Bhāṣā

Journal of South Asian Linguistics, Philology and Grammatical Traditions

Vol. 3 – Num. 1 April 2024

e-ISSN 2785-5953

Edizioni Ca'Foscari

e-ISSN 2785-5953

Bhasha

Journal of South Asian Linguistics,
Philology and Grammatical Traditions

Editor-in-chief
Andrea Drocco

Edizioni Ca' Foscari - Venice University Press
Fondazione Università Ca' Foscari
Dorsoduro 3246, 30123 Venezia
URL https://edizionicafoscari.unive.it/en/edizioni/riviste/magazen/

Bhasha

Journal of South Asian Linguistics, Philology and Grammatical Traditions

Semestral journal

Head office Ca' Foscari University of Venice | Department of Asian and North African Studies | Dorsoduro 3462, 30123 Venice, Italy | bhasha_journal@unive.it

Publisher Edizioni Ca' Foscari | Fondazione Università Ca' Foscari | Dorsoduro 3246, 30123 Venice, Italy | ecf@unive.it

Certificazione scientifica delle Opere pubblicate da Edizioni Ca' Foscari: tutti i saggi pubblicati hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima sotto la responsabilità del Comitato scientifico della rivista. La valutazione è stata condotta in aderenza ai criteri scientifici ed editoriali di Edizioni Ca' Foscari.
Scientific certification of the works published by Edizioni Ca' Foscari: all essays published in this issue have received a favourable opinion by subject-matter experts, through an anonymous peer review process under the responsibility of the Advisory Board of the journal. The evaluations were conducted in adherence to the scientific and editorial criteria established by Edizioni Ca' Foscari.

Bhasha
Vol. 3 – Num. 1 – April 2024
e-ISSN 2785-5953

Table of Contents

Passive and Causative in Sanskrit
John Lowe, Adriana Molina-Muñoz, Antonia Ruppel 5

Final Syllable Reduction in Middle Indic and Iranic
Andrew Ollett 53

The Sanskrit Paradigm of Tamil Grammar: Embrace and Resistance
E. Annamalai 107

'To Be Smeared' or 'To Be Attached'? An Investigation of Sanskrit *lipyate kāmaiḥ* and Pāli *lippati kāmesu* in Light of Their Chinese Translations
Francesco Barchi 123

URL https://edizionicafoscari.unive.it/it/edizioni/riviste/bhasha/2024/1/
DOI http://doi.org/10.30687/bhasha/2785-5953/2024/01

e-ISSN 2785-5953

Bhasha
Vol. 3 – Num. 1 – April 2024

Passive and Causative in Sanskrit

John Lowe
University of Oxford, UK

Adriana Molina-Muñoz
University of Oxford, UK

Antonia Ruppel
LMU München, Deutschland

Abstract To what extent can fine-grained statistical analysis provide evidence regarding syntactic patterns in corpus languages like Sanskrit, particularly in cases where the interaction of multiple syntactic phenomena obscures the evidence? We investigate the value of a correlation matrix for bivariate data analysis in relation to varying syntactic patterns in a relatively poorly attested yet productive construction in Classical Sanskrit: the passive of the causative ('passive causative'). The interaction of causative and passive is complex in Sanskrit, but we show that even in the case of low frequency data, syntactic conclusions can be drawn from such interactions when detailed statistical analysis is employed. In particular, our analysis speaks to the status of the ergative in Sanskrit.

Keywords Passive. Causative. Sanskrit. Ergative. Correlation matrix.

Summary 1 Introduction. – 2 Argument Structure Patterns in Active and Passive Causatives. – 2.1 The Active Causative. – 2.2 'Reduced' Active Constructions. – 2.3 The Passive Causative.– 2.4 *Ta*-Participle Causatives. – 3 Typological Comparisons and the Relation Between Active and Passive. – 4 Prior Research on Sanskrit. – 5 Restrictions on the Data. – 6 Data and Analysis. – 6.1 Predictions and Possibilities. – 6.2 Active Causative. – 6.3 Passive Causatives. – 6.4 *Ta*-Participle Causatives. – 7 Examining the correlations. – 7.1 The Active Correlations. – 7.2 The Passive and Active-Passive Correlations. – 7.3 The Participle Correlations. – 8 Discussion and Summary. – 8.1 Passive Causatives and the Ergative. – 8.2 A Note on Genre. – 8.3 Summary.

Edizioni Ca'Foscari

Peer review

Submitted 2023-07-05
Accepted 2023-09-23
Published 2024-03-01

Open access

Citation Lowe, J.; Molina-Muñoz, A.; Ruppel, A. (2024). "Passive and Causative in Sanskrit". *Bhasha*, 3(1), [1-48] 5-52.

DOI 10.30687/bhasha/2785-5953/2024/01/001

1 Introduction

Many of the basic facts about Sanskrit syntax have been known for more than a century; standard reference works are Delbrück (1888) and Speyer (1886; 1896). Yet it is only in recent years, with the advent of large numbers of digitised Sanskrit texts, that comprehensive, large-scale syntactic studies are finally becoming a possibility.[1] Sanskrit may be considered a corpus language, but its corpus is extensive, and if one is interested in major syntactic patterns, for example the relative order of major constituents in a sentence, one can now access millions of Sanskrit sentences at the touch of a button, each providing a relevant data token.

But there are many more oblique questions of Sanskrit syntax which are less easy to investigate. In this paper, we present the results of an investigation into the syntax of passive causatives (that is: passives of causatives) based on a large-scale corpus study.[2] While both the passive and the causative are highly productive categories in Sanskrit, the combination of passive and causative in the passive causative is rather less frequent (though still more common than in many languages). Moreover, due to a number of issues discussed below (§ 5), the usable data for passive causatives in Sanskrit is even more restricted than a cursory glance at the evidence might suggest. Nevertheless, we show that a bivariate correlation analysis reveals meaningful relationships in the syntax of Sanskrit passive causatives which would otherwise remain obscure.

In this paper we treat the syntax of 'Sanskrit' understood in a relatively broad sense, excluding the earliest Vedic Sanskrit, which is linguistically very different from later forms of the language, but including not only strictly Classical Sanskrit texts but also Epic and late Vedic Sanskrit, both of which are sufficiently similar to the Classical language to warrant treating them together. The Epics, the

We are grateful to Victor D'Avella, Yiming Shen, and for the comments and questions from the audiences at LFG19 (9 July 2019), Oxford's Graduate Indology Seminar (22 October 2019), the Cambridge Classics E Caucus Seminar (19 February 2020), the Oxford Graduate Philology Seminar (5 May 2020) and at WeCIEC 2019 (8 November 2019), and especially Hans Hock, Amba Kulkarni, Ian Roberts, and Rupert Thompson, among several others. This work was supported by the project 'Uncovering Sanskrit Syntax', funded as a Research Project Grant (RPG-2018-157) by the Leverhulme Trust.

1 Currently the most important online archives of digitised Sanskrit texts are GRETIL (`gretil.sub.uni-goettingen.de`), Titus (`titus.uni-frankfurt.de`), and the Digital Corpus of Sanskrit (`www.sanskrit-linguistics.org/dcs/`).

2 Our corpus comprises texts from a broad variety of genres and periods of Sanskrit, spanning a period of over 1,600 years, amounting to c. 5.5 million words. It includes c. 1.3 million words of Vedic prose, c. 1.7 million words of Epic and c. 2.5 million words of various genres of Classical (i.e. post-Pāṇinian) texts dating as late as the thirteenth century AD. Details are provided in the Appendix.

Mahābhārata and *Rāmāyaṇa*, are based on oral traditions whose origins predate Pāṇini but, in their final form, employ a language mostly following Pāṇinian rules. Strictly non-Classical features characteristic of the Epic language are sometimes later adopted in otherwise 'Classical' texts which are influenced in one way or another by the Epics (see Lowe 2017b, 288-9). The Classical Sanskrit idiom is based on a prescriptive application of the monumental grammar of Pāṇini, the *Aṣṭādhyāyī*, but the target language of this grammar was in fact the language of Vedic prose texts like the *Aitareya Brāhmaṇa* (Kulikov 2013b); thus late Vedic should not be kept entirely separate from the later classical language. Although our corpus included a number of Vedic prose texts, containing numerous instances of active causatives, the Vedic texts contributed no data on passive causatives, since passive causatives built to transitive roots are not attested before the end of the Vedic period, as noted by Kulikov (2012, 696-7).

In § 2 we present the phemonena under investigation, and address some of the complexities of Sanskrit syntax relevant to these phenomena. In § 3, we compare typological work on the causative and the relation between active and passive. In § 4, we compare previous research on causatives in Sanskrit. In § 5, we detail the restrictions on working with causative and passive causative data in Sanskrit, which results in a relatively small token count for analysis even in the case of a large corpus. In § 6, we present our quantitative data; in § 7, we present and discuss a fine-grained statistical analysis of this data. In § 8 we discuss the implications of our data for understanding the status of the ergative in Sanskrit, and draw conclusions.

2 Argument Structure Patterns in Active and Passive Causatives

We begin with some basic definitions. Causativisation is a process which takes as input a verbal predicate with a particular argument structure and returns a new version of the predicate with an augmented argument structure; specifically, causativisation adds a 'causer' argument, which becomes the semantically and grammatically most prominent argument of the predicate (surfacing as the active subject, for example). In contrast, the process of passivisation alters the argument structure of verbal predicates in almost the converse way: it demotes or suppresses the grammatically most prominent argument of a predicate (i.e. the active subject), resulting in the promotion of a less prominent argument (such as the active object), where present, to the position of greatest grammatical prominence.

For example, in the causative of an intransitive verb, a new argument, the causer, appears, filling the role of subject, while what was the subject of the non-causative becomes the object of the resulting causative:

(1) a. *sa* *khaḍga-abhihato* *'patat*
he.NOM sword-struck.NOM fall.IMPF.3SG
'Struck by the sword, he fell'. (*Rāmāyaṇa* 6.84.24)

b. *lāṅgūlena* *pradīptena* *rākṣasāṃs* *tān* *apātayat*
tail.INS blazing.INS rākṣasa.ACC.PL they.ACC.PL fall.CAUS.IMPF.3SG
'He struck down (lit. 'made fall') the rākṣasas with his blazing tail'.
(*Rāmāyaṇa* 5.51.9)

The causative applied to an intransitive base effectively creates a transitive verb, with nominative subject and accusative object; when the passive is applied to this, the subject is demoted/suppressed, and the object is (re-)promoted to subject:

(2) *tvat-kṛte* *śaṅkitair* *agnau*
you-caused afraid.INS.PL fire.LOC
munibhiḥ *pātyate* *haviḥ*
sage.INS.PL fall.CAUS.PASS.3SG oblation.NOM
'Since you have done this, the oblation is cast (lit. 'caused to fall') into the fire by the frightened sages'. (*Rāmāyaṇa* 3.29.12)

The basic possibilities for causativisation, passivisation, and their combination, in Sanskrit have been known for a long time; see the overviews in Speyer 1886, 32-8 and Renou 1961, 472-3. Detailed treatments of causativisation and especially its origins in the earliest attested stage of Sanskrit, Vedic, can be found, for example, in Cardona 1978; Hock 1981; Jamison 1983; Tichy 1980; 1993; Kulikov 2013a.

When formed to intransitive bases, there is only one pattern of causativisation, and one pattern of passivisation of the causative, as illustrated in (1) and (2) above. In the case of transitive bases, however, Sanskrit permits two competing realisations of the causative in terms of the resulting argument structure, and likewise in the passive causative two competing argument structure realisations. It is these points of argument structure variation which we are interested in this study, and so henceforth we do not consider intransitive bases further.

In terms of the morphosyntactic categories involved, we draw a primary three-way distinction between finite present-stem active causatives, causative *ta*-participles, and finite present-stem passive causatives.[3] These are the three most important and productive

3 We use 'active' here to contrast with the passive, discussed below. This also includes what is usually referred to as the 'middle' voice, which in Classical Sanskrit is more like

morphosyntactic categories of the Sanskrit verb system; we exclude from consideration causatives formed to aorist and perfect stems, since they lack morphologically distinct passive forms, and are in any case extremely rare in our corpus.

The *ta*-participle, more commonly labelled the 'past/perfect (passive) participle' will be discussed in detail below. It is the most important of the morphologically 'non-finite' verbal categories; it is mostly used as a main clause predicate, and when used as such is the most basic and common means of expressing past tense. The *ta*-participle displays an ergative-absolutive agreement pattern, in contrast to the finite verbal categories which are exclusively nominative-accusative in alignment.

2.1 The Active Causative

Crosslinguistically, when a transitive verb is causativised, the resulting argument structure may have one of two basic forms: while the object of the transitive predicate retains its morphosyntactic objecthood, in one form or another, the subject of the original transitive verb may surface in the causative either as a direct object or as an oblique or indirect object (Baker 1988, 161-7). As discussed further below, in some languages only one of the two patterns is possible, while in others both patterns are found; in the latter case, the distribution of the two patterns may be subject to lexical, semantic and/or pragmatic restrictions. In Sanskrit, both patterns are possible for all or most verbs. For example, in (3b) the subject of the base predicate in (3a) appears as the object in the causative (marked with accusative case), with the original object of the base predicate also marked in the accusative (retaining this from the noncausative, and representing a secondary object function). We call this the 'accusative-accusative' (ACC-ACC) type.

(3) a. *ahaṃ* *setuṃ* *kariṣyāmi*
I.NOM bridge.ACC make.FUT.3SG
'I will make a bridge'. (*Rāmāyaṇa* 6.15.11)

b. *nalaṃ* *setum* *akārayat*
N.ACC bridge.ACC make.CAUS.IMPF.3SG
'He had Nala make a bridge'. (*Rāmāyaṇa* 6.114.41)

Alternatively, the original subject of the non-causative may surface as an oblique argument in the causative, usually marked with

a morphologically deponent active than a functionally distinct voice. Both take the same pattern of stem formation; they differ only in the forms of the person/number endings.

instrumental case, while the original object of the base predicate remains the sole (accusative-marked) object of the causative (4).[4] We call this the 'oblique-accusative' (OBL-ACC) type.

(4) *geyaṃ* *sa* *dhīmān* *vipra-varair* *akārayat*
song.ACC this.NOM wise.NOM brahmin-best.INS.PL do.CAUS.IMPF.3SG
'This wise one made the best brahmins perform (lit. make) a song'.
(*Vāmana Purāṇa* 68.59)

Note that the causative verb form is the same in both constructions. In the rest of this paper, we use the terms *embedded subject* and *embedded object* to refer to the 'original' subject and object of the base predicate when they appear in the causative. Our *embedded subject* corresponds to what is usually called the 'causee', but we adopt the term originally used by Comrie (1976) in order to have a clear and parallel means of referring to both the 'original' subject and the 'original' object in the causative.

2.2 'Reduced' Active Constructions

In Sanskrit all arguments are in principle omissible, and more often than not at least one of the non-subject arguments of a causative are omitted. Omitted non-subject arguments may be contextually recoverable, or may be non-specific. Argument omission somewhat obscures the line between ACC-ACC and OBL-ACC causatives when, as is frequently the case, the embedded subject ('causee') is omitted:

(5) *prātar* *utthāya* *tat* *sarvaṃ* *kārayāmi* *karomi* *ca*
early.ADV rise.ABS this.ACC all.ACC do.CAUS.1SG do.1SG and
'Having risen early I have (someone) do and (myself) do all this'.
(*Mahābhārata* 13.124.15)

4 The instrumental case marking is primarily semantic, marking agency, rather than syntactically determined. With experiencer verbs like *jñā* 'know' and *śru* 'hear', the expected semantic case, dative/genitive, almost always occurs in place of the instrumental, but some examples occur where the instrumental is used in place of the semantically more appropriate case, evidencing a degree of syntactic standardisation. Whether instrumental or dative/genitive, the grammatical role of the argument is the same: it is an oblique; there is no evidence for distinguishing a separate role of 'indirect object' in Sanskrit. Moreover the variation between dative/genitive and instrumental does not affect our statistical analysis below, so we draw no distinction between these different case markings of the embedded subject.

We refer to this type as O-ACC. Alsina (1992, 519) observes a crosslinguistic tendency to the effect that object-marked embedded subjects cannot be omitted, while embedded subjects with oblique marking can be omitted, a tendency which makes sense given that oblique arguments are in general more omissible than core arguments (terms). But in Sanskrit, as mentioned, all arguments are in principle omissible, and omission of core object arguments, whether of causative or non-causative verbs, is widespread. Thus while it is likely that O-ACC more frequently, or more naturally, represents a reduced form of OBL-ACC than of ACC-ACC, it cannot be ruled out, indeed it is likely, that at least some instances of O-ACC represent reduced forms of ACC-ACC.

A third possibility for O-ACC is that it represents a separate constructional type which expresses a (likely indirect) causal sense but without increasing the valency of the verb; this would correspond to translating *kārayāmi* in (5) above as e.g. 'I have all this done'. As seen in the data below, O-ACC is extremely frequent.[5] The precise status of O-ACC causatives in relation to the other causative types remains to be established; our statistical analysis below provides evidence in relation to this question.

