relation between Sanskrit and Latin, the difference between the PoS systems in the two languages must be the result of a diachronic change. Theoretically the change can be either [N (AV)] → [N, A, V] or [N, A, V] → [N (AV)]. However, there are good reasons to think that the direction of the change is [N (AV)] → [N, A, V].

Most of the 38 primary adjective stems in the RV are the result of the autonomous storage of formerly derived items (see Section 6). Moreover, only 9 primary adjective stems are listed in Pokorny's *Indogermanisches etymologisches Wörterbuch* (1956) and only 17 primary adjective stems are found in *Nomina im Indogermanischen Lexikon* by Wodtko et al. (2008): in both cases RV Sanskrit shows more primary adjectives than those that are reconstructed for PIE,

---

Semitic, see Section 7.2): the stems, which are functionally equivalent to word-forms (fn. 10), are fully specified stems also phonetically, but the roots, which are consonantal templates, cannot be uttered nor be attached to endings if they are not endowed with the vowel patterns that specify their categorial and phonological status.

supporting the idea of a progressive lexicalization of a major class of primary adjectival stems through time.

Moreover, Ancient Greek is intermediate between RV Sanskrit and Latin, as Hittite may also be, while Avestan patterns rather with Vedic. A large class of simple adjectives is found in Greek: βαρύς 'heavy', ἀγαθός 'good', μέλας 'black', etc., but the first adjective in the extant Greek literature is a participle: μῆνιν… οὐλομένην "[Achilleus'] destructive wrath" (A.1–2), and deverbal adjectives are frequent and are formed productively in the Homeric language (κλυτός 'famous', Skt. *śrutá-*). The idea of a typological change is therefore consistent not only with the situation in Sanskrit and Latin, but also in Ancient Greek, Avestan and probably Hittite.

In addition, it is well known that the history of the IE family has been characterized by a progressive decrease in the index of gross complexity (the average number of morphemes per word) over time. This decrease is mainly due to the blurring of intra-word morpheme boundaries, especially those between the root and word-formation suffixes (including the stem vowel). In turn, this blurring triggered the change from the root-based word-formation typical of early Vedic, Hittite and Homeric Greek to the stem- or word-based word-formation typical of Latin, Old High German, English and Persian.[55] The blurring affected all types of derived words in the same way: nouns, verbs and adjectives. And probably the lexicalization of derived nouns and of derived verb stems was more evident than the autonomous storage of derived adjectives, given the larger number of nouns and verbs in the lexicon. However, from the typological point of view, the lexicalization of adjectives represented the most important part of the change, since it caused the lexicalization of a third major class of lexemes and the subsequent change from the rigid to the specialized PoS system (that is, [N (AV)] → [N, A, V]).

As a consequence, lexemes that are intermediate between true verb stems, precategorial verbal lexemes (i.e., synchronic roots) and diachronic roots (see fn. 4) are common in the history of IE languages. Coming from the same antecedent (i.e., PIE **tep-*), the Latin root **tep-* in *tepeo* 'I warm', *tepor* 'heat' and *tepidus* 'warm' and the Sanskrit root *tap-* in *tápati* 'he warms, burns', *tápas-* 'heat', *tapú-* 'hot' may

---

55.   The evidence of this blurring is discussed by Belardi (1985, 1990, 1993) and Cipriano (1988) for derived nouns; by Cipriano (2001, 2007) and Di Giovine, Flamini & Pozza (2007) for derived verbs; by Alfieri (2016, 2018) for derived adjectives; by Belardi (2002a, 2002b) and Di Giovine (2001) for their general methodological consequences. See Cowgill (1963) on the decrease of the index of gross complexity and Panagl (1982, 1987, 2006) on the decrease in the productivity of nominalizations. On the change from root-based athematic morphology to the stem-, or later word-based, thematic morphology, especially in Germanic languages, see Kastovsky (1992, 1996).

easily seem to be the same type of unit. However, Skt. *tápati, tápas-* and *tapú-* are the result of productive word-formation rules, so *tap-* is a synchronic and precategorial unit. On the contrary, Lat. **tep-* is not the input form for productive word-formation rules in Latin. Only the stems *tepe-, tepid-,* and *tepor-* play such a role: *tepe-faci-o* 'I make hot', *tepid-arium* 'heated room in Roman baths' and *tepor-us* 'hot'. The form **tep-*, therefore, was a precategorial unit at some intermediate stage between PIE and Latin, but when Caesar crossed the Rubicon, say, it was already a semi-frozen lexeme or a diachronic unit (a diachronic root), just like the English 'root' [-sijv] in *de-ceive, con-ceive, per-ceive* quoted by Bloomfield (1933: 240), which English speakers can perceive but do not process to build new words.[56]

The IE family therefore attests a deep typological change in the PoS system, namely the change from type [N (AV)], which is still well preserved in the Rig-Veda, to type [N, A, V], which is found in Latin and in almost all the other, especially modern and Western IE, languages. The factual evidence of the change is well known to specialists in IE study, but it has never been interpreted in a coherent typological framework and many scholars, consciously or otherwise, still tend to equate the root with the simple verb stem, or with a diachronic unit, as if the change in PoS typology, as well as the difference between root- and stem-based word-formation, was the result only of a diachronic change, rather than being the result of both a typological and a diachronic change.[57]

7.2   Linguistic terminology

As clearly pointed out in the CC debate, labelling a category (that is, assigning a label to a distributional set of regularities) and identifying a category (that is, selecting a specific set of distributional regularities among the almost infinite sets of distributional regularities that can be defined across languages) are totally different things.[58] Both operations are arbitrary, though in a different sense. The CCs used in research trigger the identification of certain categories among the almost infinite categories that can be defined across languages: the categories identified, therefore, are indirectly arbitrary, since they logically descend from arbitrary CCs.

---

56.   Units such as Engl. -[sijv] are 'quasi-morphemes' (Aronoff 1976: 11ff.). They are less transparent and less productive than 'blocked morphemes' (e.g., *cran-* in *cranberry,* see Mel'čuk 1982). They therefore represent the boundary between synchrony and diachrony (or between word-formation and etymology), but are already cast on the diachronic side of the boundary.

57.   On the confusion between the diachronic and the typological value of the notion of 'root' in IE linguistics, see at least Belardi (1990, 1993, 2002: I, 256ff., 2008), Alfieri (2014c) and the references in Alfieri (2016: 133–6).

58.   See LaPolla (2016: 365ff.) and Croft (2016: 387ff.) on this point.

On the other hand, the labels that refer to the categories are totally arbitrary: the same category can be labelled *verb, X_1, Jack* or *Ciro* without any consequence for the descriptive adequacy of the theory, as long as the labels selected avoid ambiguity. Any discourse on terminology, therefore, has to start from categories and discuss category labels subsequently.

If PoS-concepts are excluded, 8 categories have been defined in the paper: 3 construction classes (*Noun, Adjective, Verb*); 4 lexeme classes (*noun, adjective, verb, root*); 1 default category referring to any simple lexical unit (*lexical morph(eme)* or *lexeme*). All these categories have been defined on the basis of what Haspelmath (this volume) terms *retro-definitions* that is, definitions that "assign a precise meaning to an existing term that does not have a widely recognized precise meaning yet". Indeed, they comply with the adequacy criterion proposed by Hasplemath: "an established term should not be defined in such a way that its definition is at variance with its traditional use. It should cover the core of the phenomena designated by the term (as generally understood), it should cover at least 80% of the cases where the term has been applied, and it should not include too many cases which would not be included traditionally". However, not all these categories comply with Haspelmath's criterion in the same way, so they will be analysed individually.