It is alternatively possible to omit the embedded object argument, but retain the embedded subject. Both ACC-O and, more rarely, OBL-O structures are found:

(6) a. *mānuṣā* *mānuṣān* [...] *kārayanti* *divāniśam*
man.NOM.PL man.ACC.PL [...] do.CAUS.3PL by.day-by.night
'Men… cause men to work day and night'. (*Mahābhārata* 12.254.39)

b. *śrāvayec* *chraddadhānānāṃ* *tīrthapāda-pada-āśrayaḥ*
hear.CAUS.3SG faithful.GEN.PL Kṛṣṇa-foot-resorting.NOM
'One who resorts to the feet of Kṛṣṇa should make the faithful hear (the story of Dhruva, i.e. by reciting it)'. (*Bhāgavata Purāṇa* 4.12.50)

Granted that ACC-O and OBL-O represent reduced forms of a fuller construction, with the embedded object omitted, naturally ACC-O must be a reduced form of ACC-ACC and OBL-O a reduced form of OBL-ACC. In our data there are 180 instances of OBL-ACC and 17 instances of OBL-O, meaning that the embedded object is omitted in almost exactly 10% of instances of (what is or would be) OBL-ACC. In contrast, ACC-O makes

5 This causee-less construction became so prevalent that, in the early Middle Indo-Aryan languages, the equivalent of *kārayati* was increasingly interpreted as a simple transitive (Edgerton 1946); the ultimate fate of the *-aya-* causative is as a transitive marker in modern Indo-Aryan languages. The causative in many modern Indo-Aryan languages derives directly from the redetermined causative suffix in *-āp-aya-*, which develops as a separate formation in early Middle Indo-Aryan.

up more than 30% of instances of (what is or would be) ACC-ACC: 41 instances beside 89 instances of (unreduced) ACC-ACC. This suggests a difference in the syntactic status of the embedded object in OBL-ACC and ACC-ACC constructions. For example, it would support an analysis whereby the embedded object in the OBL-ACC causative is the core object of the causative predicate, but in the ACC-ACC causative it is the embedded subject which is the core object, while the embedded object is a secondary/indirect object or oblique argument.

Finally, it is also possible for both non-subject arguments to be omitted; we refer to this type as the 0-0 construction:

(7)	*(viśrāma-icchāṃ*	*karoty*	*atra)*	*kārayanti*
	sleep-wish.ACC	do.3SG	here	do.CAUS.3PL
	na	*te*	*bhaṭāḥ*	
	not	this.NOM.PL	servant.NOM.PL	

'(He wants to rest here but) these servants do not allow (him) to do so'. (*Garuḍa Purāṇa* 2.5.98)

In this example, both embedded subject and embedded object of the causative are directly inferable from the previous clause, but it is also possible for one or both null positions to represent indefinite null arguments. The 0-0 type could in principle be treated as a reduced form of any of the types already introduced.

2.3 The Passive Causative

As with the active (§ 2.1), there are two types of passive causative. In one, the subject of the passivised causative verb is the embedded subject of the active causative, that is the original subject of the non-causative.[6] The embedded object remains in the accusative:

(8)	*candra-āsannair*	*hi*	*nakṣatrair*
	moon-in.conjunction.INS.PL	indeed	star.INS.PL
	lokaḥ	*kāryāṇi*	*kāryate*
	world.NOM	duty.ACC.PL	do.CAUS.PASS.3SG

'People are (lit. the world is) caused to do their duties by the constellations in conjunction with the moon'. (*Bṛhatkathāślokasaṃgraha* 15.6)

6 In using the term 'subject' in relation to passivisation here we refer fundamentally to 'grammatical' subjecthood, understood first and foremost in terms of nominative case and verbal agreement. But as we discuss in § 2.4, the instrumental agent in the passive does show some syntactic subject properties.

We refer to this type as PC-S: 'passive causative (with promotion of the embedded) subject'. Here, in the basic non-causative sentence underlying (8), *loka* 'people/world' would be the subject, and *kārya* 'duty' the object. These would then be the embedded subject and object, respectively, in the causative, with the causative subject being *nakṣatra* 'constellation'. In this PC-S passive, it is the embedded subject of the causative, here *loka*, which becomes the subject of the passive causative.

In the second type, it is the embedded object which becomes the subject in the passive of the causative:

(9)	*vegavatyā*	*tataḥ*	*saha*
	V.INS	then	with
	naravāhanadattasya	*vivāhaḥ*	*kāryatām*
	N.GEN	marriage.NOM	do.CAUS.PASS.IMP.3SG

'Then let the marriage of Naravāhanadatta with Vegavatī be caused to be carried out'. (*Bṛhatkathāślokasaṃgraha* 15.13)

We refer to this type as PC-O: 'passive causative (with promotion of the embedded) object'. While the embedded object is promoted to subject, the embedded subject, if expressed, appears in the instrumental. But since this is the passive of a causative, it is also possible for the instrumental to express the demoted subject of the active causative, i.e. the causer. Examples of the PC-O passive causative in which both instrumentals are overtly expressed are extremely rare (Hock 1981, 26) and are not found in our corpus.

2.4 *Ta*-Participle Causatives

Turning now to the *ta*-participle causatives, here again we find two competing argument structure possibilities, just as with the finite active and finite passive.

(10)	*bhrāṭṛ-bhrāṭṛvya-bāndhavaiḥ*	*kāritaḥ*	*kṣetra-karma-ādi*
	brother-nephew-relative.INS.PL	do.CAUS.TA-PTC.NOM.M	field-work-etc.ACC.N

'(He) was made to do fieldwork etc. by his brothers, nephews and other kinsmen'. (*Nārada Purāṇa* 1.48.42)

As an ergative construction, (10) shows the participle predicate agreeing with the object argument (O), here the doer of the work, which would be accusative in the non-ergative active, while the transitive subject/agent argument (A), which would be nominative and controlling verbal agreement in the non-ergative active, appears in

the instrumental. This therefore corresponds to the ACC-ACC finite causative. At the same time, although the *ta*-participle is not a passive formation, and is expected to pattern syntactically as an active, the case-marking and agreement patterns seen in (10) are superficially the same as those of the PC-S finite passive. Rather than treat *ta*-participle examples such as (10) as mere variants of the active ACC-ACC construction, we label and consider them separately: we refer to the construction in (10) as NOM-ACC.

(11) *vivāhaḥ* *kārito* *mayā*
marriage.NOM do.CAUS.PASS.TA-PTC.NOM I.INS
'I had the marriage carried out (lit. 'the marriage was caused to be done by me')'. (*Bṛhatkathāślokasaṃgraha* 14.118)

The formation in (11) likewise shows the O argument, here the marriage, agreeing with the verb and standing in nominative case, while the A argument, here the first-person pronoun, appears in the instrumental. As an 'active' ergative formation, this corresponds directly to the OBL-ACC finite causative. At the same time, again, it shows the same superficial case-marking and agreement patterns as the PC-O finite passive. We refer to the construction in (11) as OBL-NOM.[7]

7 The instrumental in (11) is most naturally interpreted as the causer, and hence the A argument of the causative, but if the embedded subject were to be expressed it would likewise appear in the instrumental (just as in the OBL-ACC construction), and there could conceivably be a context in which (11) could be read in this way (i.e. 'someone had me carry out the wedding'). As noted above for the finite passive, with the *ta*-participle also both instrumental arguments are hardly ever expressed at the same time, and never in our corpus. Given that the instrumental A argument with the *ta*-participle is generally taken as a subject (i.e. this is an active construction merely with ergative morphosyntax), while the instrumental agent in the finite passive (or causer, in the passive causative) is not a subject but an oblique/adjunct, we might expect a noticeable difference in the frequency of occurrence of this instrumental argument between the two categories: it should be considerably more omissible in the case of the finite passive. Butt and Deo (2017, 651) refer to Gonda (1951, 22) in claiming that in Sanskrit the instrumental agent is hardly ever expressed in the finite passive, but rarely omitted with the *ta*-participle. In fact Gonda (1951, 22) makes a claim only about finite passives, and only in relation to a small corpus study on the Vedic Sanskrit of the *Śatapathabrāhmaṇa*. In our data for passive and *ta*-participle causatives, we find the following:

	explicit causer	unexpressed
finite	5	84
ta-ptc	68	437

Only 5.6% (5/84) of finite passive causatives have an explicit causer, whereas 13.4% (68/437) of *ta*-participles do. This difference is statistically significant (Fisher's exact test: p = 0.03598). This therefore supports the assumption that the instrumental agent is more freely omissible in the case of the finite passive than of the *ta*-participle. Even in the *ta*-participle, though, omission of A is by far the most regular situation. These observations also bear some direct relevance to the expression of the embedded subject.

3 Typological Comparisons and the Relation Between Active and Passive

Our concern in this paper is an empirical investigation of the relations between the different active, passive and ergative causative structures in Sanskrit, and what this tells us about the underlying syntax of the constructions involved. We are aware of no comparable empirical or corpus-based studies of the relation between active and passive causatives in other languages. This may be partly because passives of causatives are not particularly common constructions in languages that admit them, but also because few languages freely admit both types of causative (i.e. constructions parallel to both (3b) and (4), which, as noted above, correspond to the two main argument structure patterns for causatives crosslinguistically). According to Alsina (1992), Bantu languages like Chichewa and Kinyarwanda freely admit both types of causative, but according to Baker (1988, 161-7), the two causative structures in Chichewa correspond to two distinct dialects, and in Kinyarwanda only the equivalent of the ACC-ACC causative is possible. Turkish appears to show both patterns, but Çetinoğlu and Butt (2008) show that in fact Turkish has only one type of causative to transitive verbs in the strictest sense (the equivalent of the Sanskrit OBL-ACC causative). Tamil permits both types of passive causative (K. Sarveswaran, p.c.), and at least in some varieties permits both types of active (Davies, Rosen 1988, 78), thus coming close to the Sanskrit situation, but to our knowledge the Tamil facts have never been investigated in detail. In other languages, both patterns are found but with different sets of verbs. For example, in Marathi and some other modern Indo-Aryan languages, most verbs take the equivalent of the OBL-ACC causative, but a semantically identifiable subset of verbs, e.g. ingestive verbs, take the equivalent of ACC-ACC (Alsina, Joshi 1991). We are aware of no detailed empirical or corpus-based studies of active vs. passive alternations in the causative in these or any other language.

An early attempt to explain the alternation between the different types of active causative is by Comrie (1976). Comrie relies on the

Comparing only PC-O and OBL-NOM, where both causer and causee can be expressed in the instrumental, there is a difference in the frequency of expression of the causee: with finite passive causatives, 34.6% (18/52) of examples have explicit causees (while none have explicit causers); with the *ta*-participle, 24.5% (67/273) have explicit causees. The difference is not significant (Fisher's exact test: $p = 0.1674$), yet it still appears relevant that most of the proportional difference between the two categories can be attributed to the expression of the causer: around 30% of the *ta*-participles do have instrumentals (so considerably closer to the 34.6% of finite passive causatives with instrumentals), 15 of them being causers. So the lower number of expressed causees with the *ta*-participle may be related to the expression of the causer: expressing the causer blocks the expression of the causee, so we find causees more frequently expressed in the finite passive causative.

Noun Phrase Accessibility Hierarchy (Keenan, Comrie 1977), which ranks the arguments of a predicate as follows:

(12) Subject > Direct Object > Indirect Object > Oblique argument

Comrie argues that when a causative is formed and a new subject argument, the causer, is introduced, the original subject of the non-causative, i.e. the embedded subject, is demoted to the highest available position on the argument hierarchy. If the direct object position is not already filled (as with intransitive verbs), or if in a particular language the direct object position permits doubling, then the demoted subject becomes a direct object; if the direct object position is filled and does not permit doubling, but the indirect object position is available, the demoted subject becomes an indirect object, and so on. Such an explanation neatly accounts for languages which show fixed patterns, e.g. where the embedded subject surfaces as a direct object in the causative of intransitives, but as an indirect object in the causative of transitives. But it cannot directly account for the syntactically unconstrained alternation between the two types with transitive verbs in Sanskrit, as introduced above.

An alternative to the syntax-oriented approach of Comrie (1976) is the semantically oriented account of Cole (1983). Cole argues that the varying role of the embedded subject in causative constructions can be fully explained by semantic factors, even in languages where originally semantic alternations have been fixed according to syntactic factors. Cole argues that when the embedded subject retains agency in the causative, it is expressed with agentive marking; in the case of Sanskrit, this corresponds to the instrumental marking of the OBL-ACC causative. When the embedded subject is non-agentive, it is expressed with appropriate marking, such as patient/object marking; this corresponds to the Sanskrit ACC-ACC causative. As discussed below and as reflected in previous approaches to the Sanskrit data, semantic factors clearly play an important role in the alternation between OBL-ACC and ACC-ACC in Sanskrit, but this does not preclude the importance of syntactic factors as well.

The earliest theoretical analysis of the Sanskrit causative is that by Pāṇini in his *Aṣṭādhyāyī*. According to Pāṇini, the distribution of ACC-ACC and OBL-ACC in Sanskrit is similar to modern Indic languages like Marathi: causatives of intransitive verbs are necessarily ACC-ACC; the default structure for causatives of transitive verbs is OBL-ACC, but a semantically definable subset of transitives (verbs of motion, perception, consumption and sounding) take ACC-ACC.[8] Two verbs, *kṛ* 'do,

8 Note we are presenting Pāṇini's analysis here, and not our own. Pāṇini includes verbs of motion, which can construe with an accusative representing the goal of motion,

make' and *hṛ* 'take', are specified as taking either.[9] As we will see below, the facts in Sanskrit texts are more complicated, but we provide corpus-based evidence which at least partly supports the semantic categorisation proposed by Pāṇini.

The indigenous grammatical tradition represented by Pāṇini's *Aṣṭādhyāyī* does not explicitly treat the passive causative, presumably because it did not feel the need: the two possibilities for the passive of the causative fall out unproblematically from the ordinary interaction of the rules for the causative and the rules for the passive.[10] What this ordinary interaction implies is that the PC-S passive causative (8) is specifically the passive of the ACC-ACC active causative (3b), while the PC-O passive causative (9) is the passive of the OBL-ACC causative (4).

On an abstract level this seems intuitively reasonable. If the passive necessarily involves the promotion of the core object argument to subject, then the PC-O passive must correspond to the OBL-ACC active causative, since in the latter the embedded object is the core object, and in the former the embedded object is the subject. Similarly, if we assume that the embedded subject is the sole core object in the ACC-ACC active causative, then this should correspond only to a passive of the PC-S type.[11] A similar prediction can be derived from Comrie's (1976) account of the active causative.

This correspondence, i.e. PC-S as passive of the ACC-ACC causative, and PC-O as the passive of the OBL-ACC causative, seems intuitively reasonable, and is taken for granted by e.g. Hock (1981). Yet it need not necessarily be the case. Certain languages which show only the equivalent of the ACC-ACC active causative show both types of passive causative, PC-S and PC-O, showing that it is at least possible for a PC-O passive causative to function as passive to an ACC-ACC causative. This is the case in Setswana (Rigardt Pretorius and Ansu Berg, p.c.) and is also the pattern described for Kinyarwanda by Baker (1988, 174-80).[12] In our data, we find a variety of patterns, some of which do not

as transitives, for reasons internal to his system.

9 The ACC-ACC type and the free choice with *kṛ* and *hṛ* are specified in Pāṇini's *Aṣṭādhyāyī* 1.4.52-3. The OBL-ACC type results from more general rules.

10 Essentially, in Pāṇini's system there is usually a free choice between active and passive for the main verb when deriving a clause. The passive takes as its nominative argument (i.e. in modern terms its grammatical subject) a particular argument role labelled the *karman*, and the *karman* of any causative verb is clearly defined by the rules for the causative itself.

11 Recall the greater frequency of ACC-O over OBL-O, discussed above, which supports treating the second accusative of ACC-ACC as not being a core object argument.

12 These Bantu languages are symmetrical object languages, so these passivisation possibilities are parallel to the alternations these languages show with non-causative ditransitive verbs (cf. Bresnan, Moshi 1990).

appear to support this correspondence. For example, with the verb *jñā* 'know', the active causative is predominantly ACC-ACC (five of six examples, or 20 of 22, if we include ACC-O and OBL-O), whereas the passive shows a preference for PC-O (14 of 23 examples).

A rather different approach is taken by Kiparsky and Staal (1969). They argue, in an early generative treatment, that the OBL-ACC causative results from first applying passivisation to the base, and then applying the causative, while ACC-ACC results from applying the causative to the non-passivised base. A 'passive first' analysis of causative constructions in which the embedded subject is marked with the same oblique case as passive agents is also considered favourably by Comrie (1976), but argued against by Cole (1983). In the present context, the relevance of such a proposal is that it could not easily be integrated with an approach which associates one of the passive causative structures with the OBL-ACC active causative. If in the OBL-ACC causative the passive has already applied, we should not be able to apply it again (double passives are not possible in Sanskrit). Thus the two passive causative constructions, PC-S and PC-O, could only be both passives of the ACC-ACC causative, as is apparently the case in the Bantu languages mentioned in the previous paragraph.

Despite the support of Comrie (1976), there are typological considerations against the proposal of Kiparsky and Staal (1969): cross-linguistically, the passive is well-attested applying to causatives, but causativisation is not found applying to passives. Nevertheless, both data from certain Bantu languages and existing theoretical analyses of the Sanskrit causative cast doubt on the otherwise widespread assumption that PC-S is necessarily the passive of the ACC-ACC causative, and PC-O the passive of the OBL-ACC causative. The question requires empirical evidence, which has been hitherto lacking. The data we present in this paper allows us to fill this gap.

4 Prior Research on Sanskrit

As mentioned in the previous section, for Pāṇini and the indigenous grammatical tradition that followed him, it was only a semantically specific subset of verbs which could form ACC-ACC causatives, and all but two verbs were restricted to either ACC-ACC or OBL-ACC. It has long been observed, however, that the reality of attested Sanskrit usage is rather different: many verbs show both ACC-ACC and OBL-ACC causatives, and although different verbs may differ in how frequently they show one or the other pattern, it does not appear possible to claim that either pattern is definitely excluded for any verb. Modern treatments have therefore sought to explain the choice between the two patterns on semantic bases, based on notions such as the 'intended expression' (Speyer 1886, 37-8), the 'affectedness' or 'agency' of the embedded subject (Hock 1981; Bubeník 1987), or '(non-)contactive' causation (Bubeník 1987). Such proposals are in line with the semantically oriented approach advocated by Cole (1983), discussed above.

It is of course difficult to draw clear semantic distinctions between almost identical constructions in a language which no longer has native speakers. Indeed, the semantic distinctions drawn between the ACC-ACC and OBL-ACC causatives by previous authors do not appear immediately reconcilable. For example, Hock (1981, 21) states that "the causee [= our 'embedded subject'] marked by the instrumental seems to be less saliently the agent [than the causee marked by the accusative]", and likewise "the instrumental [causee] is marked for decreased 'agency' as compared to the accusative" (Hock 1981, 24). In apparent contrast, for Bubeník (1987, 690), "the causee in the accusative implies low retention of control [by the causee]", whereas instrumental marking "leaves greater control in the hands of the causee". Bubeník's (1987) account is in line with that of Cole (1983), while Hock's (1981) account stands in sharp contrast.

In our view, both Hock (1981) and Bubeník (1987) (and Cole 1983) are trying to describe the same difference between ACC-ACC and OBL-ACC, but do not do so clearly and compatibly because their focus is on the embedded subject alone. Although superficially the only difference between ACC-ACC and OBL-ACC is indeed the status (specifically, the case and grammatical role) of the embedded subject, the semantic difference does not rest fundamentally on the embedded subject, but on the verb itself and the verb's relation with its arguments, most importantly its object argument. As pointed out by Börjars and Vincent (2008), since at least Fillmore (1968, 25) it has been recognised that certain arguments of predicates represent

> the semantically most neutral case, the case of anything representable by a noun whose role in the action or state identified by the verb is identified by the semantic interpretation of the verb itself.

What is in Fillmorean terms the 'Objective' case, is in more modern terminology the semantic role of Theme; following Börjars and Vincent (2008, 164), "what is standardly called Theme is simply a projection into an internal argument slot of some or all of the lexical semantic content of the predicate". But to look at it the other way around, the lexical semantic content of a predicate depends on its Theme, indeed on its core internal argument(s), since Themes standardly (in the active) surface as core objects. If we take the OBL-ACC causative, then, the embedded object is the Theme of the causative predicate, and the causative predicate is understood in these terms: the interpretation of *kārayati* in (13b) is fundamentally about an event of 'making' interpreted in relation to a mat: Yajñadatta causes a mat to be made, through the agency of Devadatta. But with the ACC-ACC causative (13c), Devadatta is the core object argument of the causative predicate, and the predicate is thus interpreted in relation to Devadatta as the Theme (of the causation): Y. causes D. to act, and that action is the making of a mat.[13]

(13) a. *devadattaḥ* *kaṭaṃ* *karoti*
D.NOM mat.ACC make.3SG
'Devadatta makes a mat'.

b. *yajñadatto* *devadattena* *kaṭaṃ* *kārayati*
Y.NOM D.INS mat.ACC make.CAUS.3SG
'Yajñadatta has a mat made by Devadatta'.

c. *yajñadatto* *devadattaṃ* *kaṭaṃ* *kārayati*
Y.NOM D.ACC mat.ACC make.CAUS.3SG
'Yajñadatta makes Devadatta make a mat'.

Hock's (1981) claim that the instrumental marks 'reduced agency' of the embedded subject is thus really trying to describe a lesser focus on the action of the agent in the OBL-ACC construction, since the predicate is interpreted in primary relation to the embedded object Theme, rather than a difference in the semantic entailments associated with the embedded subject. And Bubeník's (1987) reading of 'low retention of control' by accusative-marked embedded subjects reflects the fact that in an ACC-ACC causative the predicate is interpreted primarily in relation to the embedded subject

13 We here use constructed examples, based on those offered in the grammatical tradition, to provide precise parallels, but the same points apply, *mutatis mutandis*, to (3b) and (4).

interpreted as a Theme of the event of causation (rather than as an Agent of the caused event).

In any case, the semantic difference between ACC-ACC and OBL-ACC is a separate question from that of the relations between the active and passive patterns. Whatever specific entailments influence the choice, or derive from the use, of one or another active pattern, we would expect those entailments to be reflected in whichever passive is associated with whichever active. But it is difficult, if not impossible, to use such fine semantic differences as firm empirical evidence for associations between active and passive patterns, at least in Sanskrit, because we are dealing with a corpus language and cannot rely on speaker intuitions (it would be difficult even in a language with native speakers). Below, we show that distributional data from our corpus study can provide more robust evidence in regard to the association between active and passive causative structures.

5 Restrictions on the Data

Although causatives are highly productive in Sanskrit, and although passive causatives are frequent in our corpus and formed to a large number of different verbs, the specific comparisons under investigation here could only be meaningfully compared in relation to ten Sanskrit verbs. In this section we explain why only such a small set of verbs were useable.

Based on the list of verb forms in Whitney 1885, together with other standard grammars, we identified 241 verbal roots which have attested active causatives and could conceivably also form passive causatives (which are not consistently listed by Whitney 1885).[14] Of these 241 roots, only 140 are always or sometimes transitive (in the sense of taking an accusative object argument).[15] Intransitive roots

14 For most verbs, active causatives are morphologically clearly distinct from non-causatives. However, the suffix *-aya* used in the causative is also used for some non-causative present stems; these are mostly denominatives in origin (originally with a suffix *-ya-*). For example, the verb *varṇayati* 'describes' is not a causative, but a denominative based on the noun *varṇa-* 'colour, shade'; synchronically the Indian tradition nevertheless treated it as an *-aya-* present to a root *varṇ*. There is some diachronic interaction between causatives and denominatives of this sort, and this may be a factor in the complicated picture of causative stems discussed below. For example, to the noun *kāma-* 'desire' a denominative *kāmayate* 'desires' was formed, which was reinterpreted as a causative based on a (previously non-existent) root *kam* 'desire', resulting in the formation of other finite verbal forms to this secondarily extracted root; see Jamison 1983, 75.

15 On the problematic nature of defining transitivity in the context of Sanskrit, see e.g. Kulikov 2012b; Lowe 2017a, 4-34. For the present purposes it is sufficient to take 'accusative object argument' to mean an accusative argument which regularly becomes the subject in the passive (thus excluding goal accusatives).

(including verbs of motion which can take goal accusatives, and verbs which take only clausal complements) cannot display the patterns under investigation, so those roots were the first to be excluded.

Forty-five of the remaining 140 roots are morphologically ambiguous in the passive: their passive causative is, or would be if attested, formally identical to their basic passive. For example, the root *āp* 'obtain' forms an active present stem *āpnoti* 'obtains', a basic causative *āpayati* 'causes to obtain', and a basic passive *āpyate* 'is obtained'. If a passive were to be formed to the causative stem, it would have the form *āpyate*, indistinguishable from the basic passive. Such a passive causative would be clearly identifiable only if it were PC-S, i.e. if it had the sense 'was caused to obtain', since its subject argument would then be different from the basic passive 'is obtained'. But a PC-O passive 'was caused to be obtained' would in most cases be impossible to distinguish from the basic passive. That is, it is often contextually unproblematic to read a basic passive as if it were a PC-O passive causative, and if we cannot always be absolutely sure of the difference based on context, our figures for the PC-O passive causative would be inflated by cases of the simple (non-causative) passive. Thus we are left with only 95 morphologically reliable, transitive roots which are recognised to form causatives.

Our study sought not only to compare finite active with passive, but also to compare *ta*-participle with both finite active and passive. It was therefore necessary to restrict our study further to only those roots which are attested in all three categories. Fifty of the remaining 95 roots are unattested in either the finite passive or the *ta*-participle (mostly the former) in our corpus.[16] For example, the root *likh* 'write', a common verb widely attested in the active causative and in the causative *ta*-participle *lekhita-* 'caused to write / caused to be written', is not attested in the expected finite causative passive **lekhyate*.

Of the remaining 45 roots, 35 are semantically problematic, falling broadly into two groups. Twenty roots form causatives which, while being genuine morphological causatives, are not semantically causative, functioning rather as simple transitive stems alongside the existing transitive stems of the root. For example, the root *kṛt* 'cut' forms a present *kṛntati* 'cuts', and a morphological causative *kartayati*, which however means the same as the simple present. There also is a passive of the latter, again entirely regular in its morphology, *kartyate* 'is cut'. At times it would be possible to force a PC-O passive causative reading onto *kartyate*, but it is never necessary:[17]

16 The majority of these appear to be unattested in the relevant category outside our corpus as well.

17 Several of the relevant roots form their regular simple present with a nasal element, either a nasal infix in the root (e.g. *kṛt*, *kṛntati* or *kartayati* 'cuts'; *stambh*,

(14) *sa* *taiḥ* *saha* *śastreṇa*
he.NOM they.INS with knife.INS
yudhyate *yāvat* *tathā* *kartyate*
fight.PASS.3SG until thus cut.CAUS.PASS.3SG
'With the knife he fights with them, until he is thus hacked to pieces (lit. 'is (caused to be?) cut')'. (*Śikṣāsamuccaya* 4)

Since the active causative is simply transitive, like the causative of an intransitive verb, and since the (morphological) passive causative need never be treated as anything other than a passive to a transitive stem, the data for this and parallel roots can tell us no more than the data for unambiguously intransitive roots, and must be excluded.

In some cases, the problem lies in the basic verb, which can be either transitive or intransitive. For example, the root *vah* has a simple present *vahati* which can be intransitive, 'travels, is conveyed', or transitive, 'conveys, transports'. The causative *vāhayati* can function as the causative of the basic verb in its transitive or intransitive sense, and the passive causative *vāhyate* is likewise ambiguous. Again, certain constructions with the active or passive causative are clearly identifiable as causatives to the transitive base, such as ACC-ACC actives and PC-S passives. But OBL-ACC, O-ACC and passive PC-O cannot reliably be distinguished from causatives based on the intransitive sense of the root, as can be seen immediately below. The root as a whole must therefore be excluded.

(15) *saṃvāhyantāṃ* *ca* *śakaṭair* *naukābhir* *mā* *vilambatha*
travel.CAUS.PASS.3PL and cart.INS.PL ship.INS.PL neg take.2PL
'and let them be conveyed (lit. 'be caused to travel / be made to be carried') by carts (and) ships; do not delay'. (*Brahma Purāṇa* 47.9)

The other broad group of semantically problematic verbs are those whose causatives have unpredictable idiomatic or lexicalised meanings, which do not correspond to the expected sense of a causative to the base verbal sense. For example, the causative of *ā-jñā* 'perceive, understand' has undergone a semantic development which means that it no longer functions as a regular causative: *ājñāpayati* does not

stabhnoti or *stambhayati* 'stops, supports'; *lup*, *lumpati* or *lopayati* 'breaks'), or (synchronically) a suffix (*dṝ*, *dṛṇāti* or *dārayati* 'tears'; *stṝ*, *stṛṇāti* or *stārayati* 'spreads'; *vṛ*, *vṛṇoti* or *vārayati* 'covers'). Many of these alternative *-aya* forms have a historical explanation; see Jamison 1983, 178-89, for the early history of *-aya* and its transitive use, and also Renou 1961, 473-4, on the variable sense of some *-aya* formations.

merely mean 'causes to perceive/understand', but 'commands'.[18] Similarly the causative of *vac* 'speak' regularly refers to a sound being produced, e.g. someone playing an instrument or reading out a letter. That is, the causative of *vac* can mean, literally, that someone is causing something (a lute/a letter) to speak, as in the following example.

(16) *vācite* *lekhe* *siddhārthena*
speak.CAUS.TA-PTC.LOC letter.LOC S.INS
'when this letter was read by Siddhārtha' (*Kathāsaritsāgara* 8.1.161)

Yet the corresponding usage does not exist in the basic verb: letters, musical instruments, etc. are not said to 'speak' using the non-causative of *vac*; the causative is therefore not the regular causative of the basic verb.[19]

Some causatives also show additional argument structure patterns which cannot be subsumed under causative syntax proper. The verb *bhojayati*, causative of *bhuj* 'to eat, enjoy', has two broadly synonymous constructions, one a genuine causative 'make eat', the other more naturally translated 'feed (with)'. The genuine causative shows the ACC-ACC pattern ('make someone (ACC) eat something (ACC)'), while the alternative constructions shows an instrumental of the substance fed ('feed someone (ACC) with something (INS)'), resulting in an 'ACC-OBL' pattern which is otherwise not found with causatives. Although it is possible to distinguish the 'ACC-OBL' construction (17a) from the genuine causatives (17b) when the foodstuff is expressed, there is no way to distinguish the two constructions when it is omitted (17c).

(17) a. *tato* *'nnena* *avaśeṣeṇa* *bhojayed* *atithīn* *api*
then food.INS remaining.INS eat.CAUS.OPT.3SG guest.ACC.PL too
'Then he should feed the guests too with the remaining food'. (*Mahābhārata* 13.100.17)

18 In this case, it is only the forms of *ā-jñā* which were excluded. Causatives of the simplex root *jñā* 'know' or *jñā* in combination with other preverbs are semantically regular; for example, to *jñā* 'know' the causative, *jñāpayati*, means 'informs, lets know'.

19 Regular uses of the causative of *vac* are found, but they cannot always be securely distinguished from the specialised sense, and so the root as a whole was excluded. On the causative of the similar root *vad* 'speak', which may have influenced the causative of *vac*, see Kulikov 2012a, 697.

b. *ikṣūṃś* *ca* *madhu-lājāṃś* *ca*
sugarcane.ACC.PL and honey-grain.ACC.PL and
bhojayanti *sma* *vāhanān*
eat.CAUS.3PL past draft-animal.ACC.PL
'They made the draft animals eat both sugarcane and honied grains'. (*Rāmāyaṇa* 2.85.52)

c. *dvijān* *dvādaśa* *bhojayet*
brahmin.ACC.PL twelve.ACC eat.CAUS.OPT.3SG
'He should feed the twelve brahmins'. (Lit. 'he should make the twelve brahmins eat') (*Nārada Purāṇa* 1.121.19)

Since genuine cases of ACC-O could not therefore be distinguished from a reduced form of the non-causative ACC-OBL construction, the data for this root was omitted.

Semantic specialisation is of course a gradient notion, and it is not easy to draw a clear line. As in the case of *bhuj*, we have been guided by syntactic factors as well as semantic, and this leads us to include two causatives which are sometimes considered semantically specialised. The causative of *dṛś* 'see' appears both with an ACC-ACC argument structure ('make someone (ACC) see something (ACC)'), which is unproblematic for a true causative, and an OBL-ACC structure, which is more naturally translated as 'show' than 'make seen': 'show something (ACC) to someone (GEN/DAT)'). But even if there is arguably a degree of semantic specialisation here (from 'make see(n)' to 'show'), for our purposes the argument structure patterns shown by *darśayati* are entirely consistent with a true causative, and there is no syntactic argument for excluding the root. We also include the causative *ghātayati*, morphologically suppletive to the root *han* 'slay', which is sometimes considered semantically specialised in the sense 'have killed, have executed', usually O-ACC. Again, *ghātayati* does show entirely regular causative argument structures, including instances of ACC-ACC (which can only be treated as genuine causatives), and so there is no syntactic reason to exclude it.

For the reasons discussed in this section, only ten roots proved viable for our investigation.[20] The ten verbs included in our study range across the categories that Pāṇini specifies as ACC-ACC, OBL-ACC, or both, permitting us also to compare our data with Pāṇinian prescriptions. Altogether, our corpus contains 1660 relevant tokens for these ten verbs (991 finite actives/middles, 77 finite passives and 592 *ta*-participles). We present and offer an initial discussion of this data in § 6, and then turn to a statistical analysis in § 7.

20 As noted above, certain clearly distinguishable subsets of data for these roots, such as the causative of *ā-jñā*, were excluded.

6 Data and Analysis

Before we introduce our data, we briefly summarise the relevant issues and the potential correlations we might expect to find. We state a list of explicit predictions, so that each issue can be easily referred back to later on.

6.1 Predictions and Possibilities

Above we introduced six types of active causative: ACC-ACC, OBL-ACC, 0-ACC, ACC-0, OBL-0 and 0-0. Firstly, we would predict that ACC-0 positively correlate with ACC-ACC, since the former can only be a reduced form of the latter. That is, verbs which more frequently form ACC-ACC active causatives should be statistically more likely to form ACC-0 causatives, since they are underlyingly the same formation.

- Prediction 1: positive correlation between ACC-ACC and ACC-0.

For the same reason, the equivalent correlation should hold between OBL-ACC and OBL-0.

- Prediction 2: positive correlation between OBL-ACC and OBL-0.

More tentatively, we may expect a closer correlation between 0-ACC and OBL-ACC than between 0-ACC and ACC-ACC, based on the idea that oblique arguments are more omissible than core object arguments. That is, while 0-ACC may in principle represent a reduced form of either OBL-ACC or ACC-ACC, if the embedded subject of the ACC-ACC causative is a core object argument, this should be less frequently omissible than the oblique argument of the OBL-ACC causative.