The categories of *noun, verb* and *adjective* defined in Latin, and Western grammar generally, are the product of a "lexicalist" approach, which neutralizes (or, more precisely, fails to distinguish) the difference between PoS-constructions and PoS-lexemes (or, more precisely, between stem classes and morpheme classes) in the classical, inflectional notion of *word*, which can be simple or derived with no difference.[59] In the paper, therefore, we have confined ourselves to dividing the two levels and identifying the prototypical function of the traditional notions of *noun, verb* and *adjective*, with little change in the canonical reference of these labels. The sum of the noun lexemes and the Noun constructions in Latin, therefore, covers most of the phenomena designated through the label of *noun* in traditional grammars, probably about 80%, as Haspelmath's criterion requires (only non-prototypical nouns meaning Actions and Quality being excluded from the comparative definitions of Noun/noun). And the same holds true for the traditional labels of *adjective* and *verb*.

---

59. Saying that the Latin-based PoS theory and, generally, all the linguistic theory up to the 19th century is strictly lexicalist is a commonplace (see the literature in Alfieri 2013b, 2014c): before Bopp the word was not only the most important unit of language, it was also its minimum unit, and was divided into the universal classes of nouns, verbs and adjectives.

Slightly more complex is the use of the label of *root*. We have already said (see fn. 4), that the term *root* can refer to both a synchronic and a diachronic unit. The synchronic use of the term *root* is traced to Indian native grammar and to the Western grammars of Sanskrit from the 17th century onwards, where it refers to Sanskrit simple verbal units, which – as we showed – are synchronic and precategorial. Moreover, the term *root* is common in Arabic and Hebrew grammar (Ar. *aṣl-* 'root, trunk') and in Western grammars of the Semitic languages from the 16th century onwards, where it refers to a simple verbal unit. However, Arabic grammars usually admit that the adjective is a class of derived stems (Wright 1979: 106): the adjectives *kabīr* 'big', *sādiq* 'true', *sa'b* 'hard', etc. are formed with the same vowel patterns found in derived nouns (*ra'īs* 'chief', *kātib* 'writer', *dars* 'studying') and the roots *k.b.r.* 'become big, grow up', *s.d.q.* 'tell the truth', and *s.'.b.* 'be hard' can be used verbally, just like the verbal roots *d.r.s.* 'to study' and *k.t.b.* 'to write', at least in the classical language (Ar. *huwa kabura* 'he grew old [3sɢ.ᴍ big<ᴘsᴛ>3sɢ.ᴍ]', lit. 'he became big', from *k.b.r.* '(to be) big').[60] Therefore Arabic and, more generally, Semitic roots can be considered precategorial, like Sanskrit roots, and Arabic – especially Classical Arabic – can easily fall into the same PoS type [N (AV)] as RV Sanskrit. In this case, we simply redefined the unit traditionally termed as *root* in Indian and Arabic grammar from a functional-typological point of view, without any change in its original reference. My label of *root* thus covers almost 100% of the phenomena where the term has been applied in Indian and Arabic native grammar, as well as in Indian and Arabic grammars written by western scholars.

Much more complex is the case of the label used to refer to the 'simple lexical unit'. Dictionaries of linguistics and manuals of morphology usually define the *root* as "that part of a word-form which remains when all inflectional […] and derivational […] affixes have been removed" (Bauer 2003: 340, following Bopp 1824: 126) or "the simplest possible form of a lexical morpheme" (Trask 1993: 224). On this basis, Haspelmath (2012), following Dixon (2004: 2), proposed the use of the term *root* to refer to 'any simple lexical unit', rather than the labels of *lexeme* or *lexical morpheme*, as we have done in this paper. On the one hand, Haspelmath's proposal is natural, since the ambiguity of the terms *morph(eme)* and *lexeme* are

---

60. See Stassen (1997: 158) on the verbal coding of the Arabic quality predicate; Jelinek & Deemers (1994: 710ff.) on the precategoriality of Semitic roots; the references in Alfieri (2016: 130, fn. 1) for the psycholinguistic reality of the root in Semitics; Troupeau (1984), Rousseau (1989) and Alfieri (2017) on the history of the notion of root in Semitic studies; Alfieri (2014c) on the history of this notion in Sanskrit philology. Harris (1946: 166) noticed that stems are units of lexical storage in Latin, though not in Sanskrit and in Arabic, where they are only secondarily derived units, bar the existence of a few primary noun stems.

well known (see fn. 2), speaking of *lexical roots* meaning 'lexical morph(eme)s' is common, and in many languages with rich morphology distinguishing the inflectional stem from the root (i.e., the stem without the stem forming affix) is useful, although the root so defined is not always a completely synchronic unit (see Section 7.1). On the other hand, such a proposal is highly ambiguous, since it does not comply, or complies only partly, with the general adequacy criterion for retro-definitions proposed by Haspelmath himself.

The standard definition of the term *root* in dictionaries is unquestionable in practice, because it refers to any lexical simple unit, and the theory of general linguistics so far has included such a unit. However, this use contrasts with the traditional use of the term *root* in Sanskrit and Arabic grammar, where only verbal items are defined as *roots*, simple nouns, pronouns, adjectives and particles being labelled with different terms. On the other hand, it is very ambiguous from a theoretical point of view, since it hides the difference between a diachronic unit reconstructed by the linguist, e.g., Engl. -[sijv] and Lat. *tep- (or Lat. *am- in *amāre*, which is the paramount example of the notion of *root* in dictionaries), and a synchronic unit used by speakers as an input form for productive word-formation rules, e.g., Lat. *tepē-* or Engl. *sing-*. Moreover, in Sanskrit and Arabic, roots are synchronic and precategorial, so they differ both from the Latin verbal units such as *tepē-*, which are monocategorial, and from the Proto-Latin or English roots such as *tep- or -[sijv] which are diachronic units. If the term *root* was used to refer to any simple lexical unit, in other words, it would hide the difference between a diachronic unit (Proto-Lat. *tep-, Engl. -[sijv]), a synchronic monocategorial unit (Lat. *tepē-*, Engl. *sing-*), and a synchronic precategorial unit (Skt. *tap-*, Ar. *d.r.s.*), triggering the unwarranted inference that Proto-Lat. *tep-, Engl. -[sijv], Lat. *tepē-*, Engl. *sing-*, Skt. *tap-* and Ar. *d.r.s.* represent the same type of unit, since they are referred to with the same label, which is clearly not the case.[61] Haspelmath's use of the label of *root*, therefore, assumes the dictionary definition of the term *root* as the "traditional use", but he does not realize that this traditional use is intrinsically biased by the same confusion between single language grammar, diachrony and typology on which Bopp's original definition of the notion of root was based, a confusion that can be avoided if (and only if) our CC of root is retro-defined starting from the "traditional use" of such term in Indian and Arabic grammar.[62]

---

61.   A part of this ambiguity was acknowledged by Mugdan (2015: 257), who noticed that many definitions of the term *root*, especially those cast in a Generative framework, are in fact definitions of what is commonly considered a *stem*.

62.   On the confusion between synchrony, diachrony and typology in the definition of the *root* in 19th century IE linguistics, see Alfieri (2013b, 2014c).

7.3   Further research prospects

The PoS theory proposed in Section 4 was applied only to two genetically related languages. Still, if the method employed to classify the lexeme class system in Latin and in Sanskrit is consistent, it can be applied to other cases as well. In fact, even if we confine ourselves to unmarked correlations (Noun, Adjective and Verb), at least five different types of lexical inventories can be defined consistently.

If these three constructions define three different classes of simple lexemes, the lexical structure is [N, A, V] and the simple lexemes typically associated with Noun, Adjective and Verb constructions are termed as *nouns, adjectives* and *verbs* respectively. This is the situation in Latin and English, but also in Amele, Bukip and Dyirbal.[63] On the other hand, if the three constructions projected on the map are prototypically filled with only two classes of lexemes, then the lexical structure can be [N (AV)] if a single class of lexemes enters both Verb and Adjective constructions, as in RV Sanskrit, Garo, Lao and Yimas, or [(NA) V] if a single class of lexemes enters both Noun and Adjective constructions, as in Quechua.[64] In the latter case, precategorial lexemes can be called *nominals* rather than nouns, while in the latter they are called *verbal roots* or verbal *types* rather than *verbs*, because in both cases these lexemes are distributionally different from standard nouns and verbs.[65] If the three environments Noun, Verb and Adjective define a single class of lexemes, which can enter each of these constructions without differences, then a single class of lexical items is found, as may be the case in Late Archaic Chinese, Riau Indonesian and Kharia.[66] In this case, the lexical structure is [(NAV)] and the single class of lexemes is termed as *contentive*. To these, also type [(NA) (AV)] may

---

63.   See Croft (2001: 88–9) on English and Alfieri (2014a) on the other cases.

64.   See Alfieri (2014a) on Garo, Lao and Yimas and Quechua. Floyd (2011) showed that nominals meaning qualities and nominals meaning actions are distributionally different in Quechua. But none of their differences concern the prototypical Noun and Adjective constructions, so Quechua nominals are interchangeable in these two environments. On this point, see also Haspelmath (2012: 116ff.).

65.   The term *type* was used by Broschart (1997) and Vogel (2000) with reference to the precategorial verbal units that are found in the analytic languages of South-east Asia, while *root* is limited to the Sanskrit and Semitic languages. However, from the functional point of view *types* are *roots* are the same type of precategorial unit, and the term root has often been used also with reference to the analytic and precategorial verbal morphemes in Chinese (e.g., Gabelentz 1811: 90).

66.   The first candidates for this type were Salish, Tongan and Mundari (see Hengeveld, Rijkhoff & Siewierska 2004 and Hengeveld & Rijkhoff 2005), but the monocategorial interpretation of these languages has been rejected in recent years (see the references in fn. 11 for Salish and Mundari and Völkel 2017 for Tongan). However, Late Archaic Chinese (Bisang

be added, if the ADJECTIVE slot is split between two constructions with almost the same frequency and the lexemes are divided in two classes, roots and nominals, as in Japanese and Wari'.[67] The lexical inventories discussed so far are summed up in the following table (Table 11):

Table 11. The major lexeme inventories across languages

Lexical inventories	Languages
[N, A, V]	English, Amele, Bukip, Dyirbal
[N (AV)]	Lao, Garo, Yimas
[(NA) V]	Quechua
[(NAV)]	Kharia, Late Archaic Chinese, Riau Indonesian
[(NA) (AV)]	Japanese, Wari'

At present, not all these inventories are equally agreed: types [N, A, V], [N (AV)] and [(NA) V] are often accepted in the literature on adjectival typology, while type [(NA) (AV)] is rarely found and type [(NAV)] is looked at suspiciously outside the Amsterdam Typology. On a future occasion, I will try to show that all the types are acceptable on the cross-linguistic level. However, at present we can confine ourselves to saying that if the PoS theory proposed in Section 4 is accepted, a similar typology of lexical inventories across languages is possible in principle.

In turn, if a similar typology is accepted, at least in its application in IE languages, a better understanding of the relation between DCs and CCs can be obtained. If the theory of LANGUAGE really aims to be general, it must have a space for all the units that are found in single languages. This does not mean that the categories in the two domains coincide. Quite the contrary. All the concepts required to descend from the study of the LANGUAGE to the study of groups of languages defined by some kind of coding similarities are CCs, rather than DCs. But CCs can refer to different levels of generality, and at the lowest level a CC is not very different from a DC shared by two or more languages. Therefore, any DC must have its own space in (the lowest level of) the general theory of language, although at this level it has a space not as an individual, with all the idiosyncratic features of individuals, but rather as a member of a group of units that function alike with respect to some abstract, functional feature arbitrarily identified by the researcher.

2008), Kharia (Peterson 2005, 2011, 2013) and Riau Indonesian (Gil 2000, 2013) are still good candidates.

67. On Wari', see Everett & Kern (1997); on Japanese, see Uehara (1998), Lombardi Vallauri (2000) and Backhouse (2004).

If this view is accepted, CCs are not unrelated to DCs, rather they may also act as a tool for clarifying aspects of single language grammar (e.g., the structure of Sanskrit lexeme class system, the secondary nature of its adjective class, the functional difference between the Sanskrit root and the Latin verb stem, the neutralization of simple and derived stems in Latin grammar, the different functions and different productivity of word-formation in Latin and in Sanskrit, etc.), for studying the dynamics of language change, which is necessarily a history of individuals (Saussure 1922: 134), and for arriving at a better understanding of contemporary linguistic terminology.

In sum, at present we do not know which functional features are needed to describe the DCs in single languages consistently, nor which type of CCs should be used in each case. We may presume that each linguistic phenomenon should be described through a different system of CCs in terms of number, type and level of generality, but an agreed method to operationalize such idea has yet to be found. However, if the interaction of PoS-concepts, PoS-constructions and PoS-lexemes is accepted as a consistent method to analyse the continuum that goes from LANGUAGE to individual languages in the field of PoS, the above PoS theory or, more precisely, the above theory of lexeme class division may play a pivotal role in the general theory of language (that is, in the empirical-typological version of UG defined in the Introduction), just it played a pivotal role in the foundations of single language grammar in Greek-Latin times.[68]

List of abbreviations

ABL	ablative	ADJ	adjectivalizer
ACC	accusative	AFF	affix
AG	ʀagreement	AOR	aorist

---

68. I prefer the label *General Grammar* (GG) to *Universal Grammar* (UG), so as to stress that the theory of LANGUAGE is general, though not universal, as it is based on arbitrary CCs and on historically determined empirical generalizations, which may change if the CCs used to detect them are changed or improved with the enlargement of the data. Moreover, the labels *PoS theory* and *theory of lexical inventories* are not equivalent. PoS theory is the sum of two theories, a theory of lexeme class division and a theory of the levels of language analysis at which constructions are fixed (Lazard 1999; Alfieri 2014a). Iroquoian languages have been said to lack nouns, but also to define the noun above the lexicon, thanks to the cooperation of various morphological and syntactic constructions (Mithun 2000). However, the morphology-syntax divide, which is the only theory of the levels of language analysis available at present, is inconsistent across languages (Haspelmath 2011, but already Saussure 1922: 186ff.), so also this aspect of PoS theory has not arrived at any consensus.