- Prediction 3: 0-ACC is more closely correlated with OBL-ACC than with ACC-ACC.

The question of a correlation between ACC-ACC and OBL-ACC is complex. If our data were to follow Pāṇini's prescriptions, that is if all but two verbs were to form exclusively either an ACC-ACC or an OBL-ACC causative, we would expect either no correlation or an inverse correlation. If there is a positive correlation between ACC-ACC and OBL-ACC, this would mean that the more a verb forms ACC-ACC causatives, the more it is also likely to form OBL-ACC causatives. This would therefore speak against lexical constraints, or even strong lexical preferences, for one active causative type over another, suggesting a more contextual semantic basis for the choice between ACC-ACC and OBL-ACC, rather than a grammatical or lexical semantic basis. As discussed above, contextual semantic distinctions are

assumed to underlie the alternation in most modern work (Speyer 1886; Hock 1981; Bubeník 1987).

- Prediction 4: OBL-ACC and ACC-ACC are positively correlated.

We can then consider the two finite passive categories, PC-S and PC-O. If the widespread assumption that PC-S is the passive of the ACC-ACC causative and PC-O the passive of the OBL-ACC causative is correct, then we might expect to find two things: firstly, positive correlations both between PC-S and ACC-ACC, and between PC-O and OBL-ACC; and secondly, a similar degree of correlation between PC-S and PC-O as between ACC-ACC and OBL-ACC. That is, we expect verbs which form ACC-ACC active causatives to correspondingly form PC-S passives more frequently, and the equivalent, *mutatis mutandis*, for OBL-ACC and PC-O. On the other hand, under the approach of Kiparsky and Staal (1969), according to which the OBL-ACC causative involves an underlying passivised base, we might expect positive correlations between ACC-ACC and both PC-S and PC-O, and no significant correlation between the passive categories and OBL-ACC. For the sake of argument, we frame our predictions in terms of the more widespread assumptions.

- Prediction 5: positive correlation between PC-S and ACC-ACC.
- Prediction 6: positive correlation between PC-O and OBL-ACC.
- Prediction 7: a similar degree of correlation between PC-S and PC-O as between ACC-ACC and OBL-ACC.

Considering now the *ta*-participle types NOM-ACC and OBL-NOM, we would expect these categories to correlate with the finite active ACC-ACC and OBL-ACC categories, respectively. In fact, if we assume that the *ta*-participle is a paradigmatically active (or at least distinctly non-passive) formation, we should find not only strong correlations between ACC-ACC and NOM-ACC, and between OBL-ACC and OBL-NOM, but also similar correlations between each member of these pairs and the other (e.g. passive) categories.

- Prediction 8: positive correlation between ACC-ACC and NOM-ACC.
- Prediction 9: positive correlation between OBL-ACC and OBL-NOM.
- Prediction 10: similar correlations between ACC-ACC and NOM-ACC, and between OBL-ACC and OBL-NOM, with other categories.

6.2 Active Causative

We now turn to the data itself. In § 2 we introduced the different categories of active causative. Here we repeat the examples for ease of reference:

(18) a. OBL-ACC:
geyaṃ *sa* *dhīmān* *vipra-varair* *akārayat*
song.ACC this.NOM.M wise.NOM.M brahmin-best.INS.PL do.CAUS.3SG
'The wise one made the best brahmins perform (lit. make) a song'.
(*Vāmana Purāṇa* 68.59)

b. ACC-ACC:
nalaṃ *setum* *akārayat*
N.ACC bridge.ACC make.CAUS.3SG
'He had Nala make a bridge'. (*Rāmāyaṇa* 6.114.41)

c. 0-ACC:
prātar *utthāya* *tat* *sarvaṃ* *kārayāmi* *karomi* *ca*
early rise.ABS this.ACC.N all.ACC.N do.CAUS.1SG do.1SG and
'Having risen early I have (someone) do and (myself) do all this'.
(*Mahābhārata* 13.124.15)

d. ACC-0:
mānuṣā *mānuṣān…* *kārayanti* *divā-niśam*
man.NOM.PL man.ACC.PL do.CAUS.3PL by.day-by.night
'Men… cause men to work day and night'. (*Mahābhārata* 12.254.39)

e. OBL-0:
śrāvayec *chraddadhānānāṃ* *tīrthapāda-pada-āśrayaḥ*
hear.CAUS.3SG faithful.GEN.PL Kṛṣṇa-foot-resorting.NOM
'One who resorts to the feet of Kṛṣṇa should make the faithful hear (the story of Dhruva, i.e. by reciting it)'. (*Bhāgavata Purāṇa* 4.12.50)

f.
karoty *atra)* *kārayanti* *kārayanti*
do.3SG here do.CAUS.3PL do.CAUS.3PL
na *te* *bhaṭāḥ*
not they.NOM.PL servant.NOM.PL
'(He wants to take rest here but) the servants do not allow (him) to do so'.
(*Garuḍa Purāṇa* 2.5.98)

We begin with ACC-ACC and OBL-ACC. Table 1 presents the figures for these patterns for the ten roots included in our study **[tab. 1]**. For comparison, the roots are grouped according to Pāṇini's categories: those that should prescriptively show OBL-ACC only, the two verbs that can take either OBL-ACC or ACC-ACC, and verbs which should prescriptively take ACC-ACC only.

Table 1 Active causatives

Root	**ACC-ACC**		**OBL-ACC**		**ACC-ACC**	**Pāṇini**
	(full)	ACC-0	(full)	OBL-0	proportion	
han 'kill'	2	0	9	0	.18	OBL-ACC only
pac 'cook'	0	0	1	0	0	
grah 'seize'	12	4	26	2	.36	
dā 'give'	4	3	0	0	1	
kṛ 'do'	14	1	6	0	.7	either
hṛ 'carry'	5	0	4	0	.56	
jñā 'know'	5	15	1	1	.91	ACC-ACC only
śru 'hear'	27	7	3	3	.85	
paṭh 'recite'	3	2	0	0	1	
dṛś 'see'	17	9	130	11	.16	
Total	**89**	**41**	**180**	**17**	**.40**	

Table 1 shows that both ACC-ACC and OBL-ACC are attested across a range of roots, beyond the restrictions observed by Pāṇini. (19) gives two examples of this, ACC-ACC with the theoretically OBL-ACC only verbs *han* 'strike, kill' and *dā* 'give' respectively.

(19) a. *evaṃ* *daśa* *sutās* *tasya*
thus ten daughter.ACC.PL he.GEN
kaṃsas *tān* *aghātayat*
K.NOM they.ACC kill.CAUS.IMPF.3SG
'Thus, Kaṃsa caused them to kill that one's ten daughters'.
(*Brahmāṇḍa Purāṇa* 2.71.182)

b. *taṃ* *nija-svāminaṃ* *siṃhaṃ*
him.ACC own-master.ACC lion.ACC
tasya *abhayam* adāpayat
him.GEN without-fear.ACC give.CAUS.IMPF.3SG
'He made his own master, the lion, give safe passage to him'.
(*Kathāsaritsāgara* 10.4.69)

Although Pāṇini's prescriptions are not universally observed, as was already well known, table 1 shows that there are some tendencies in

that direction: three of the four roots which for Pāṇini are ACC-ACC only show a greater than 0.8 proportion of ACC-ACC structures, the three highest proportions of ACC-ACC in our data, while three of the four roots with the lowest proportion of ACC-ACC, all below the average of 0.4, are found with roots which for Pāṇini should be OBL-ACC only.[21] In addition, the two roots for which Pāṇini licences both structures show proportions roughly in the middle. Two roots, *dṛś* and *dā*, go against expectations. This partial conformance to Pāṇini's norms could be a result of conscious attempts to follow Pāṇinian grammar by some authors, or may reflect a deeper feature of Sanskrit grammar which Pāṇini himself recognised and over-prescriptively incorporated into the *Aṣṭādhyāyī*, or indeed, a combination of both. If it were only the former, i.e. conscious conformity with Pāṇini, then one might have expected that the more strictly Classical texts in our corpus would conform more closely with the Pāṇinian norms, and that the linguistically freer Epics, for example, might show more deviation.[22] However, we found no pattern in the distribution of the data in terms of more or less 'Pāṇinian' Sanskrit; forms violating Pāṇinian norms are evenly distributed across the different genres in our corpus. This may suggest that these tendencies are a more ingrained aspect of Sanskrit grammar, which Pāṇini merely imperfectly reflected in his grammar.

The greater part of the active data involves structures which are ambiguous between ACC-ACC and OBL-ACC due to the omission of the first argument; in particular, 0-ACC alone makes up more than half the active causative data, as shown in table 2 **[tab. 2]**.

Table 2 Active causative distributions with omission of first argument

Root	0-0	0-acc	Pāṇini
han 'kill'	8	75	OBL-ACC only
pac 'cook'	2	18	
grah 'seize'	0	16	
dā 'give'	1	23	
kṛ 'do'	4	114	either
hṛ 'carry'	5	58	
jñā 'know'	6	6	ACC-ACC only
śru 'hear'	30	25	
paṭh 'recite'	1	2	
dṛś 'see'	32	238	
Total	**89**	**575**	

21 Excluding *dṛś* (since the high token frequency of OBL-ACC with this root overwhelms the other figures), this contrast between Pāṇini's OBL-ACC only and ACC-ACC only roots is statistically significant (Fisher's exact test $p = 5.75 \times 10^{-9}$).

22 Such a distribution is found with other phenomena in Sanskrit, cf. Lowe 2017b.

There are no clear patterns in the distribution of O-ACC here; it is frequent with all roots.

6.3 Passive Causatives

Table 3 presents the distribution of all finite passive causative forms of the ten roots attested in our corpus in terms of PC-S vs. PC-O. The ordering of roots in the table is the same as that in tables 1 and 2. Examples of passive causatives for all ten roots are provided in the appendix.

Table 3 Passivisation patterns in finite passive causatives

Roots	pc-s	pc-o	PC-S Proportion
han 'kill'	0	1	0
pac 'cook'	0	1	0
grah 'seize'	1	2	.33
dā 'give'	4	3	.57
kṛ 'do'	8	13	.38
hṛ 'carry'	2	3	.4
jñā 'know'	9	14	.39
śru 'hear'	1	1	.5
paṭh 'recite'	1	0	1
dṛś 'see'	0	13	0
Total	**26**	**51**	**.34**

As table 3 shows, the number of attested finite passive causatives for most roots is very small, except for *kṛ* 'do', *jñā* 'know', and *dṛś* 'see'. Nevertheless some observations can be made. According to previous studies, PC-O is highly restricted: Speyer (1886, 37-8) states that this pattern (i.e. the embedded object becoming the subject in the passive) is very rare compared with PC-S; Bubeník (1987) finds PC-O to be frequent only with the verb *han*. In contrast, table 3 shows that PC-O is attested with all but one of our ten roots, and is more common overall than PC-S.

Table 3 also shows considerable variation between the different roots in terms of their relative preference for PC-S or PC-O. Excluding *dṛś*, which is again an outlier, all roots which are attested more than once show a relatively even distribution of PC-S vs. PC-O: between 0.33 and 0.57 proportion of PC-S. The three roots which are attested only once all pattern in the 'expected' directions, based on Pāṇini's prescriptions: *han* and *pac* as OBL-ACC only roots show only PC-O, and *paṭh* as ACC-ACC only shows only PC-S. But since we are dealing with lone attestations for each root, this may be nothing more than chance.

6.4 *Ta*-Participle Causatives

The figures for this category, shown in table 4, are higher than those for the previous categories, due to the high productivity of the *ta*-participle in Sanskrit **[tab. 4]**.

Table 4 Patterns in ta-participle causatives

Roots	nom-acc	obl-nom	NOM-ACC Proportion
han 'kill'	0	73	0
pac 'cook'	0	6	0
grah 'seize'	12	4	.75
dā 'give'	6	7	.46
kṛ 'do'	38	73	.34
hṛ 'carry'	1	28	.03
jñā 'know'	66	4	.94
śru 'hear'	34	17	.66
paṭh 'recite'	3	0	1
dṛś 'see'	11	209	.05
Total	**171**	**421**	**.29**

With *ta*-participles, OBL-NOM is more common overall than NOM-ACC; the proportion is similar to the proportion of PC-O with the finite passive causatives. The two roots unattested in NOM-ACC are both in Pāṇini's 'OBL-ACC only' category (= OBL-NOM in the ergative participle, cf. §2.4), while the two roots with the highest proportion of NOM-ACC are those in Pāṇini's 'ACC-ACC only' category (= NOM-ACC in the ergative). However, the root *dṛś*, for which 95% of occurrences are OBL-NOM, goes strongly against Pāṇinian expectations, although this is in line with its behaviour in the finite active and passive. The root *grah*, showing 75% NOM-ACC, also goes against Pāṇinian expectations, although less severely. Also, the root *hṛ* is surprisingly rare in NOM-ACC, given its otherwise fairly even distribution.

In this section we have offered merely a few superficial observations, particularly relating to the Pāṇinian analysis. In the following section, we investigate the correlations in our data in more detail.

7 Examining the Correlations

Our data is complex and varied, involving distinct patterns (OBL-ACC vs. ACC-ACC, PC-S vs. PC-O, etc.) in different but related data sets for a set of ten verbal roots which do not all pattern together and which show considerable variation in frequency of attestation. In this section we use correlation matrices to identify patterns in this complex dataset.

The 'correlation' in this term refers to the measure of the dependence between two variables. In a correlation matrix, this is expressed in the form of a correlation coefficient, which is measured on a scale from −1 to +1. The closer the value is to +1 or −1, the more closely the two variables are related. While the value of the coefficient tells us about the strength of the relationship, the sign (+/−) of the coefficient indicates the direction of the relationship, positive or negative respectively. A positive correlation coefficient means that the two variables correlate in the same direction: an increase in one is accompanied by an increase in the other and a reduction in one is accompanied by a reduction in the other. A negative correlation coefficient represents a negative correlation: when one variable increases, the other decreases, and vice versa. A correlation coefficient of 0 means that there is no correlation between the two variables: they are not related to each other at all. The correlation coefficients are not the same as p-values, but p-values for each coefficient can be (and were) generated. In the correlation matrices we mark coefficients with significant p-values (≤ 0.05), and discuss some other p-values below.

Our data consists of frequencies of token instances of verbs in particular syntactic constructions. Frequency data tends to follow a Zipfian distribution, which means that there are few very high-frequency items and many low-frequency items (Piantadosi 2014). This is a non-linear distribution, while traditional correlations (e.g. Pearson's correlation coefficient) are linear. We therefore applied a logarithmic transformation (the natural logarithm, base *e*, approximately 2.718282)[23] to the data in order to normalise the frequency data, in other words, to reduce the distance between the smallest and the highest data points, which otherwise would skew the statistical analysis. Furthermore, to address the zero values in our data set, we applied the 'Add-k' smoothing method, with k = 0.1 added to all values (Bellégo, Benatia, Pape 2021; Criscuolo, Overman, Van Reenen 2019; Jurafsky, Martin 2020).

23 The reason for using the natural logarithm is that higher bases tend to pull extreme values in more drastically than lower bases (Osborne 2002). For comparison, we also performed a transformation using base 10 and the correlations were not significantly different.

Table 5 Full correlation matrix for (transformed) data from tables 1-4

	0-0	0-ACC	ACC-0	OBL-0	ACC-ACC	OBL-ACC	PC-S	PC-O	NOM-ACC	OBL-NOM
0-0	1	0.48	0.06	0.30	0.22	0.28	-0.22	0.30	0.08	0.57
0-ACC	0.48	1	-0.21	0.20	0.28	**0.72**	-0.30	0.60	-0.07	**0.96**
ACC-0	0.06	-0.21	1	**0.72**	**0.64**	0.04	0.39	0.29	**0.87**	-0.15
OBL-0	0.30	0.20	**0.72**	1	0.58	0.62	-0.16	0.36	0.55	0.27
ACC-ACC	0.22	0.28	**0.64**	0.58	1	0.39	0.41	0.35	**0.78**	0.28
OBL-ACC	0.28	**0.72**	0.04	0.62	0.39	1	-0.40	0.50	0.08	**0.73**
PC-S	-0.22	-0.30	0.39	-0.16	0.41	-0.40	1	0.30	**0.69**	-0.25
PC-O	0.30	0.60	0.29	0.36	0.35	0.50	0.30	1	0.47	**0.69**
NOM-ACC	0.08	-0.07	**0.87**	0.55	**0.78**	0.08	**0.69**	0.47	1	-0.01
OBL-NOM	0.57	**0.96**	-0.15	0.27	0.28	**0.73**	-0.25	**0.69**	-0.01	1

Bold: Correlations with significant p-value (≤ 0.05)

Table 5 shows the correlation matrix (generated using R) for the transformed values of the data from tables 1-4 **[tab. 5]**. That is, the variables underlying the correlation matrix are the log-transformed numbers of finite PC-S, finite PC-O, *ta*-participle NOM-ACC etc. in our corpus for each root. The figures in the table show the correlations between those variables. For example, the 0.48 figure near the top left represents a positive correlation between the number of (finite active) 0-0 tokens and the number of (finite active) 0-ACC tokens in our corpus; that is, the larger the number of 0-0 tokens for any root, in general the larger the number of 0-ACC, and vice versa.[24]

For convenience, the correlation matrix in table 5 can be considered in three sections, indicated by the dashed lines, partly corresponding to the relations between the different supercategories of our data.[25] The top left quadrant of table 5 shows the correlations between the six finite active categories, that is between ACC-ACC and OBL-ACC, ACC-ACC and ACC-0, etc. The bottom right quadrant includes three (smaller) sets of correlations: the correlation between the two passive categories, PC-S and PC-O; the correlation between the two *ta*-participle categories, NOM-ACC and OBL-NOM; and the correlations between these passive and participial categories. The bottom left and top right quadrants (which include the same information) show the correlations between the active categories and both the passive and participial categories.