CC	comparative concept	OPT	optative
COMP	comparative	PERS	person
SBJ	subjunctive	(P)IE	(Proto)-Indo-European
GEN	genitive	PF	perfect
DC	descriptive categories	PL	plural
DAT	dative	POS	parts of speech
DU	dual	PRS	present
F	feminine	PRE	prefix
INS	instrumental	PRS	present
IPT	imperative	PST	past
LOC	locative	PTC	participle
M	masculine	RV	Rg-Veda
MD	middle	SG	singular
NM	nominalizer	INS	instrumental
NOM	nominative	SUP	superlative
NT	neuter	VOC	vocative

References

Alfieri, Luca. Forthcoming. The lexicalization of the adjective as an innovative feature in the Indo-European Family. *Poznań Studies in Contemporary Linguistics* 36(3): 379–412.

Alfieri, Luca. 2018. La definizione tipologica della radice e la teoria del segno lessicale. In *Linguistica filologia e storia culturale. In ricordo di Palmira Cipriano*, Alfieri Luca, Benvenuto Maria Carmela, Ciancaglini Claudia Angela, De Angelis Alessandro, Milizia Paolo & Pompeo Flavia (eds), 25–44. Roma: Il Calamo.

Alfieri, Luca. 2017. A proposito di Thomas Lindner, *200 Jahre Indogermanistik* (Salzburg-Wien 2016), ovvero la rilevanza della *Grammer* di Ravio (1648) per la storia della linguistica indoeuropea. *Archivio Glottologico Italiano* 102(2): 225–237.

Alfieri, Luca. 2016. The typological definition of the (apparently historical) notion of root. *Archivio Glottologico Italiano* 102(1): 129–169.

Alfieri, Luca. 2014a. Qualifying modifier encoding and adjectival typology. In Simone & Masini (eds), 119–139.

Alfieri, Luca. 2014b. The birth of a grammatical category: The case of the adjective class. *Studi e Saggi Linguistici* 52(1): 141–175.

Alfieri, Luca. 2014c. The arrival of the Indian notion of root into Western linguistics. *Rivista degli Studi Orientali* 87(2): 59–84.

Alfieri, Luca. 2013a. Review of: Ansaldo, Umberto, Don, Jan & Pfau, Roland (eds.). 2010. *Parts of Speech: Empirical and Theoretical Advances. Special issue of Studies in Language* 37(2): 425–434.

Alfieri, Luca. 2013b. L'ipotesi indoeuropea di Bopp e il problema del contatto tra grammatiche. In *Le lingue del Mediterraneo antico. Culture, mutamenti, contatti*, Mancini Marco & Lorenzetti Luca (eds), 15–35. Roma: Carocci.

Alfieri, Luca. 2009. La categoria dell'aggettivo in vedico. *Archivio Glottologico Italiano* 94(1): 3–41.

Anslado, Umberto & Don, Jan & Pfau, Roland (eds). 2010. *Parts of Speech. Empirical and Theoretical Advances. Special issue of Studies in Language* 32(3), 2008. See Alfieri (2013) for review.

Anward, Jan. 2000. A dynamic model of part-of-speech differentiation. In Vogel & Comrie (eds), 3–45.

Anward, Jan & Moravcsik, Edith & Stassen, Leon. 1997. Parts of speech: A challenge for typology. *Linguistic Typology* 1(2): 167–183. https://doi.org/10.1515/lity.1997.1.2.167

Aronoff, Mark. 2012. Morphological stems: What William of Ockham really said. *Word Structure* 5: 28–51. https://doi.org/10.3366/word.2012.0018

Aronoff, Mark. 1994. *Morphology by Itself. Stem and Inflectional Classes*. Cambridge MA: The MIT Press.

Aronoff, Mark. 1976. *Word-Formation in Generative Grammar*. Cambridge MA: The MIT Press.

Backhouse, Anthony E. 2004. Inflected and uninflected adjectives in Japanese. In Dixon & Aikhenvald (eds), 50–73.

Baker, Mark C. 2003. *Lexical Categories. Verbs, Nouns, and Adjectives*. Oxford: OUP. https://doi.org/10.1017/CBO9780511615047

Bauer, Laurie. 2003. *Introducing Linguistic Morphology*. Washington DC: Georgetown University Press.

Beck, David. 2016. Some language-particular terms are comparative concepts. *Linguistic Typology* 20(2): 995–402. https://doi.org/10.1515/lingty-2016-0013

Beck, David. 2013. Unidirectional flexibility and the noun-verb distinction in Lushootseed. In Rijkhoff & van Lier (eds), 185–220.

Beck, David. 2002. *The Typology of Parts of Speech Systems: The Markedness of Adjectives*. London: Routledge.

Belardi, Walter. 2008. Le "unità di lingua concrete", la parola e la frase. *Incontri Linguistici* 31: 11–39.

Belardi, Walter. 2002a. *L'etimologia nella storia della cultura occidentale*, 2 vols. Roma: Il Calamo.

Belardi, Walter. 2002b. *Il tema del segno lessicale nella diacronia linguistica*. Roma: Il Calamo.

Belardi, Walter. 1993. Sulla tipologia della struttura formale della parola nelle lingue indoeuropee. *Rendiconti dell'Accademia Nazionale dei Lincei, s. 9, v. 4, f. 4*: 525–570.

Belardi, Walter. 1990. Genealogia, tipologia, ricostruzione, leggi fonetiche. In *Linguistica, filologia e critica dell'espressione*, Walter Belardi (ed.), 155–218. Roma: Il Calamo.

Belardi, Walter. 1985. Considerazioni sulla ricostruzione dell'indoeuropeo. In *Tra linguistica storica e linguistica generale. Scritti in onore di Tristano Bolelli*, Ambrosini Roberto (ed.), 39–66. Pisa: Giardini.

Benedetti, Marina (ed.). 2001. *Fare etimologia. Passato, presente e future nella ricerca etimologica. Atti del convegno tenutosi presso l'Università per Stranieri di Siena, 2–3 ottobre 1998*. Roma: il Calamo.

Benveniste, Émile. 1962[3][1936[1]]. *Origines de la formation des noms en indo-européen*. Paris: Adrienne-Maisonneuve.

Bertram, Raymond & Schreuder, Robert & Baayen, Harald R. 2000. The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy and productivity. *Journal of Experimental Psychology: Learning, Memory and Cognition* 26(2): 489–511.

Bhat, Darbhe Narayana Shankara. 2000. Word classes and sentential function. In Vogel & Comrie (eds), 47–64.

Bhat, Darbhe Narayana Shankara. 1994. *The Adjectival Category. Criteria for Differentiation and Identification* [Studies in Language Companion Series 24]. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.24

Bisang, Walter. 2013. Word-classes. In *The Oxford Handbook of Linguistic Typology*, Song Jae Jung (ed.), 280–302. Oxford: OUP.

Bisang, Walter. 2008. Precategoriality and syntax based parts of speech: The case of Late Archaic Chinese. *Studies in Language* 32(3): 65–86. https://doi.org/10.1075/sl.32.3.05bis

Bloomfield, Leonard. 1933. *Language*. New York NY: Henry Holt.

Blevins, James P. 2016. *Word and Paradigm Morphology*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199593545.001.0001

Booij, Gert. 2007 [2015][1]. *The Grammar of Words. An Introduction to Linguistic Morphology*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199226245.001.0001

Booij, Gert, Lehman, Christian & Mugdan, Joachim (eds). 2000–2004. *Morphology. An International Handbook on Inflection and Word-formation*, Vol. 1 (2000); Vol. 2 (2004). Berlin: Mouton de Gryuter.

Bopp, Franz. 1827. *Ausführliches Lehrgebäude der Sanscrita-Sprache*. Berlin: Dümmler.

Bopp, Franz. 1824. Vergleichende Zergliederung des Sanskrits und der mit ihr verwandten Sprachen. *Abhandlungen der Königlichen Akademie der Wissenschaften zu Berlin, philosophische-historische Klasse* 1824: 117–148 (Repr. 1972. *Kleine Schriften zur vergleichende Sprachwissenschaft*, 1–33. Leipzig: Zentralantiquariat der deutschen demokratischen Republik).

Bossong, Georg. 1992. Reflections on the history of the study of universals. The case of the *partes orationis*. In *Meaning and Grammar. Cross-linguistic Perspectives*, Michel Kefer & Johan van der Auwera (eds). *Belgian Journal of Linguistics* 4: 27–51.

Bozzone, Chiara. 2016. The origin of the Caland system and the typology of adjectives. *Indo-European Linguistics* 4: 15–52. https://doi.org/10.1163/22125892-00401003

Broschart, Jürgen. 1997. Why Tongan does it differently: Categorial distinctions in a language without nouns and verbs. *Linguistic Typology* 1: 123–165. https://doi.org/10.1515/lity.1997.1.2.123

Brereton, Joel P. & Jamison, Stephanie W. 2014. *The Rig-Veda. The Earliest Religious Poetry of India*. Oxford: OUP.

Brugmann, Karl. 1904. *Kurze vergleichende Grammatik der indogermanischen Sprachen*. Straßburg: Trubner. https://doi.org/10.1515/9783111556857

Burrow, Thomas. 1955. *The Sanskrit Language*. London: Faber & Faber.

Cardona, George. 1997a[1988][1]. *Pāini. His Work and its Tradition*. Delhi: Motilal Banarsidass.

Cardona, George. 1997b[1976][1]. *Pāini. A Survey of Research*. Delhi: Motilal Banarsidass.

Chomsky, Noam. 1970. Remarks on nominalizations. In *Readings in English Transformational Grammar*, Roderick A. Jacobs & Peter S. Rosenbaum (eds), 184–221. Waltham MA: Ginn.

Cipriano, Palmira. 2007. Evoluzione tipologica e mutamento fonologico nell'area del persiano. *Rendiconti dell'Accademia Nazionale dei Lincei. Classe di Scienze Morali, Storiche e Filologiche s. 9, v. 18, f. 1*: 21–80.

Cipriano, Palmira. 2001. Il ruolo delle etimologie iraniche nello studio dell'indoeuropeo preistorico. In Benedetti (ed.), 107–120.

Cipriano, Palmira. 1988. Le implicazioni metodologiche e fattuali della teoria di W. Belardi sull'indoeuropeo. *Studi e Saggi Linguistici* 28: 101–126.

Corbett, Greville G. 2006. *Agreement*. Cambridge: CUP.

Coseriu, Eugenio. 2001. Les universaux linguistiques (et les autres). In *L'Homme et son langage*, Hiltraud Dupuy-Engelhardt, Jean-Pierre Durafour & Françoise Rastier (eds), 69–107. Louvain: Peeters.

Cowgill, Warren. 1963. A search for universals in Indo-European diachronic morphology. In *Universals of Language*, Joseph H. Greenberg (ed.), 114–141. Cambridge MA: The MIT Press.

Cristofaro, Sonia. 2009. Grammatical categories and relations: Language-universality vs. language-specificity and construction-specificity. *Language and Linguistic Compass* 3(1): 441–479. https://doi.org/10.1111/j.1749-818X.2008.00111.x

Croft, William. 2016. Comparative concepts and language-specific categories: Theory and practice. *Linguistic Typology* 20(2): 377–393. https://doi.org/10.1515/lingty-2016-0012

Croft, William. 2005. Word-classes, parts of speech and syntactic argumentation. *Linguistic Typology* 9: 431–441.

Croft, William. 2001. *Radical Construction Grammar*. Oxford: OUP.
https://doi.org/10.1093/acprof:oso/9780198299554.001.0001

Croft, William. 2000. Parts of speech as language universals and language-particular categories. In Vogel & Comrie (eds), 65–102.

Croft, William. 1991. *Syntactic Categories and Grammatical Relations*. Chicago IL: University of Chicago Press.

Croft, William & van Lier, Eva. 2012. Language universals without universal categories. *Theoretical Linguistics* 38(1): 57–72.

Crystal, David. 1967. English word classes. *Lingua* 17: 24–56.
https://doi.org/10.1016/0024-3841(66)90003-9

Dahl, Östen. 2016. Thoughts on language-specific and cross-linguistic entities. *Linguistic Typology* 20(2): 427–437. https://doi.org/10.1515/lingty-2016-0016

Dahl, Eynsten. 2010. *Time, Tense and Aspect in Early Vedic Grammar: Exploring Inflectional Semantics in the Rigveda*. Leiden: Brill. https://doi.org/10.1163/ej.9789004178144.i-475

Delbrück, Bertold. 1888. *Altindische Syntax*. Halle an der Saale: Waisenhause.

Demirdache, Hamida & Matthewson, Lisa. 1995. On the universality of syntactic categories. *Proceedings of the North-West Linguistic Society* 25: 70–93.

Di Giovine, Paolo. 2001. Etimologia indoeuropea ed etimologia romanza: Due metodi a confronto. In Benedetti (ed.), 285–293.

Di Giovine, Paolo & Flamini, Sara & Pozza, Marianna. 2007. Internal structure of verbal stems in the Germanic languages. In *Europe and the Mediterranean as a Linguistic Area. Convergencies from a Historical and Typological Perspective* [Studies in Language Companion Series 88], Paolo Ramat & Elisa Roma (eds), 49–62. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.88.05gio

Dixon, Robert M.W. 2004. Adjective classes in typological perspective. In Dixon & Aikhenvald (eds), 1–50.

Dixon, Robert M.W. 1977. Where have all adjectives gone? *Studies in Languages* 1: 19–77. (Repr. 1982. Berlin: Mouton De Gruyter).

Dixon, Robert M.W. & Aikhenvald, Alexandra Y. (eds). 2004. *Adjective Classes: A Cross-linguistic Typology*. Oxford: OUP.

Dressler, Wolfgang U. 1987. Word-formation as a part of natural morphology. In Dressler et al. (eds), 99–126.