24 Most of the correlations in table 5 are positive, and the negative correlations that there are are all low. This is an artefact of the type of data we are considering, and has no significant consequences.

25 The grouping of finite passive with *ta*-participle is simply because these together constitute four categories, vs. the six categories of finite active, permitting the table to be divided into four roughly equal parts. If preferred, one may ignore the dashed lines and take the table as a whole.

7.1 The Active Correlations

We begin with the active correlations. We find three statistically significant correlations in this (top left) quadrant of the matrix. Above, we presented a number of predictions (§ 6.1): that we should expect a positive correlation between ACC-ACC and ACC-O (prediction 1), reflecting the fact that the latter is a reduced form of the former, and similarly a positive correlation between OBL-ACC and OBL-O (prediction 2); also, that we should expect a closer correlation between O-ACC and OBL-ACC than between O-ACC and ACC-ACC (prediction 3), based on the idea that oblique arguments are less obligatory than core object arguments; and that we might expect, given the claims of previous literature, a positive correlation of some sort between ACC-ACC and OBL-ACC (prediction 4), assuming that the choice between the two is based more on context than on grammatical or lexical semantic constraints.

ACC-ACC positively correlates with ACC-O (0.64, $p = 0.0443$), supporting prediction 1. In terms of prediction 2, OBL-ACC does positively correlate with OBL-O, but the p-value is marginally above the threshold for significance (0.62, $p = 0.0584$). The lack of significance here is likely due to the very small number of OBL-O tokens. O-ACC positively correlates with OBL-ACC (0.72, $p = 0.0177$); while there is a minor positive correlation between O-ACC and ACC-ACC, it is not significant. This supports prediction 3.

In relation to prediction 4, although there is a positive correlation between ACC-ACC and OBL-ACC, it is not significant. Interestingly, there is a significant positive correlation between ACC-O and OBL-O (0.72, $p = 0.0192$). Given that these categories respectively represent reduced forms of ACC-ACC and OBL-ACC, this could be taken to represent the correlation between the two non-reduced categories. But the question remains why this correlation is not found to the same degree with the non-reduced categories themselves, nor indeed between OBL-O and ACC-ACC, or between ACC-O and OBL-ACC. The strong correlation is specifically between the reduced categories; this may indicate that in our data certain roots disprefer reduced structures of either kind, while other roots license reduced structures of either kind, but it does not necessarily tell us anything about the correlation between ACC-ACC and OBL-ACC.

The data therefore does not allow us to make definitive claims regarding the relative importance of contextual considerations vs. lexical or grammatical constraints in the choice between ACC-ACC and OBL-ACC. As our raw data shows, most of our verbs attest both active causatives, meaning that hard grammatical constraints of the kind proposed by Pāṇini (i.e. that most verbs can only form one or another causative) are not (at least generally) at play. But still the lexical semantics of different verbs may predispose them to favour one or another type of causative, above and beyond purely contextual considerations.

To the extent that it may be possible to make such fine distinctions in a corpus-based language, this issue awaits future work.

7.2 The Passive and Active-Passive Correlations

Above, we predicted that we should find positive correlations between PC-S and ACC-ACC (prediction 5), and between PC-O and OBL-ACC (prediction 6), based on the assumption that PC-S is the passive of the ACC-ACC causative and PC-O the passive of the OBL-ACC causative. An alternative possibility was raised that OBL-ACC works differently, in which case we would expect correlations between ACC-ACC and both PC-S and PC-O.

There are no statistically significant correlations between the finite active and passive categories. In terms of the non-significant correlations, it is noticeable that PC-S is positively correlated with both ACC-ACC and ACC-O (0.41 and 0.39), but negatively correlated with all the other finite categories: OBL-ACC, O-ACC, OBL-O and O-O; note that our data primarily involves positive correlations, and these are four of only nine negative correlations out of the forty-five correlations we are investigating.[26] PC-O is also significantly correlated with OBL-ACC. Taking this all together with the significant correlation between O-ACC and OBL-ACC (which suggests that the former is often a reduced form of the latter), we can conclude that there is a closer relation between PC-S and ACC-ACC (and its variants) than between PC-S and OBL-ACC (or its variants), and likewise a closer relation between PC-O and OBL-ACC (and its variants) than between PC-O and ACC-ACC (or its variants). This provides statistical support for the traditional assumption that PC-S is the passive of the ACC-ACC causative, and PC-O the passive of OBL-ACC. However, the correlations are not absolute, and there appears to be considerable freedom in the formation of passives relative to actives.

The highest non-significant correlation between the finite active and passive categories is between PC-O and active O-ACC (0.60, $p = 0.0696$). Since O-ACC is ambiguous between ACC-ACC and OBL-ACC, this correlation does not contribute anything to our assessment of predictions 5 and 6. It was however noted above that the O-ACC construction may to some extent represent a separate construction, and not merely a reduced form of ACC-ACC and/or OBL-ACC. Considered as such, O-ACC would necessarily form a PC-O passive, so this near significant positive correlation may support that understanding of it. Note also that the correlation

26 Cf. fn. 23. PC-S is also negative correlated with OBL-NOM, meaning that the majority of negative correlations in our data (including the four most extreme negative correlations) concern PC-S.

between PC-O and O-ACC speaks against Kiparsky and Staal's (1969) analysis of OBL-ACC as a causative applied to a passive base.

Prediction 7 was that we might find a similar degree of correlation between PC-S and PC-O as between ACC-ACC and OBL-ACC, based again on the idea that PC-S is the passive of the ACC-ACC causative, and PC-O the passive of OBL-ACC. We do in fact find this: the degree of correlation between PC-S and PC-O is 0.30, while the correlation between ACC-ACC and OBL-ACC is 0.39. Since this is only a similarity in degrees of correlation, rather than a correlation itself, it must be interpreted cautiously, but it does add further support to the idea that PC-S and PC-O alternate as passives in correspondence, at least to some extent, with the alternation of ACC-ACC and OBL-ACC as actives.

7.3 The Participle Correlations

Above, we predicted that NOM-ACC should correlate positively with ACC-ACC (prediction 8), that OBL-NOM should correlate positively with OBL-ACC (prediction 9), and also that the correlations between NOM-ACC and ACC-ACC, and between OBL-ACC and OBL-NOM, and the other (e.g. passive) categories should be similar (prediction 10). These predictions are based on the assumption that the ergative *ta*-participle is a syntactically active category, albeit with ergative-absolutive morphosyntactic alignment.

These predictions largely hold. Participial NOM-ACC positively correlates with the active category ACC-ACC (0.78, $p = 0.0075$) and with its reduced form ACC-O (0.87, $p = 0.0011$); both these correlations are significant. In broad terms the degrees of correlation with different categories shown by NOM-ACC are all similar to those shown by ACC-ACC.

Similarly, participial OBL-NOM positively correlates with OBL-ACC (0.73, $p = 0.0167$) and O-ACC (0.96, $p<0.0001$); it shows broadly the same degrees of correlation with other categories as OBL-ACC.

Another more interesting difference between the finite active and the *ta*-participle categories may be observed, however. While there were no significant correlations between the main finite active and finite passive categories – that is, there was no significant correlation between ACC-ACC and PC-S, nor between OBL-ACC and PC-O, as originally predicted – the corresponding correlations between the participial categories and the passive are statistically significant. So NOM-ACC is positively correlated with passive PC-S (0.69, $p = 0.0264$), while OBL-NOM is positively correlated with PC-O (0.69, $p = 0.0275$).

This shows that the *ta*-participle categories pattern more closely with the passive than the finite active categories do. And this in turn provides evidence that the *ta*-participle category shows syntactic affinities with the passive beyond what would be expected from a standardly active formation.

To this point we have followed the mainstream modern understanding of the *ta*-participle formation as fundamentally non-passive: a morphosyntactically ergative formation, but syntactically aligned with the active, just like its descendant (the perfective aspect) in many modern Indo-Aryan languages. But to understand the position of the *ta*-participle in the causative data considered here, we must adopt a more nuanced understanding of the formation, which we address in the next section.

8 Discussion and Summary

8.1 Passive Causatives and the Ergative

The status of the *ta*-participle in Sanskrit is complicated, and there remains no agreement in the literature as to its origin and early development in the history of Sanskrit.[27]

The *ta*-participle is traditionally labelled a 'passive' participle in Western grammars (e.g. Whitney 1896, 340), but in fact this is a misnomer; the participle shows ergative-absolutive morphosyntactic alignment: with intransitive verbs the participle agrees in case, gender and number with the single (subject) argument (S), which appears in the nominative, while with transitive verbs the participle agrees with the object argument (O), which appears in the nominative, while the transitive subject/agent argument (A) gets instrumental case marking.

(20) *te* *'dya* *suptā* *mahītale*
they.NOM.PL.M today sleep.TA-PTC.NOM.PL.M ground.LOC
'Today they have fallen asleep on the ground'. (*Mahābhārata* 1.138.15d)

(21) *madra-rājaḥ* *kṛtaḥ* *śalyo*
M-king.NOM.M make.TA-PTC.NOM.M Ś.NOM
dhārtarāṣṭreṇa *mādhava* *senā-patiḥ*
Dh_son.INS M.VOC army-commander.NOM
'The son of Dhṛtarāṣṭra has made Śalya, king of the Madras, the commander of the army'. (*Mahābhārata* 9.6.22b–c)

27 Relevant literature includes: Pray 1976; Klaiman 1978; Andersen 1986; Bubeník 1989; 2001; Hook 1991; Peterson 1998; Jamison 2000; Butt 2001; Bynon 2005; Condoravdi, Deo 2014; Butt, Deo 2017; Patel-Grosz 2021.

The ergative when applied to a transitive verb is superficially identical to the passive of a transitive verb:[28] what would ordinarily be the object appears in the nominative case, the default subject case, and what would ordinarily be the subject appears in the instrumental, the same case as the demoted subject in the passive. Compare (21) with the finite passive in (22).

(22) *kriyate* *'yaṃ* *vidhir* *mayā*
do.PASS.3SG this.NOM.M ordinance.NOM.M me.INS
'This ordinance is observed by me'. (*Mahābhārata* 2.6.2d)

This superficial similarity does not make the *ta*-participle a passive, since the similarity is restricted to transitive verbs. The intransitive is crucially different. The alignment in (20) is superficially identical to the corresponding finite active (e.g. $sa_{[\text{NOM}]}$ *svapiti* 'he sleeps') and is different from the impersonal finite passive (e.g. $tena_{[\text{INS}]}$ *supyate* lit. 'it is slept by him').

Despite the persistence of the idea that the ergative construction in Indo-Aryan (including in Sanskrit) derives historically from a passive (most recently Patel-Grosz 2021), the fact is that as far back as we are able to reconstruct, the *ta*-participle and its ancestor, the Proto-Indo-European verbal adjective in *-*to*, never showed passive alignment in the sense of 'demoting' both the A of transitive verbs and the S of intransitives. Rather, it always showed ergative alignment in the sense of agreeing with the S of intransitives and the O of transitives.

The question of morphosyntactic alignment is, however, distinct from the question of paradigmatic alignment ('alignment' in the less technical sense). As discussed by Butt and Deo (2017), the *ta*-participle in the earliest attested Sanskrit is an adjectival stative-resultative formation. This would not have been specifically aligned with either the active or passive voice in the verbal system, since it was not originally a verbal construction.[29] At some stage in the history of Sanskrit (at the latest by the Epic Sanskrit period, according to Butt and Deo, but perhaps much earlier), it was reanalysed as a verbal formation. It came to serve as a basic past tense predicate from the late Vedic Sanskrit period onwards.

Although the *ta*-participle formation is fundamentally not a passive formation, but rather an active formation with ergative

28 Excepting a very few verbs which may – but need not – show nominative/accusative alignment in the *ta*-participle, but not the finite passive. These are all complex verbs formed of transitivising preverbs with intransitive bases.

29 And so we disagree with Butt and Deo's (2017) use of the term 'passive' to describe it at this period.

morphosyntactic alignment, we have seen evidence above from the causative and passive causative correlations showing that the *ta*-participle categories pattern more closely with the finite passive than the finite active categories do. This suggests that it is overly simplistic to treat the *ta*-participle as a purely active (or non-passive) category, and to assume that there is no synchronic paradigmatic association between the *ta*-participle and the passive in Sanskrit.

In fact, there are other signs of influence between the two. Although it is the less common pattern by far, the *ta*-participle can show 'passive' alignment with intransitive verbs. For example, in (23) the *ta*-participle of the intransitive verb *sthā* 'stand, stay' is exceptionally treated as an impersonal passive, exactly parallel to the (exceptionlessly impersonal) finite passive of the same verb, shown in (24).

(23) *virāṭanagare* *pārthaiḥ* *kathaṃ*
V.LOC Pāṇḍava.INS.PL how
mūḍhātmabhiḥ *sthitam*
foolish.INS.PL stay/be.TA-PTC.NOM.SG.NTR
lit. 'how could it be stayed in V. by the foolish Pāṇḍavas': 'how could… stay in V'. (*Bṛhatkathāślokasaṃgraha* 22.305)

(24) *yāvad* *anena* *muninā* *sthīyate*
while this.INS sage.INS stay.PASS.3SG
'while this sage stays (alive)' (lit. 'while it is stayed by this sage') (*Hitopadeśa* 4.5)

Here the S argument of the intransitive verb appears in the instrumental and does not show verbal agreement, just like the A argument of transitive verbs in both the ergative and the passive. This therefore reflects a nominative-accusative alignment of the participle (since S is treated like A), and more specifically a morphosyntactic patterning identical to a passive. This relatively rare use appears to represent the influence of the finite passive on the *ta*-participle, based on their superficial identity with transitive verbs. It appears, then, that despite the non-passive nature of the *ta*-participle, it could nevertheless be associated with the passive, even to the extent of being reconfigured to align more closely with it.

This evidence of association and influence between ergative and passive is perhaps surprising, since even in relation to transitive verbs, where ergative and passive are superficially identical in terms of morphosyntactic alignment, they should still be fundamentally distinct in terms of more purely syntactic phenomena: with the passive of a transitive verb, we expect the O argument to be the syntactic subject, while in the ergative we expect the A argument to show properties of syntactic subjecthood. However, in Sanskrit all syntactic

tests for subjecthood beyond agreement and case assignment target agents (or the most agentive argument), even in the passive.[30]

For example, one widely used test for identifying subjects relates to control of the subject of verbal adjuncts. In an ordinary active finite sentence, it is the nominative argument (of transitive or intransitive verb) that controls the subject of an absolutive clause:

(25)	*pāduke*	*ca*	*asya*	*rājyāya*	*nyāsaṃ*	*dattvā*
	sandal.ACC.DU	and	his	royal.power.DAT	symbol.ACC	give.ABS
	punaḥ	*punaḥ*	*nivartayām-āsa*	*tato*	*bharataṃ*	
	again	again	return.PERF.CAUS.3SG	then	B.ACC	

'But having$_{ij/*}$ given his sandals as a symbol of his royal power, Bharata's older brother$_i$ repeatedly urged Bharata$_j$ to return'. (*Rāmāyaṇa* 1.1.30)

With the *ta*-participle, it is standardly the nominative subject of an intransitive verb and the instrumental argument of a transitive verb which control the absolutive:

(26) a.	*guror*	*udayanaḥ*	*śrutvā*	
	teacher.GEN	U.NOM.SG.M	hear.ABS	
	nāga-lokaṃ	*gatas*	*tataḥ*	
	nāga-world.ACC	go.TA-PTC.NOM.SG.M	then	

'After listening$_i$ to his guru, Udayana$_i$ went to the world of the snake-people'. (*Bṛhatkathāślokasaṃgraha* 5.142)

b.	*ehi*	*iha*	*ca*	*mayā*
	come.IMPV.2SG	here	and	I.INS
	āhūya	*spṛṣṭaḥ*	*pṛṣṭhe*	*nirāmayaḥ*
	summon.ABS	touch.TA-PTC.NOM.M	back.LOC	healthy.NOM.M

'Having$_{ij/*}$ summoned (him) 'come here!', I$_i$ touched his$_j$ back (and he$_j$ was) healed'. (*Bṛhatkathāślokasaṃgraha* 9.79)

This looks like clear evidence for the subjecthood of the instrumental, which is what we expect for the (non-passive) ergative *ta*-participle. However, in the finite passive too it is the agent, not the grammatical subject, which controls the subject of the absolutive:[31]

30 See e.g. Hock 1982; 1986; 1987; 1990; 1991a; 1991b. As discussed by Hock (1986), there are infrequent examples of 'loose' constructions violating the standard patterns discussed here, but they do not undermine the significance of the generalisations.