Dressler, Wolfgang U. & Mayerthaler, Willi & Panagl, Oswald & Wurzel, Wolfgang U. (eds). 1987. *Leitmotivs in Natural Morphology* [Studies in Language Companion Series 10]. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.10

Dryer, Matthew S. 2016. Cross-linguistic categories, comparative concepts, and the Walman diminutive. *Linguistic Typology* 20(2): 305–331.
https://doi.org/10.1515/lingty-2016-0009

Dryer, Matthew S. 1997. Are grammatical relations universal? *Essays on Language Function and Language Type Dedicated to T. Givón*, Joan L. Bybee, John Haiman, SandraA. & Thompson (eds), 115–144. Amsterdam: John Benjamins. https://doi.org/10.1075/z.82.09dry

Evans, Nicholas. 2000. Word classes in the world's languages. In Booij et al. (eds), Vol. I, 708–731.

Everett, Daniel L. & Kern, Barbara. 1997. *Wari: The Pacaas Novos Language of Western Brazil*. London: Routledge.

EWAia = Mayrhofer, Manfred. 1986–1996. *Etymologisches Wörterbuch des Altindoarischen*. Heidelberg: Carl Winter.

Floyd, Simeon. 2011. Re-discovering the Quechua adjective. *Linguistic Typology* 15: 25–63.
https://doi.org/10.1515/lity.2011.003

von der Gabelentz, Georg. 1881. *Chinesische Grammatik*. Leipzig: T.O. Weigel.

Gil, David. 2016. Describing languoids: When incommensurability meets the language-dialect continuum. *Linguistic Typology* 20(2): 439–462.
https://doi.org/10.1515/lingty-2016-0017

Gil, David. 2013. Riau Indonesian: A language without nouns and verbs. In Rijkhoff & van Lier (eds), 89–130.

Gil, David. 2000. Syntactic categories, cross-linguistic variation and Universal Grammar. In Vogel & Comrie (eds), 173–216.

Grassmann, Hermann. 1976[5]. *Wörterbuch zum Rig-Veda*. Wiesbaden: Otto Harrassowitz.

Gren-Eklund, Gunilla. 1978. *A Study of Nominal Sentences in the Oldest Upaniad*. Uppsala: Almqvist & Wiksell International.

Gross, Maurice. 1979. On the failure of generative grammar. *Language* 55(4): 859–885.
https://doi.org/10.2307/412748

Haegeman, Liliane. 1994[2]. *Introduction to Government and Binding Theory*. Oxford: Blackwell.

Halle, Morris. 1973. Prolegomena to a theory of word formation. *Linguistic Inquiry* 4: 3–16.

Harris, Zelig. 1946. From morpheme to utterance. *Language* 22: 161–183.
https://doi.org/10.2307/410205

Haspelmath, Martin. Forthcoming. *The morph as a minimum linguistic form*. Draft, May 2019.

Haspelmath, Martin. 2021. Toward standardization of morphosyntactic terminology for general linguistics. (This volume) https://doi.org/10.1075/tsl.132.02has

Haspelmath, Martin. 2014[2003][1]. The geometry of grammatical meanings: Semantic maps and cross-linguistic comparison. In *The New Psychology of Language. Cognitive and Functional Approaches to Language Structure*, vol. 2, Michael Tomasello (ed.), 211–242. New York NY: Psychology Press.

Haspelmath, Martin. 2012. How to compare major word-classes across languages. In *Theories of Everything in honor of Edward Keenan* [UCLA Working Papers in Linguistics 17, Article 16], Thomas Graf, Denis Paperno, Anna Szabolcsi & Jos Tellings (eds), 109–130. Los Angeles CA: UCLA. <http://phonetics.linguistics.ucla.edu/wpl/issues/wpl17/wpl17.html> (11 November 2020).

Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1): 31–80. https://doi.org/10.1515/flin.2011.002

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86(3): 663–687. https://doi.org/10.1353/lan.2010.0021

Haspelmath, Martin. 2007. Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11: 119–132. https://doi.org/10.1515/LINGTY.2007.011

Haspelmath, Martin. 1996. Word-class-changing-inflection and morphological theory. In *Yearbook of Morphology 1995*, Geert Booij & Jaap van Marle (eds), 43–67. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-017-3716-6_3

Hengeveld, Kees. 2013. Parts-of-speech system as a basic typological determinant. In Rijkhoff & van Lier (eds), 31–55.

Hengeveld, Kees. 1992. *Non-verbal Predication: Theory, Typology, Diachrony*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110883282

Hengeveld, Kees & Rijkhoff, Jan & Siewierska, Anna. 2004. Parts-of-speech and word-order. *Journal of Linguistics* 40(3): 527–570. https://doi.org/10.1017/S0022226704002762

Hengeveld, Kees & Rijkhoff, Jan. 2005. Mundari as a flexible language. *Linguistic Typology* 9(3): 406–431.

Hopper, Paul J. & Thompson, Sandra A. 1984. The discourse basis for lexical categories in Universal Grammar. *Language* 60(4): 703–752. https://doi.org/10.1353/lan.1984.0020

Janda, Richard D. & Joseph, Brian D. 2003. On language, change and language change – Or of history, linguistics and historical linguistics. In *Handbook of Historical Linguistics*, Richard D. Janda & Brian D. Joseph (eds), 4–113. London: Blackwell. https://doi.org/10.1002/9780470756393.ch

Jelinek, Eloise & Deemers, Richard. 1994. Predicates and pronominal arguments in Strait Salish. *Language* 70(4): 697–737. https://doi.org/10.2307/416325

Ježek, Elisabetta & Ramat, Paolo. 2009. On parts-of-speech transcategorization. *Folia Linguistica* 43: 391–416. https://doi.org/10.1515/FLIN.2009.011

Joos, Martin (ed.). 1957. *Readings in Linguistics I*. Chicago IL: University of Chicago Press.

Joshi, Shivaram D. 1967. Adjectives and substantives as a single class in the parts of speech. *Journal of the University of Poona, Humanities Section* 25: 19–30.

Kastovsky, Dieter. 1996. Verbal derivation in English: A historical survey. Or much ado about nothing. In *English Historical Linguistics 1994* [Current Issues in Linguistic Theory 135], Derek Britton (ed.), 93–117. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.135.09kas

Kastovsky, Dieter. 1992. Typological reorientation as a result of level interaction: The case of English morphology. In *Diachrony within Synchrony: Language History and Cognition*, Günter Kellermann & Michael D. Morrissey (eds), 411–428. Frankfurt: Peter Lang.

Lander, Yury & Arkadiev, Peter. 2016. On the right of being a comparative concept. *Linguistic Typology* 20(2): 403–416. https://doi.org/10.1515/lingty-2016-0014

LaPolla, Randy J. 2016. On categorization: Stick to the facts of the languages. *Linguistic Typology* 20(2): 365–375. https://doi.org/10.1515/lingty-2016-0011

Launey, Michel. 1994. *Une grammaire omniprédicative: Essai sur la morphosyntaxe du nahuatl classique*. Paris: CNRS Editions.

Lazard, Gilbert. 2001a. On the grammaticalization of evidentiality. *Journal of Pragmatics* 33: 359–367. https://doi.org/10.1016/S0378-2166(00)00008-4

Lazard, Gilbert. 2001b. Transitivity revisited as an example of a more strict approach in typological research. *Folia linguistica* 36: 141–200.