31 Note that in this example, the instrumental agent of the passive not only controls the absolutive, but is also the antecedent of the reflexive pronoun *sva-* in *sva-śvasaiḥ*

(27) *mat-paraṃ* *durlabhaṃ* *matvā*
me-after hard.to.obtain.ACC think.ABS
nūnam *āvarjitaṃ* *mayā*
now offered.NOM me.INS
payaḥ *pūrvaiḥ* *sva-niḥśvāsaiḥ*
libation.NOM ancestor.INS.PL self-sigh.INS.PL
'The libation offered by me is now partaken of by my ancestors$_i$, made tepid by their$_i$ sighs, (on) thinking$_i$ that (it will be) hard to obtain after my death'. (*Raghuvaṃśa* 1.67)

Thus the ergative of transitive verbs and the passive are not distinguished by this syntactic test, and the same applies to all other syntactic tests for subjecthood known to us.[32] The status of the ergative *ta*-participle in Classical Sanskrit is therefore somewhat problematic. Although it is not a passive construction, it was morphosyntactically and syntactically indistinguishable from the passive when formed to transitive verbs, and shows evidence of occasional reanalysis and realignment as a passive. We might therefore say that while the *ta*-participle is not passive, it is less emphatically not a passive than the finite active; that is, the *ta*-participle is a non-passive (active) formation which nevertheless shows some association with the passive.

Returning to our data, although the participle is closer to the active than to the passive, the finite passive is also closer to the participle than it is to the finite active. So rather than seeing finite active and participle as constituting a single 'active' category opposed to the passive, we may rather be seeing an opposition between active and passive in which the finite categories stand most distant, and the participle stands in between, closer to the active, but nevertheless closer to the passive than the finite active.

8.2 A Note on Genre

In this paper we have treated our data as representing 'Sanskrit' understood as a relatively homogenous language form, while acknowledging that the texts from which our data comes vary considerably in terms of date and genre. To some extent this is a reasonable simplification, since in many respects Sanskrit changed relatively little after the late Vedic period, when Pāṇini's grammar of the language was codified. At the same time, variation in idiom between different

'by their own sighs'. Further examples are offered by Hock (1982, 132; 1986, 22), and Söhnen-Thieme (2019, 7), but in all of their examples the agent phrase is unexpressed.

32 See Hock 1986.

genres of Sanskrit can be significant at all periods of the language.

We noted above one small respect in which we found no significant variation: the adherence, or otherwise, to the Pāṇinian prescriptions restricting particular patterns to particular roots (i.e. ACC-ACC and its relatives to a specific subset of transitive roots). That is not to say that such patterns do not exist somewhere in Sanskrit literature, only that they did not appear in our data.

One small respect in which we do see variation between genres is in the relative frequency of the passive constructions in comparison with active causatives. We categorised part of our data according to three broad genres: Classical kāvya (including dramatic texts and verse narrative literature like the *Kathāsaritsāgara*), Epic Sanskrit (i.e. the *Mahābhārata* and *Rāmāyaṇa*), and *purāṇa* literature. Of 570 causative tokens from the kāvya literature, 44 (7.7%) were finite passives, 234 (41.1%) participles, and the rest active causatives. For the epics, of 498 tokens only 12 were finite passives (2.4%), and 158 (31.7%) participles. The purāṇas show a relative frequency of finite passives between that of the epics and the kāvya literature, but a higher proportion of participles than either: of 427 tokens for the purāṇas, 21 are finite passives (4.9%), and 196 participles (45.9%). The notably lower figures for the epics – less than a third of the proportion found in kāvya, for the finite passive – perhaps reflects something of the earlier pattern, found in Vedic, where passive causatives to transitive roots are largely unattested. In other respects, however, the three genres are not significantly different. For example, in comparing the relative frequency of ACC-ACC and related constructions (i.e. ACC-O, PC-S and NOM-ACC) vs. OBL-ACC and related constructions (OBL-O, PC-O and OBL-NOM), there is no significant difference: 136 (23.9%) of the kāvya tokens represent the ACC-ACC 'family', compared with 109 (21.9%) of the epic tokens, and 82 (19.2%) for the purāṇas.

It was not the aim of this paper to investigate chronological or genre variation in detail, however, and beyond these small observations, we leave it as a task for future work to investigate the variation between different texts and genres in respect of the causativisation and passivisation patterns discussed in this paper.

8.3 Summary

In this paper we have offered a detailed investigation of the interaction of causative and passive in Sanskrit. The corpus-based nature of the language, and practical complications in the data restricting usable token counts, combined with the multifaceted nature of the phenomena in question, render the underlying patterns obscure. Using fine-grained statistical analysis, we have demonstrated that even this problematic data can provide clear support for otherwise obscured syntactic patterns.

Our analysis has firstly adduced statistical support for the assumption of a relation between the ACC-ACC active causative and the PC-S passive causative, and between the OBL-ACC active and the PC-O passive. While it is clear that there are no hard grammatical constraints on the formation of the different types of active or passive causatives in Sanskrit, our data does not provide sufficient evidence to draw firm conclusions regarding the importance of contextual vs. lexical or grammatical constraints in the choices between ACC-ACC and OBL-ACC, and between PC-S and PC-O. We have also adduced statistical support for the presumed but never proven associations between OBL-ACC and O-ACC and between ACC-ACC and ACC-O.

In addition, we have shown that in the causative the *ta*-participle correlates both with the finite active and with the finite passive: NOM-ACC correlates with ACC-ACC and PC-S, and OBL-NOM correlates with OBL-ACC and PC-O. The participle correlates more closely with the active than with the passive, as expected given its status as an ergative (non-passive) construction, but also correlates more closely with the passive than does the active, supporting some paradigmatic association between *ta*-participle and passive, and showing that the *ta*-participle stands paradigmatically somewhere between active and passive.

Overall, we have shown that detailed statistical analysis of corpus data can yield new and interesting results for Sanskrit syntax, even on complex topics such as the syntax of the causative or the status of the *ta*-participle.

Data sources

The figures presented in this paper were based on data extracted (using Python) from a corpus of electronic texts collated from a number of sources by the authors and processed to produce a standardised encoding. The following list includes only those texts which yielded examples of causatives from the ten roots treated in sections 6-7 above.

From GRETIL: (gretil.sub.uni-goettingen.de):

Mahābhārata, Rāmāyaṇa, Agni Purāṇa, Bhāgavata Purāṇa, Brahma Purāṇa, Brahmāṇḍa Purāṇa, Garuḍa Purāṇa, Kūrma Purāṇa, Liṅga Purāṇa, Mārkaṇḍeya Purāṇa, Matsya Purāṇa, Nārada Purāṇa, Narasiṁha Purāṇa, Revakhanda of the Vāyu Purāṇa, Śiva Purāṇa, Vāmana Purāṇa, Viṣṇu Purāṇa, Śarīputraprakaraṇa, Veṇīsaṃhāra, Priya darśikā, Rātnavalīnāṭikā, Abhijñānaśākuntala, Mattavilāsaprahasana, Pādatāḍitaka, Bṛhatkathāślokasaṃgraha, Daśakumāracarita, Hitopadeśa, Kathāsaritsāgara, Pañcatantra, Śukasaptati, Tantrākhyāyika, Kubjikāmātātantra, Mātṛkābhedatantra, Mṛgendrāgama, Paraśurāma Kalpasūtra, Sārdhatriśatikālottarāgama, Ṭoḍalatantra,

Vinasikhatantra, Brahma Saṁhitā, Kṛṣṇāmṛṭamahārṇava, Sātvatatantra, Vaikhānasamantraprašna, Nāradasmṛti, Nyāyakusumāñjali, Śikṣasamuccaya, Yājnavalkyasmṛti, Atharvaveda, Gopatha Brāhmaṇa, Kauṣītaki Brāhmaṇa, Maitrāyaṇī Saṃhitā, Pañcaviṃśa Brāhmaṇa, Śatapatha Brāhmaṇa; Śiva Upaniṣad.

From Titus (titus.uni-frankfurt.de):

Aitareya Āraṇyaka; Aitareya Brāhmaṇa, Jaiminīya Brāhmaṇa, Kapiṣṭhala-Kaṭha-Saṃhitā, Taittirīya Brāhmaṇa, Taittirīya Samhitā; Bṛhadāraṇyaka Upaniṣad, Chāndogya Upaniṣad, Kauṣītaki Upaniṣad, Taittirīya Upaniṣad.

Courtesy of Matthias Ahlborn:

Abhiṣekanāṭaka, Avimāraka, Bālacarita, Cārudatta, Dūtaghaṭotkaca, Karṇabhāra, Mudrārākṣasa, Pañcarātra, Pratijñāyaugandharāyaṇa, Pratimānāṭaka, Svapnavāsavadatta, Ūrubhaṅga.

Bibliography

Alsina, A. (1992). “On the Argument Structure of Causatives”. *Linguistic Inquiry*, 23(4), 517-55.

Alsina, A; Joshi, S. (1991). “Parameters in Causative Construction”. Dobrin, L.; Nichols, L.; Rodriguez, R.M. (eds), *Proceedings from the 27th Regional Meeting of the Chicago Linguistic Society (CLS)*. Chicago: Chicago Linguistic Society, 1-16.

Andersen, P.K. (1986). “Die *ta*-Partizipialkonstruktion bei Aśoka: Passiv oder Ergativ?”. *Zeitschrift für vergleichende Sprachforschung*, 99, 75-94.

Baker, M.C. (1988). *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.

Bellégo, C.; Benatia, D.; Pape, L.D. (2021). “Dealing with Logs and Zeros in Regression Models”. *CREST – Série des Documents de Travail no. 2019-13*. http://dx.doi.org/10.2139/ssrn.3444996.

Börjars, K.; Vincent, N. (2008). “Objects and OBJ”. Butt, Holloway King 2008, 150-68.

Bresnan, J.; Moshi, L. (1990). “Object Asymmetries in Comparative Bantu Syntax”. *Linguistic Inquiry*, 21(2), 147-85.

Bubeník, V. (1987). “Passivized Causatives in Sanskrit and Prākrits”. *Linguistics*, 25, 687-704.

Bubeník, V. (1989). “On the Origins and Elimination of Ergativity in Indo-Aryan Languages”. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, 34(4), 377-98.

Bubeník, V. (2001). “On the Actualization of the Passive-to-Ergative Shift in Pre-Islamic India”. Henning, A. (ed.), *Actualization: Linguistic Change in Progress*. Amsterdam: Benjamins, 95-118.

Butt, M. (2001). “A Reexamination of the Accusative to Ergative Shift in Indo-Aryan”. Butt, M.; Holloway King, T. (eds), *Time over Matter: Diachronic Perspectives on Morphosyntax*. Stanford: CSLI Publications, 105-41.

Butt, M.; Deo, A. (2017). "Developments into and Out of Ergativity: Indo-Aryan Diachrony". Coon, J.; Massam, D.; Demena Travis, L. (eds), *The Oxford Handbook of Ergativity*. Oxford: Oxford University Press, 530-52.

Butt, M.; Holloway King, T. (eds) (2008). *Proceedings of the LFG08 Conference*. Stanford: CSLI Publications.

Bynon, T. (2005). "Evidential, Raised Possessor and the Historical Source of the Ergative Construction in Indo-Iranian". *Transactions of the Philological Society*, 103(1), 1-72.

Cardona, G. (1978). "Relations Between Causatives and Passives in Indo-Iranian". *Studies in the Linguistic Sciences*, 8(2), 1-42.

Çetinoğlu, Ö.; Butt, M. (2008). "Turkish Non-Canonical Objects". Butt, Holloway King 2008, 214-34.

Cole, P. (1983). "The Grammatical Role of the Causee in Universal Grammar". *International Journal of American Linguistics*, 49(2), 115-33.

Comrie, B. (1976). "The Syntax of Causative Constructions: Cross-Language Similarities and Divergences". Shibatani, M. (ed.), *The Grammar of Causative Constructions*. New York: Academic Press, 261-312.

Condoravdi, C.; Deo, A. (2014). "Aspect Shifts in Indo-Aryan and Trajectories of Semantic Change". Gianollo, C.; Jaeger, A.; Penka, D. (eds), *Language Change at the Syntax-Semantics Interface*. Berlin: Mouton de Gruyter, 261-92.

Criscuolo, R.M.; Overman, H.G.; Van Reenen, J. (2019). "Some Causal Effects of an Industrial Policy". *American Economic Review*, 109, 48-85.

Davies, W.; Rosen, C. (1988). "Unions as Multi-Predicate Clauses". *Language*, 64(1), 52-88.

Delbrück, B. (1888). *Altindische Syntax*. Halle: Verlag der Buchhandlung des Waisenhauses.

Edgerton, F. (1946). "Indic Causatives in *-āpayati* (*-āpeti*, *-āvei*)". *Language*, 22(2), 94-101.

Fillmore, C.J. (1968). "The Case for Case". Bach, E.; Harms, R.T. (eds), *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston, 1-88.

Gonda, J. (1951). *Remarks on the Sanskrit Passive*. Leiden: Brill.

Hock, H.H. (1981). "Sanskrit Causative Syntax: A Diachronic Study". *Studies in the Linguistic Sciences*, 11(2), 9-33.

Hock, H.H. (1982). "The Sanskrit Passive: Synchronic Behaviour and Diachronic Development". *South Asian Review*, 6, 127-37.

Hock, H.H. (1986). "'*P*-Oriented' Constructions in Sanskrit". Krishnamurti, B. (ed.), *South Asian Languages: Structure, Convergence and Diglossia*. Delhi: Motilal Banarsidass, 15-26.

Hock, H.H. (1987). "Reduced-Clause and Clause-Union Absolutives and Participles in Vedic Prose". Bashir, E.; Deshpande, M.M.; Hook, P.E. (eds), *Select Papers from SALA-7*. Indiana: Indiana University Linguistics Club, 182-97.

Hock, H.H. (1990). "Oblique Subjects in Sanskrit". Verma, M.K.; Mohanan, K.P. (eds), *Experiencer Subjects in South Asian Languages*. Stanford: CSLI Publications, 119-39.

Hock, H.H. (1991a). "Causees, Passive Agents, or Instruments? Instrumental NPs with Causatives in Early and Later Vedic Prose". Hock, H.H. (ed.), *Studies in Sanskrit Syntax: A Volume in Honor of the Centennial of Speijer's Sanskrit Syntax, 1886-1986*. Delhi: Motilal Banarsidass, 71-94.

Hock, H.H. (1991b). *Principles of Historical Linguistics*. 2nd ed. Berlin: Mouton de Gruyter.

Hook, P.E. (1991). "On Identifying the Conceptual Restructuring of Passive as Ergative in Indo-Aryan". Desphande, M.M.; Bhate, S. (eds), *Paninian Studies: Professor S.D. Joshi Felicitation Volume*. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan, 177-99.

Jamison, S.W. (1983). *Function and Form in the -áya-Formations of the Rig Veda and Atharva Veda*. Göttingen: Vandenhoeck & Ruprecht.

Jamison, S.W. (2000). "Lurching Towards Ergativity: Expressions of Agency in the Niya Documents". *Bulletin of the School of Oriental and African Studies*, 63(1), 64-80.

Jurafsky, D.; Martin, J.H. (2023). *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. Draft 3rd ed. https://web.stanford.edu/~jurafsky/slp3/.

Keenan, E.L.; Comrie, B. (1977). "Noun Phrase Accessibility and Universal Grammar". *Linguistic Inquiry*, 8(1), 63-99.

Kiparsky, P.; Staal, J.F. (1969). "Syntactic and Semantic Relations in Pāṇini". *Foundations of Language*, 5(1), 83-117.

Klaiman, M.H. (1978). "Arguments Against a Passive Origin of the IA Ergative". Farkas, D.; Jacobsen, W.M.; Todrys, K.W. (eds), *Papers from the Fourteenth Regional Meeting, Chicago Linguistic Society* (Chicago, 13-14 April 1978). Chicago: Chicago Linguistic Society, 204-16.

Kulikov, L. (2012a). *The Vedic -ya-Presents: Passives and Intransitivity in Old Indo-Aryan*. Amsterdam: Rodopi.

Kulikov, L. (2012b). "Vedic Preverbs as Markers of Valency-Changing Derivations". *Studies in Language*, 36(4), 721-46.

Kulikov, L. (2013a). "Constraints on the Causative Derivation in Early Vedic: Evidence for a Diachronic Typology of Transitivity". *Poznań Studies in Contemporary Linguistics*, 49(1), 79-101.

Kulikov, L. (2013b). "Language Vs. Grammatical Tradition in Ancient India: How Real Was Pāṇinian Sanskrit? Evidence from the History of Late Sanskrit Passives and Pseudo-Passives". *Folia Linguistica Historica*, 34, 59-91.

Lowe, J.J. (2017a). *Transitive Nouns and Adjectives: Evidence from Early Indo-Aryan*. Oxford: Oxford University Press.

Lowe, J.J. (2017b). "The Sanskrit (Pseudo)Periphrastic Future". *Transactions of the Philological Society*, 115(2), 263-94.

Osborne, J.W. (2002). "Notes on the Use of Data Transformations". *Practical Assessment, Research & Evaluation*, 8(6), 1-7.

Patel-Grosz, Pritty. (2021). "Ergativity in Indo-Aryan". Aronoff, M. (ed.), *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press.

Peterson, J.M. (1998). *Grammatical Relations in Pāli and the Emergence of Ergativity in Indo-Aryan*. München: LINCOM Europa.

Piantadosi, S.T. (2014). "Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions". *Psychonomic Bulletin & Review,* 21(5), 1112-30.

Pray, B.R. (1976). "From Passive to Ergative in Indo-Aryan". Verma, M.K. (ed.), *The Notion of Subject in South Asian Languages*. Madison, WI: South Asian Studies, University of Wisconsin-Madison, 195-211.

Renou, L. (1961). *Grammaire Sanscrite*. Paris: Adrien-Maisonneuve.