Lazard, Gilbert. 1999. La question de la distinction entre nom et verbe. *Folia Linguistica* 33(2): 389–418.

Lazard, Gilbert. 1997. Pour une terminologie rigoureuse: Quelques principes et propositions. *Mémoires de la Société de Linguistique de Paris (nouvelle série)* 6: 111–133.

Lazard, Gilbert. 1992. Y a-t-il des catégories interlanguagières? In *Texte, Sätze, Wörter und Moneme. Festschrift für Klaus Heger zum 65. Geburtstag*, Susanne R. Anschüz (ed.), 427–434. Heidelberger Orientverlag.

Lazzeroni, Romano. 2013. Divagazioni sul comparativo indoeuropeo. Ἀλεξάνδρεια – *Alessandria Rivista di Glottologia* 6–7: 265–276.

Lazzeroni, Romano. 2008. Il vedico tra varianti e standardizzazione. In *Standard e non standard tra scelta e norma. Atti del XXX convegno della Società Italiana di Glottologia, Bergamo 20–22 ottobre 2005*, Molinelli Piera (ed.), 109–116. Roma: Il Calamo.

Lazzeroni, Romano. 2005. Fra mondo indiano e mondo mediterraneo: Categorie scalari e gradi di comparazione. *Archivio Glottologico Italiano* 95: 1–18.

Law, Vivien. 1993. Process of assimilation. European grammars of Sanskrit in the early decades of the nineteenth century. In *La linguistique entre mythe et histoire*, Daniel Droixhe & Chantal Grell (eds), 237–261. Münster: Nodus.

Lehmann, Christian. 2008. Roots, stems and word classes. *Studies in Language* 32(3): 546–567. (Reprint: Ansaldo et al. 2010, 43–64). https://doi.org/10.1075/sl.32.3.04leh

Lieber, Rocher. 2010. *Introducing Morphology*. Cambridge: CUP.

Lombardi Vallauri, Edoardo. 2000. Gli aggettivi giapponesi fra nome e verbo. *Studi Italiani di Linguistica Teorica e Applicata* 29(2): 311–345.

Lyons, John. 1977–1979. *Semantics*, Vol. 1 (1977), Vol. 2 (1979). Cambridge: CUP.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *Penn Working Papers in Linguistics 4.2. Proceedings of the 21st Annual Penn Linguistics Colloquium*, Alexis Dimitriadis, Laura Siegel, Clarissa Surek-Clark & Alexander Williams (eds), 201–225. Philadelphia PA: University of Pennsylvania.

Martinet, André. 1980[1960][1]. *Élements de linguistique générale*. Paris: Armand Colin.

MacDonell, Arthur A. 1975[1910][1]. *Vedic Grammar*. Delhi-Varanasi: Bhartiya Publishing House.

Matthews, Peter H. 1974. *Morphology. An Introduction to the Theory of Word-Structure*. Cambridge: CUP.

Matthews, Peter H. 1972. *Inflectional Morphology: A Theoretical Study Based on Aspects of Latin Verb Conjugation*. Cambridge: CUP.

Matthews, Peter H. 1967. Latin. *Lingua* 17:153–181. https://doi.org/10.1016/0024-3841(66)90008-8

Mayerthaler, Willi. 1981. *Morphologisches Natürlichkeit*. Wiesbaden: Akademische Verlagsgesellschaft Athenaion.

Meillet, Antoine. 1903[1934][7]. *Introduction à l'étude comparative des langues indo-européennes*. Paris: Hachette.

Mel'čuk, Igor. 1982. *Towards a Language of Linguistics. A System of Formal Notions for Theoretical Morphology*, Philip A. Luelsdorf (ed.). München: Fink.

Mithun, Marianne. 2000. Noun and verb in Iroquoian. In Vogel & Comrie (eds), 397–415.

Moravcsik, Edith A. 2016. On linguistic categories. *Linguistic Typology* 20(2): 417–425. https://doi.org/10.1515/lingty-2016-0015

Morgenroth, Wolfgang. 1977[2]. *Lehrbuch des Sanskrit. Grammatik – Lektionen – Glossar*. Leipzig: Max Hueber.

Mosel, Ulrike. 2017. Teop – An Oceanic language with multifunctional nouns, verbs and adjectives. *Studies in Language* 41(2): 255–293. https://doi.org/10.1075/sl.41.2.02mos

Mugdan, Joachim. 2015. Units of word-formation. In *Word Formation: An International Handbook of the Languages of Europe*, Vol. I, Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds), 235–300. Berlin: Mouton De Gruyter.

Mugdan, Joachim. 1986. Was ist eigentlich ein Morphem? *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 39(1): 29–43.

Panagl, Oswald. 2006. Zur verbale Konstruktion deverbativer Nomina. In *Word-Classes and Related Topics in Ancient Greek*, Emilio Crespo, Jesus de la Villa & Antonio R. Revuelta (eds), 47–57. Louvain la Neuve: Peeters.

Panagl, Oswald. 1987. Productivity and diachronic change in morphology. In Dressler et al. (eds), 127–152.

Panagl, Oswald. 1982. Produktivität in der Wortbildung. *Folia Linguistica* 16: 224–240.

Peterson, John. 2013. Parts of Speech in Kharia: A formal account. In Rijkhoff & van Lier (eds), 131–168.

Peterson, John. 2011. *A Grammar of Kharia*, 2 Vols. Leiden: Brill. https://doi.org/10.1163/ej.9789004187207.i-474

Peterson, John. 2005. There's a grain of truth in every "myth", or, Why the discussion on Mundari isn't quite over yet. *Linguistic Typology* 9: 351–390.

Plank, Franz. 1997. Word classes in typology: Recommended readings (a bibliography). *Linguistic Typology* 1(2): 185–192. https://doi.org/10.1515/lity.1997.1.2.185

Pokorny, Julius. 1959. *Indogermanisches etymologisches Wörterbuch*. Bern: Francke.

Pontillo, Tiziana & Candotti, Maria P. 2011. Discriminare tra aggettivo e sostantivo: Appunti sulla tradizione pāṇiniana. *Atti del Sodalizio Glottologico Milanese (n.s.)* 6: 66–84.

Radicchi, Anna. 1973–1974. Le parti del discorso nella tradizione grammaticale indiana. In *Materiali dell'Istituto di Glottologia dell'Università di Cagliari*, 1–64. Cagliari: Pubblicazioni dell'istituto di glottologia.

Ramat, Paolo. 2014. Categories, features and values in the definition of word-classes. *Studi e Saggi Linguistici* 52(2): 9–24.

Ramat, Paolo. 1999. Linguistic categorization and linguists' categories. *Linguistics* 37(1): 157–180. https://doi.org/10.1515/ling.1999.002

Renou, Louis. 1965. Remarques générales sur la phrase védique. In *Symbolae linguisticae in honorem Georgii Kuryłowicz*, Drewniak Stanisław (ed.), 230–234. Wrocław, Warszawa & Kraków: Zakład Narodowy Imienia Ossolińskich.

Renou, Louis. 1952. *Grammaire de la langue védique*. Lyon: Centre National de la Recherche Scientifique.