Söhnen-Thieme, R. (2019). "How to Translate a Gerund? The 'Gerund/Absolutive' in Epic-Puranic and Classical Sanskrit Literature". Deshpande, M.M.; Houben, J. (eds), *Proceedings of the 17th World Sanskrit Conference.*

Section 2: Linguistics (Vancouver, Canada, July 9-13, 2018). Vancouver: Department of Asian Studies, University of British Columbia.
Speyer, J.S. (1886). *Sanskrit Syntax*. Leiden: Brill.
Speyer, J.S. (1896). *Vedische und Sanskrit-Syntax.* Strassburg: K.J. Trübner.
Tichy, E. (1980). "Zum Kasusgebrauch bei Kausativa transitiver Verben". *Die Sprache*, 26, 1-18.
Tichy, E. (1993). "Transponierte Rollen und Ergänzungen beim vedischen Kausativ". Meiser, G. (Hrsg.), *Indogermanica et Italica: Festschrift für Helmut Rix zum 65. Geburtstag*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck, 436-60.
Whitney, W.D. (1885). *The Roots, Verb-Forms and Primary Derivatives of the Sanskrit Language*. Leipzig: Breitkopf & Härtel.
Whitney, W.D. (1896). *A Sanskrit Grammar: Including Both the Classical Language, and the Older Dialects, of Veda and Brahmana*. 3rd ed. Leipzig: Breitkopf & Härtel.

Appendix: Examples of Finite Passive Causatives

(1) han 'kill'
(a) PC-S: none.
(b) b) PC-O:

itarau	*tu*	*hasti-bala-kāmukau*
other.NOM.DU	but	elephant-strength-longing.NOM.DU
hastinā	*eva*	***ghātyetām***
elephant.INS	indeed	beat.CAUS.PASS.IMPV.3SG

'But **let** the other two, longing for the strength of elephants, **be beaten** by the elephant'. (*Mudrārākṣasam* 5.208)

(2) pac 'cook'
(a) PC-S: none.
(b) PC-O:

sa-nāma-homa-saṃpāta-ghṛte	***pācyatām***
with-name-oblation-residue-ghee.LOC	cook.CAUS.PASS.IMPV.3SG

'Let it (an effigy) **be fried** in the ghee residue of the named oblation'. (*Nārada Purāṇa* 1.90.113)

(3) grah ‘seize’
(a) PC-S:

taiḥ	*kila*	*asāv*	*ittham*	***agrāhyata***
they.INS.PL	apparently	he.NOM	as-follows	seize.IMPV.CAUS.PASS.3SG

‘Apparently he **has been persuaded** (lit. ‘caused to seize (their argument)’) by them as follows’. (*Daśakumāracarita* 2.4)

(b) PC-O:

yo	*niyogī*	*na*	*jānāti*	
which.NOM	official.NOM	not	know.3SG	
kṛtsna-āyāsena		*saṃyuktaḥ*	*sa*	
complete-fatigue.INS.SG		linked.NOM.M	this.NOM.M	
nṛpa-bhaktān		*varān*		*kṣitau*
king-worshippers.ACC.PL		best.ACC.PL		earth.LOC
tair	***nigrāhyate***		*punaḥ*	
they.INS.PL	seize.CAUS.PASS.3SG		again	

‘The official who does not know the best king-worshippers on earth, he is **held back** (lit. ‘caused to be seized’) by them again, having been befallen by complete fatigue’. (*Nārada Purāṇa* 2.6.11)

(4) dā ‘give’
(a) PC-S:

tena	*deva*	*yadi*	*nyāyyaṃ*	*pitṛ-draviṇam*		*āvayoḥ*
thus	king.VOC	if	justice.NOM	paternal-wealth.ACC		us.GEN.DU
bhrātṛjāyā		*tataḥ*	*sā*	*nau*	*vyutthitā*	***dāpyatām***
sister-in-law.NOM.F		thus	this.NOM.F	us.DAT.DU	swerving NOM.F	give.CAUS.PASS.3SG

‘Thus, your majesty, if there is justice let the wayward sister-in-law **be caused to give** the patrimony to both of us’. (*Bṛhatkathāślokasaṃgraha* 4.25)

(b) PC-O:

atha	*tat-siddhi-lubdhatvād*		*avocaṃ*
then	this-success-desire.ABL		say.AOR.1SG
upadeśo	*mama*	*apy*	*eṣa*
advice.NOM.M	I.GEN	also	this.NOM.M
tāḥ	*sakhīr*	*aham*	
they.ACC.F.PL	friends.ACC.F.PL	I.NOM	
yuṣmābhir	***dāpyatām***		
you.INS.PL	give.CAUS.PASS.IMP.3SG		

‘Then I said, desiring to succeed, to my dear friends: **let** this my advice also **be** (caused to be) **given** by you’. (*Kathāsaritsāgara* 3.6.106)

(5) kṛ ‘do’
(a) PC-S:

candra-āsannair	*hi*	*nakṣatrair*
moon-in.conjunction.INS.PL	indeed	star.INS.PL
lokaḥ	*kāryāṇi*	***kāryate***
world.NOM	duty.ACC.PL	do.CAUS.PASS.3SG

‘People **are** (lit. the world is) **caused to do** their duties by the constellations in conjunction with the moon’. (*Bṛhatkathāślokasaṃgraha* 15.6)

(b) PC-O:

vegavatyā	*tataḥ*	*saha*
V.INS	then	with
naravāhanadattasya	*vivāhaḥ*	***kāryatām***
N.GEN	marriage.NOM	do.CAUS.PASS.IMPV.3SG

‘Then **let** the marriage of Naravāhanadatta with Vegavatī **be caused to be carried out**’. (*Bṛhatkathāślokasaṃgraha* 15.13)

(6) hṛ ‘carry’
(a) PC-S:

kaccid	*abhyāgatā*	*dūrād*	*vaṇijo*	*lābha-kāraṇāt*
perhaps	arrived.NOM.PL	far.ABL	merchant.NOM.PL	gain-purpose.ABL
yathā-uktam	***avahāryante***	*śulkaṃ*	*śulka-upajīvibhiḥ*	
as-stated	put-down.CAUS.PASS.3PL	tax.ACC	tax-subsister.INS.PL	

‘Perhaps the merchants who come from afar for gain **are caused to put down** (i.e. pay) the stated tax by the tax collectors’. (*Mahābhārata* 2.5.103)

(b) PC-O:

dhana-hīnena	*deho*	*’pi*	***hāryate***
wealth-deprived.INS	body.NOM	even	take.CAUS.PASS.3SG

‘Even the body **is caused to be taken (away)** by/from the man deprived of wealth’. (*Kathāsaritsāgara* 3.5.28)

(7) jñā ‘know’
(a) PC-S:

tena	*hi*	*mad-vacanād*	***vijñāpyatām***	*upādhyāyaḥ*	*somarātaḥ*
thus	indeed	my-command.ABL	know.CAUS.PASS.IMPV.3SG	teacher.NOM	S.NOM

‘Thus indeed let the teacher Somarāta **be informed (lit. caused to know)** at my command’. (*Abhijñānaśākuntala* 5.5)

(b) PC-O:

yad-arthaṃ *vayam* *āhūtās* *tat* ***samājñāpyatām***
why we.NOM.PL summoned.NOM.PL that.ACC know.CAUS.PASS.IMPV.3SG
'Let it **be made known** why we have been called here'.
(*Bṛhatkathāślokasaṃgraha* 8.35)

(8) śru 'hear'
(a) PC-S:

pālakaḥ ***śrāvyatāṃ*** *sūnor* *vṛttāntam*
prince.NOM hear.CAUS.PASS.IMPV.3SG son.GEN news.ACC
'Let the prince **be caused to hear** the news of his son'.
(*Bṛhatkathāślokasaṃgraha* 3.79)

(b) PC-O:

tāta *ghoṣavatī-ghoṣa-saṃgītaṃ* ***śrāvyatām***
father.VOC lute-sound-concert.NOM hear.CAUS.PASS.IMPV.3SG
'Father, **let** the concert of lutes **be caused to be heard**'.
(*Bṛhatkathāślokasaṃgraha* 5.144)

(9) paṭh 'recite'
(a) PC-S:

na *vyāpāra-śatena* *api* *śukavat* ***pāṭhyate*** *bakaḥ*
not action-hundred.INS even parrot-like recite.CAUS.PASS.3SG heron.NOM
'The heron **is not caused (i.e. taught) to recite** like a parrot (is), even if one tries a hundred times'. (*Hitopadeśa* 0.42)

(b) PC-O: none.

(10) dṛś 'see'
(a) PC-S: none.
(b) PC-O:

Padmāvatī *tato* *devī* ***darśyate*** *ca* *acirāt*
P.NOM then queen.NOM see.CAUS.PASS.3SG and soon
'Then queen Padmāvatī **is** also soon **revealed**'. (*Kathāsaritsāgara* 3.1.117)

e-ISSN 2785-5953

Bhasha
Vol. 3 – Num. 1 – April 2024

Final Syllable Reduction in Middle Indic and Iranic

Andrew Ollett
University of Chicago, USA

Abstract The mutual influence of Indic and Iranic languages in the 'contact zone' of the upper Indus valley and the adjoining mountain ranges has long been known to scholarship. Previous scholarship has observed lexical borrowings in the ancient period, and phonological and morphological features in the modern period. Outside of this area, the Indic and Iranic languages developed in different ways, largely in accordance with the prominence of the syllable and the word, respectively, in their phonologies. The important Middle Indic languages of Gandhari and Apabhramsha, however, appear to have participated in certain structural changes that affected all of the contemporary Middle Iranic languages, above all the reduction of final syllables. The fact that the earliest attestation of these changes in Middle Indic occurs in languages strongly associated with the Indo-Iranian contact zone suggests that contact may have played a critical role. The chronology of these changes in Middle Indic also suggests that they occurred in the 'Saka-Kuṣāṇa Age' in the early centuries of the common era.

Keywords Middle Indic. Gandhari. Apabhramsha. Middle Iranic. Historical phonology. Language contact.

Summary 1 Introduction. – 2 The Long History of Indo-Iranian Contact. – 3 Key Isoglosses between Middle Indic and Middle Iranic. – 4 Overview of the Data. – 4.1 Sogdian and the Rhythmic Law. – 4.2 Middle Persian. – 4.3 Bactrian. – 4.4 Khotanese. – 4.5 Gandhari. – 4.6 Apabhramsha. – 5 Analysis. – 6 The Development of Gandhari. – 7 The Origins of Apabhramsha. – 8 Lexical Evidence. – 8.1 Words Related to Writing. – 8.2 Words Related to Governance. – 8.2.1 *kṣatrapa-* 'governor'. – 8.2.2 *bhaṭāraka-* 'lord'. – 8.2.3 Words Ending in *-pati*. – 8.3 Military Terms. – 8.4 Names. – 9 Conclusions.

Edizioni Ca'Foscari

Peer review

Submitted 2023-12-14
Accepted 2024-04-22
Published 2024-05-24

Open access

Citation Ollett, A. (2024). "Final Syllable Reduction in Middle Indic and Iranic". *Bhasha*, 3(1), 53-106.

DOI 10.30687/bhasha/2785-5953/2024/01/004

1 Introduction

This paper makes three connected arguments. First, it identifies a sound change that affected all of the Middle Iranic languages and some of the Middle Indic languages, which I call Final Syllable Reduction (FSR).[1] FSR itself is a well-known feature of many Middle Iranic languages and of the Middle Indic languages Gandhari and Apabhramsha, but with one exception, scholars have not attempted to bring these changes under a single unified analysis. Second, it argues that the application of FSR to the Middle Indic languages in question, Gandhari and Apabhramsha, is due to contact with speakers of Iranic languages. Scholars have previously taken these changes to be internal developments within the Indic language family. Given the long history of contact between Indic and Iranic languages in the 'contact zone' where Gandhari and Apabhramsha appear to have originated, a finer-grained account of the sound changes in question is desirable for evaluating the plausibility of contact as an explanation. In that connection, the third argument of this paper is that FSR in Gandhari is due to its use as a *lingua franca*, starting in the first century BCE, by speakers of Iranic languages, and that similar developments explain the characteristic features of Apabhramsha. The long-term impact of these changes was profound, since these features of Apabhramsha subsequently spread to most of the North Indian languages, and resulted in the North Indian languages having phonological systems that differed greatly from those of their ancestors (Common Middle Indic) and their neighbours (the Dravidian languages).

2 The Long History of Indo-Iranian Contact

The contact between representatives of the Indic and the Iranic language families is full of both historical and theoretical significance. To start with its theoretical significance, it has long been recognised that the contact zone between these families forms part of larger 'linguistic areas'. A linguistic area is a zone in which certain linguistic features are found across the boundaries of language 'families' as constituted by historical linguistics. Retroflex consonants, for example, are found throughout the South Asian linguistic area, in the Indic, Iranic,

1 I follow Kümmel (2018) in using the term 'Iranic' to refer to the language family in distinction to 'Iranian' (used in a wide variety of other contexts). Similarly, I use 'Indic' to refer to the language family that other scholars call 'Indo-Aryan', especially since it is extremely difficulty to pry the word 'Aryan' away from the racial (and racist) ideas with which it has long been associated.

Bhasha
3, 1, 2024, 53-106
e-ISSN 2785-5953

Dravidian, and Munda language families.[2] The boundary between the Indic and Iranic language families is ostensibly the Indus river, with Iranic languages (Balochi and Pashto) on the western side and Indic languages (Sindhi, Seraiki, and Punjabi) on the eastern side, terminating in a dense tangle of languages where the Indus descends from the Hindu Kush. Yet many phonological, morphological, and syntactic features integrate these languages into a larger South Asian linguistic area.

The Indic and Iranic languages are, of course, related by common descent from Proto-Indo-Iranian. In contrast to two clearly unrelated languages, in the case of neighbouring Indic and Iranic languages, it can be difficult to determine whether a common linguistic feature has been maintained by both from the protolanguage or transmitted from one daughter language to another (cf. Morgenstierne 1975). A term that may prove helpful for language in the contact zone is 'linkage', which François (2015, 170-1) defines as "a network of dialects which remained in contact with each other for an extended period of time".

Linkages are actually predicted by the traditional 'family tree' model of language relationships, given two premises. The first is that the speakers of the languages in question stay more or less in one place – in other words, that the increasing distance between languages is due to the gradual but differential accumulation of sound changes by different social groups, rather than a sudden dislocation of one group of speakers (due, for example, to migration). A second and related premise is that the languages, and social groups, that have separated in this way can and often do remain in contact. It is a well-known weakness of the 'family tree' model that the situation of contact that it predicts as the general rule cannot be represented in the model itself. This is one of the reasons why the family tree model has fallen out of favour for accounting for and representing the internal differentiation of language families (cf. François 2015; Korn 2019).

Whatever model of language change we ultimately adopt, the Indo-Iranian contact zone is characterised by 'differentiation plus contact', i.e. a situation of contact between genetically related languages. At the time of our oldest historical documentation of the Iranic and Indic languages, they were already very similar to each other: "There must, in the sixth and fifth centuries B.C., have been hundreds of the most commonly used words which were practically identical on both sides of the linguistic border" (Morgenstierne 1974, 271; cf. also Chatterji 1960, 126). As these languages gradually diverged from each other due to sound changes, contact phenomena may have brought them back together.

2 Cf. Emeneau 1956; 1980 (focused on phonological features) and Masica 1976 (focused on syntax) for the South Asian linguistic area; Tikkanen 2008 focuses specifically on the 'transit zone' between South Asia and Central Asia.

One major obstacle to the study of language contact over the long term is the lack of documentation. In the case of Indo-Iranian, we are exceptionally lucky to have representatives of the language families in question attested for thousands of years. But even so there are long breaks in the historical record. Even if Yaghnobi might continue Sogdian in some sense, or Khowar might continue Gandhari, there is nevertheless a gap of many centuries. Sanskrit has played an important role in the reconstruction of Indo-Iranian contact in part because it provides a much greater quantity of evidence, thanks to its continuous use over a wide geographic area for thousands of years; but for precisely the same reason, this evidence needs to be used with care, and the conservatism of Sanskrit has meant that the only types of contact phenomena visible from this perspective are loanwords (cf. Chatterji 1960; Morgenstierne 1974). As a somewhat crude generalisation, we may say that scholars have approached Indo-Iranian contact using two very different sets of evidence: the modern languages (cf. Bashir 2016), each of which presents, synchronically, the outcome of various contact phenomena; and the ancient languages, which could in principle attest to the before-and-after stages of contact phenomena, but which have nevertheless been treated in much the same way as the modern languages, viz. as a 'snapshot' of contact phenomena that happens to have been taken earlier in time.

Ironically, although the modern evidence provides strong evidence for contact-induced sound changes, in the form of phonological isoglosses between neighbouring languages, I am not aware of any scholarship that attributes sound changes in the ancient languages to contact. That is due partly to the 'lexical bias' of the Sanskrit evidence, noted above, but also to the fundamental impossibility of ascertaining the causes of sound change outside of relatively broad parameters; the areal isoglosses that make contact a possible or likely explanation are often precisely what is lacking in the case of ancient languages.

Let us now turn to the historical significance of Indo-Iranian contact. Anyone who wants to travel between South Asia and the west by land must go through the Indo-Iranian contact zone. This region, and especially the mountain passes in the Karakorum range, have for this reason been considered one of the 'crossroads of Asia'. It has been the primary corridor within which new peoples, languages, cultures, religions, and technologies have entered South Asia, and through which South Asia, in turn, has been integrated into other Asian political and cultural formations.