Rijkhoff, Jan. 2016. Crosslinguistic categories in morphosyntactic typology: Problems and prospects. *Linguistic Typology* 20(2): 333–363. https://doi.org/10.1515/lingty-2016-0010

Rijkhoff, Jan. 2007. Word classes. *Language and Linguistic Compass* 1(6): 709–726. https://doi.org/10.1111/j.1749-818X.2007.00030.x

Rijkhoff, Jan & van Lier, Eva. 2013. Flexible word classes in linguistic typology and grammatical theory. In Rijkhoff & van Lier (eds), 1–30.

Rijkhoff, Jan. 2013. (ed.). *Flexible Word-Classes. Typological Studies of Underspecified Parts of Speech*. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199668441.001.0001

Robins, Robert H. 1964. *General Linguistics. An Introductory Survey*. Bloomington IN: Indiana University Press.

Roth, Heinrich. 1660–1668. *Grammatica linguae sanscretanae bracmanum Indiae orientalis*. Facsimile edition Arnulf Camp & Jean Claude Muller. 1988. *The Sanskrit Grammar and Manuscripts of Father Heinrich Roth S.J. (1620–1668)*. Biblioteca Nazionale, Roma, Mss. Or. 171 e 172. Leiden: Brill.

Rousseau, Jean. 1984. La racine arabe et son traitement par les grammairiens européens (1505–1810). *Bulletin de la Société de Linguistique de Paris* 79(1): 285–321.

Sapir, Eduard. 1921. *Language. An Introduction to the Study of Speech*. New York NY: Harcourt, Brace and Company.

Sasse, Hans-Jürgen. 2001. Scales of nouniness and verbiness. In *Language Typology and Linguistic Universals: An International Handbook*, Vol. 2, Martin Haspelmath, Ekkehard König, Wulf Österreicher & Wolfgang Raible (eds), 495–509. Berlin: De Gruyter.

Sasse, Hans-Jürgen. 1993a. Das Nomen – Eine universale Kategorie? *Sprachtypologie und Universalienforschung* 46(3): 187–221.

Sasse, Hans-Jürgen. 1993b. Syntactic categories and sub-categories. In *Syntax: An International Handbook of Contemporary Research*, 2 Vols, Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (eds), 646–686. Berlin: De Gruyter.

de Saussure, Ferdinand. 1922[1916][1]. *Cours de linguistique générale*. Paris: Payot. [Italian translation: T. De Mauro (eds.). 1999. *Corso di linguistica generale*, Roma-Bari: Laterza].

de Saussure, Ferdinand. 1878. *Mémoire sur le système primitif des voyelles dans les langues indo-européennes*. Leipzig: Teubner.

Scalise, Sergio. 1984. *Generative Morphology*. Dordrecht: Foris. https://doi.org/10.1515/9783112328040

Schachter, Paul & Shopen, Timothy. 2007[2][1985][1]. Parts of speech systems. In *Linguistic Typology and Syntactic Description, Vol. 1: Clause Structure*, Timothy Shopen (ed.), 1–60. Cambridge: CUP. https://doi.org/10.1017/CBO9780511619427.001

Simone, Raffaele. 2007. Constructions and categories in verbal and signed languages. In *Verbal and Signed Languages. Comparing Structures, Constructs, and Methodologies*, Elena Pizzuto, Paola Pietrandrea & Raffaele Simone (eds), 198–252. Berlin: Mouton de Gruyter.

Simone, Raffaele & Masini, Francesca (eds). 2014. *Word Classes. Nature, Typology and Representations* [Current Issues in Linguistic Theory 332]. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.332

Speijer [alias Speyer], Jakob S. 1998[1886][1]. *Sanskrit Syntax*. Delhi: Motilal Banarsidass.

Speijer [alias Speyer], Jakob S. 1974[1896][1]. *Vedische und sanskrit Syntax*. Graz: Akademische Druck.

Stassen, Leon. 1997. *Intransitive Predication*. Oxford: OUP.

Thieme, Paul. 1967. Kṛṣṭí und carṣaṇí. *Zeitschrift für vergleichende Sprachforschung* 81: 233–244. (Reprint: Thieme Paul. 1971. *Kleine Schriften*, 2 Vols, G. Buddruss (ed.), Vol. 1, 247–258. Wiesbaden: Franz Steiner).

Thompson, Sandra A. 1988. A discourse approach to the cross-linguistic category 'adjective'. In *Explaining Language Universals*, John Hawkins (ed.), 167–185. Oxford: Blackwell.

Touratier, Christian. 2009. Questions surrounding the basic notions of word, *lexie*, morpheme and lexeme. In *Form and Function in Language Research. Papers in Honor of Christian Lehmann, Trends in Linguistics*. [Studies and Monographs 210], Johannes Helmbrecht (ed.), 157–166. Berlin: De Gruyter.

Trask, Robert L. 1993. *A Dictionary of Grammatical Terms in Linguistics*. London: Routledge.

Troupeau, Gérard. 1984. La notion de 'racine' chez les grammairiennes arabes ancien. In *History of Language Science. An International Handbook of the Evolution of the Study of Language*

from the Beginnings to the Present, Vol. I, Sylvain Auroux, E.F. Konrad Koerner, Hans-Josef Niederehe & Kees Versteegh (ed.), 239–246. Berlin: De Gruyter.

Uehara, Satoshi. 1998. *Syntactic Categories in Japanese: A Cognitive and Typological Introduction.* Tokyo: Kurosio.

VIA = Werba, Chlodwig H. 1997. *Verba Indoarica. Die primären und sekundären Wurzeln der Sanskrit Sprache, Pars I: Radices Primariae.* Wien: Verlag der Österreichischen.

van Lier, Eva (ed.). 2017 Lexical flexibility in Oceanic Languages. *Studies in Language* 41(2): 241–254 (monographic issue on *Lexical Flexibility in Oceanic Languages*). https://doi.org/10.1075/sl.41.2.01van

van Lier, Eva. 2009. *Parts of Speech and Dependent Clause.* Utrecht: LOT.

Vogel, Petra M. 2000. Grammaticalization and part of speech systems. In Vogel & Comrie (eds), 259–284.

Vogel, Petra M. & Comrie, Bernard. 2000. *Approaches to the Typology of Word Classes.* Berlin: Walter De Gruyter. https://doi.org/10.1515/9783110806120

Völkel, Svenja. 2017. Word classes and the scope of lexical flexibility in Tongan. *Studies in Language* 41(2): 445–495. https://doi.org/10.1075/sl.41.2.06vol

Wakernagel, Jakob. 1905. *Altindische Grammatik, Vol. II.1: Einleitung zur Wortlehre. Nominalkomposition.* Göttingen: Vandenhoek & Ruprecht.

Wetzer, Harrie. 1996. *The Typology of Adjectival Predication.* Berlin: De Gruyter. https://doi.org/10.1515/9783110813586

Whitney, William D. 2000[1979][1]. *Sanskrit Grammar.* Delhi: Motilal Banarsidass.

Wodtko, Dagmar S. & Irslinger, Britta & Schneider, Carolin (eds). 2008. *Nomina im Indogermanischen Lexikon.* Heidelberg: Winter.

Wright, William. 1979[1896][1]. *A Grammar of the Arabic Language*, translated from the German of Caspari and edited with numerous additions and corrections by W. Wright, LL.D. 3rd ed. revised by W. Robertson Smith & M.J. de Goeje. Cambridge: CUP.