The mountain ranges of Eastern Afghanistan, Northern Pakistan, and Northwest India are known for their linguistic diversity as well as the density and complexity of language contact. Bashir (2016) calls the region PHKKK (Pamir-Hindukush-Karakorum-Kohistan-Kashmir). Two anthropologists, Augusto and Alberto Cacopardo, have

argued that the same region, which they call Peristan, is culturally distinct from its lowland neighbours (Cacopardo 2016). In Zoller's interpretation, "these differences must be explained for the greater part in terms of earlier and later Indo-Aryan immigrations" (2018, 178; cf. also 248). Zoller appears to favour the hypothesis that the Indic languages in the PHKKK range represent an "earlier" wave of migration of Indic speakers that maintained its cultural identity in the face of subsequent waves of migration.

We turn now from prehistory to history. Gandhāra – the region corresponding to modern Peshawar and the adjoining valleys – was added to the Achaemenid Empire during the reign of Darius I (r. 522-486 BCE). The Achaemenid kings ruled the region through satraps until Alexander's victory over Darius III in 331 BCE. It was probably during the latter part of this period that the Sanskrit grammarian Pāṇini produced his *Aṣṭādhyāyī* (Eight Chapters), in Śālātura, now Chota Lahor in northern Pakistan. Sanskritists can easily forget that Pāṇini was a subject of a vast empire that stretched to the Mediterranean Sea. The representatives of the Achaemenid state in Gandhāra – satraps, administrators, bureaucrats, engineers, and soldiers – probably reflected the multiethnic and multilingual composition of the empire, and likely included speakers of Iranic languages (Persian, Median, etc.). It is to this "first period" of Indo-Iranian contact (according to Chatterji 1960) that we can trace the earliest secure examples of borrowings from one language into another, for example *dipi-*/*lipi-* for 'writing' (see below). In fact the technology of writing itself was one of the borrowings: the Kharōṣṭhī script, used in the former Achaemenid provinces of the Northwest, was based on the Achaemenid Aramaic script.

After Alexander, the Northwest came under the control of the Mauryas, and after the Mauryas, the Greeks. It is notable that Iranians continued to play a role in the administration of the Maurya and Indo-Greek states during this period: one *yavanarāja* (Western King) Tuṣāspha, evidently with the Iranic onomastic element *-aspa*, is credited with restoring a tank during the reign of Aśōka.[3]

Iranic-speaking peoples returned to the centre stage of political history in the Northwest starting in the first century BCE, which starts the "second period" of borrowing, according to Chatterji (1960). Groups began to move into South Asia from across the mountains, driven in part by large-scale population movements in Central Asia and the Iranian plateau. The most important group of newcomers were the Sakas, who spoke Eastern Iranian languages and who had lived in the plains of Central Asia. The Sakas under Maues (Moga) took power in the Punjab in around 80 BCE. Maues' successors,

3 Junāgaṛh inscription of Rudradāman, l. 8 (Kielhorn 1905-06).

including Azes (Aya), held the area until the Parthian ruler Gondophares took it over around 20 CE. By this time, however, Saka rulers had established themselves further east, around Mathura, and another group of Sakas (known by the family name Kṣaharāta) had established themselves in coastal Gujarat. The Mathura-based Sakas were absorbed into the Kuṣāṇa empire in the later first century CE, and around the same time, the Kṣaharātas of Gujarat were succeeded by another Saka family, the Kārdamakas. The Kārdamakas would continue to rule in Gujarat and Western Madhya Pradesh until their defeat by the Guptas in the later fourth century.

The reign of the Sakas and especially the Kuṣāṇas represents a period in which Northern India was not just ruled by Iranic-speaking peoples but closely connected to the predominantly Iranic-speaking regions to the east. The Sakas must have spoken an Eastern Iranic language, closely related to Khotanese and Tumshuqese, which are attested later; all of their coins and inscriptions, however, made use of other languages (Greek, Gandhari, and Sanskrit, in that order). The ethnicity and language of the early Kuṣāṇas is still under discussion, but it is clear that Kaniṣka (r. 126-150 CE) adopted Bactrian, an Eastern Iranic language, for public inscriptions and coinage. During the 'Saka-Kuṣāṇa Age', we have dozens of inscriptional attestations of Iranic names in Gandhari and Sanskrit, especially in the Indo-Iranian contact zone. The borrowing of political and military terminology from Iranic into Indic languages continued during this period as well (see below).

Even after the disappearance of Iranic-speaking dynasties from the plains of North India in the third century, considerable Iranic influence can be seen in the colophons of the Sanskrit manuscripts produced during the reign of the Palola Ṣāhis in Gilgit (von Hinüber 1980).

The Saka-Kuṣāṇa Age has a historical significance beyond its role, in this article, as the backdrop for contact between Indic and Iranic languages. Max Müller is only the most prominent of a number of scholars who posited a radical break in Indian literary and cultural history around the turn of the common era. Müller's chronology was quite different from the one that scholars generally accept today, but if we update his argument, he would state that there was a properly 'ancient' period that extended until about the fourth or third century BCE, which includes the entirety of the Vedas as well as some version of the canonical texts of the Buddhist and the Jains, and a 'revivalist' period that starts from the second century CE or so, which includes classical Sanskrit literature (Aśvaghōṣa, Kālidāsa, Bhāravi, etc.). Müller himself attributed the break to a "Turanian invasion" (1883, 85).

Müller's theory was embedded in an orientalist narrative, with its fetishisation of the very distant past (a time of advanced humanity which has been on the decline ever since) and its racialisation of

culture. These tropes must be resisted and refuted with vigilance. But the idea of an important articulation in the linguistic and cultural history of South Asia, falling within a century or two of the year zero, continues to find acceptance, explicitly or implicitly, even if Müller's name is hardly ever associated with it. The distinction between 'Vedic' and 'Classical' Sanskrit is one instance; we will see below that it articulated Mayrhofer's etymological dictionary of Sanskrit into two parts. And the evidence for migration into the subcontinent from Central Asia on the part of groups who identified as Sakas (Müller's "Turanians") between the first century BCE and the second century CE has only accumulated since then.

Some version of Müller's hypothesis may help to account for two important phenomena of which Müller himself had almost no knowledge whatsoever. One of these is the literature in the language that Bailey called "Gandhari", now known to be represented by hundreds of inscriptions, wooden and leather documents, and birch-bark manuscripts dating from the first century BCE to about the third century CE (Bailey 1946; Salomon 1999b). This literature was only discovered in the early twentieth century, with a large cache of additional materials coming to light in the final decade of that century. Another is the literary language called Apabhramsha (literally 'degradation') in premodern India, which was used mostly from the eight to the twelfth century CE. This language was known from references in Sanskrit texts already in Müller's time, but actual literature in the language was not published until the turn of the twentieth century. What Müller called the "Turanian invasion", which we might call the 'Saka-Kuṣāṇa Age', is of decisive importance to the development of both languages. Its importance to the development of Classical Sanskrit literature was already outlined, albeit controversially, by Lévi (1903), who similarly had no knowledge of Gandhari, and little knowledge of Apabhramsha, at the time.

3 Key Isoglosses between Middle Indic and Middle Iranic

A sound change that I will call Final Syllable Reduction (FSR) affected all of the Middle Iranic languages and two Middle Indic languages originating in the Indo-Iranian contact zone. FSR is a two-stage process. First, the quantity of the final syllable is reduced, such that all heavy final syllables become light. Secondly, the quality of the final syllable is lost. This might mean that they were pronounced similarly or identically, or it might mean that they were not pronounced at all. This second stage might itself have unfolded in two separate moments, i.e. the loss of final vowel quality followed by the loss of final vowels, or the loss of final vowels might have followed directly from the loss of final vowel quantity. The evidence in general does

not allow us to decide between these two steps or scenarios; what we observe, instead, in languages wherein final vowels are written, is the systematic confusion of forms of a word that historically had different final vowels.

The oldest-attested languages of the Indo-Iranian language family, such as Vedic Sanskrit, Avestan, and Old Persian, were all highly inflected. Nouns, adjectives, pronouns, and verbs were inflected according to their grammatical category, and most of these inflections were found at the end of the word. The final syllable of a word therefore contained indispensable information about its part of speech and grammatical category. Hence, although FSR is a 'mere' sound change, by targeting the final syllable, it resulted in the collapse and partial reconstruction of the inherited system of inflection. FSR is therefore implicated in a wide range of structural changes in the Middle Indic and Middle Iranic languages, including the loss of distinctions in gender, number, and case.

One important corollary to FSR is the development of 'extended' endings by attaching a pleonastic suffix (deriving from Proto-Indo-Iranian *-*ka*-) to nominal forms, i.e. nouns, adjectives, and pronouns. Although the conditions for the addition of this suffix differ from language to language, all the languages discussed in this paper ended up with secondary inflections.

A secondary feature of Middle Indo-Iranian that I identify here, although much more tentatively than FSR, is the fricativisation of consonants. Already in Proto-Iranian, a number of inherited stops had taken on a fricative pronunciation, specifically: (1) voiceless aspirated stops in all positions (/k^h/, /t^h/, /p^h/ → /x/, /θ/, /f/), and (2) voiceless unaspirated stops before other consonants (/k/, /t/, /p/ → /x/, /θ/, /f/ / _/C/). In the analysis below I connect both FSR and fricativisation to the shift from 'syllable' phonology to 'word' phonology in the relevant languages, which has been perceptively noted already by Kümmel (2014).

4 Overview of the Data

The languages that I argue were affected by the changes outlined above include:[4]

- Middle Iranic:
  - 'Eastern':
    - Sogdian: attested from the fourth to about the tenth century, in Sogdia (the area between the Zeravshan and Oxus rivers) and generally along the Silk Road. Written in a consonantal script derived from Aramaic, with a few documents also in the Brāhmī script.
    - Bactrian: attested in coins and inscriptions of the Kushana period (late first to early third century), and in leather documents up until the ninth century, in Bactria (northern Afghanistan) and neighbouring areas. Written in the Greek script.
    - Khotanese: attested in Khotan, in modern Xinjiang UAR, from the fourth through the eight century. Written in the Brāhmī script.
  - 'Western':
    - Middle Persian: the language of southwestern Iran, used as a literary language by Zoroastrians, Christians, and Manichaeans, attested in documents from Iran and Turfan (in Xinjiang UAR), from roughly the third century; I do not separately discuss Parthian, a 'Western' Iranic language that is in many ways similar to Middle Persian. Both are written in consonantal scripts derived from Aramaic.
- Middle Indic:
  - Gandhari: attested in the 'greater Gandhara' region (around modern Peshawar) from the first century BCE to the late third century CE, with some earlier inscriptions written in the same script (Kharōṣṭhī), as well as a collection of third century documents from Niya, in the Tarim basin.
  - Apabhramsha: a literary language first associated with certain communities of northwestern India (Gujjars and Ahirs), attested from the eight century onwards.

This list includes all of the attested Middle Iranic languages, but only two of the Middle Indic languages. In other words, the features that I take to be diagnostic of contact between the Middle Indic and Middle Iranic languages are general features of Middle Iranic, but their presence in Middle Indic is exceptional and requires explanation. None

4 Cf. Kümmel 2018; Korn 2019 for the conventional but controversial distinction between 'Western' and 'Eastern' Iranic languages.

of the other attested Middle Indic languages, including Prakrit, Ardhamagadhi, Pali, or Epigraphic Middle Indic exhibit either FSR or fricativisation. (Of course, final syllables in all these languages were reduced in the case that they originally consisted of superheavy syllables, but all superheavy syllables, not just those at the end of a word, were eliminated in Middle Indic; indeed this is one of the defining characteristics of Middle Indic). The loss of single intervocalic consonants in these languages, however, is sometimes said to have passed through a fricative phase; this will be discussed below in reference to Gandhari.

In the conclusion I suggest that the changes associated with FSR are also represented in modern languages. I am not a specialist of the modern Indic or Iranic languages, so I cannot speak with any authority about the range of languages affected by these changes. I will note, however, that most of the Indic languages of North India, including Hindi, Gujarati, Marathi, and Bengali, exhibit the loss of final vowels, and show clear evidence for FSR in their early records. Hence it appears that the changes that were generalised within the Iranic family of languages already by the beginning of the common era came to be generalised within the Indic family, too, albeit centuries later.

4.1 Sogdian and the Rhythmic Law

FSR is corroborated by the following features of Sogdian:

- Inherited final vowels are shortened. This is a postulate, because the quantity of vowels is not marked in Sogdian writing. I note, however, that all instances of final vowels can be interpreted as short, and this accords with the following generalisation about inherited final consonants.
- Inherited final consonants are lost.
  - The sequence */am#/ becomes /u#/ (Gershevitch 1954, §§ 349, 1171, 1194).
  - The sequence */a:m#/ becomes /u#/ (Gershevitch 1954, §§ 1173, 1192) or /a#/ (Gershevitch 1954, § 1175).
  - The sequence */i:m#/ becomes /i#/ (Gershevitch 1954, §§ 350. 1174, 1197).
  - The sequence */ah#/ becomes /i#/ (Gershevitch 1954, §§ 402, 1168, 1191).
  - The sequence */a:t#/ becomes /a#/ (Gershevitch 1954, § 1179).
  - The sequence */a:h#/ becomes /a#/ (Gershevitch 1954, § 1185).

An important exception to the above rules is monosyllables like /xo:/ ‘he’ and /xa:/ ‘she’.

The foregoing sound changes operate throughout Sogdian. Their results serve as input to another set of sound changes which,

however, are limited to certain types of words. I refer to the well-known 'Rhythmic Law' of Sogdian. Paul Tedesco was the first to notice that the inflectional endings of a word differ according to some prosodic feature of their base. Some bases are 'light', and some are 'heavy'. The endings of 'heavy' bases are reduced or elided versions of the endings of 'light' bases. See table 1 for an example.

Table 1 Examples of the Sogdian Rhythmic Law

	Light stem	**Heavy stem**
written	⟨βγy⟩	⟨β'γ⟩
pronounced	/βaɣi/	/βaːɣ/
category	nom.sg.	nom.sg.
meaning	"god"	"garden"

The nominative masculine singular ending /-i/ has been deleted after the heavy base /βaːɣ-∅/ but not after the light base /βaɣ-i/. This difference has been accounted for, by Gershevitch and Sims-Williams, by formulating a rule that deleted final unstressed vowels. This rule should target the final syllable of */βaːɣ-i/ but not of */βaɣ-i/. Hence the final vowel of the former should be unstressed (*/'βaːɣ-i/), whereas the final vowel of the latter should be stressed (*/βaɣ-'i/).

Sims-Williams explained the assignment of stress in the following terms:[5]

- stress a syllable containing a long vowel;
- otherwise, stress the final syllable.

Hence in Sims-Williams' account what makes a stem heavy is simply the occurrence of a long vowel anywhere within the stem. (According to him, a following consonant in the same syllable – which, in many languages, make the syllable prosodically heavy – does not make a stem heavy for the purposes of the Rhythmic Law, except in those cases where a consonant, such as 'ṁ' or 'r', can be interpreted as a nasalised or rhotacised vowel segment. I am not in a position to disagree with Sims-Williams about the patterns found in the language, but I do find the phonetic characterisation of preconsonantal 'ṁ' and 'r' to be idiosyncratic). Sims-Williams later (1996, 312) noted a tendency to omit "short vowels in initial open syllables" when Sogdian is written in the Brāhmī script (e.g. *mdhu* for [məðu]), which he took to support his idea that final syllables are stressed as a rule.

5 "[A] heavy syllable may therefore be defined very simply as a syllable which contains a long vowel or diphthong" and "it should be possible to define a heavy stem as one which contains at least one heavy syllable" (Sims-Williams 1984, 213).

The stress rule allegedly underlying the Rhythmic Law is unique to Sogdian. In fact, according to Novák (2013, 74), it did not operate in Yaghnobi, Sogdian's closest ancestor, at all. Novák, who accepts the Gershevitch-Sims-Williams account of the Rhythmic Law, argues for four distinct types of stress pattern in Sogdian:

- Stress I: the Proto-Iranian (and Proto-Indo-Iranian) 'free stress' system, where the location of prominence within a word is determined by lexical and morphological factors.
- Stress II: stress a heavy penultimate syllable, or if the penultimate is light, the antepenultimate syllable. Novák (2013) dates this shift prior to the divergence of Sogdian and Yaghnobi, hence it operated at least in the 'Proto-Sogdic' period.
- Stress III: the Gershevitch-Sims-Williams system described above, i.e. stress a syllable containing a long vowel, or if there is none, stress the final syllable. Novák notes that this system operated only in Sogdian, and not in Yaghnobi.
- Stress IV: stress the final syllable.

In the following I will propose a reanalysis of the Rhythmic Law that completely dispenses with Stress III. In this analysis, the Rhythmic Law of Sogdian does not presuppose any change in stress vis-à-vis other Middle Iranic languages. Rather, I consider it to be a deletion rule that operated in Sogdian but presupposes the same stress pattern shared by most of the other Middle Iranic languages, namely, Novák's Stress II. There are four motivations for this analysis.

First, Stress III is posited only in order to explain the Sogdian Rhythmic Law. There is, to my knowledge, no other evidence for it, and in fact the evidence for Sogdian written in Brāhmī (Sims-Williams 1996) supports Stress IV rather than Stress III. As Novák's summary shows, Stress III was not a feature of Proto-Sogdic, either. I also know of no other Middle Iranic languages that exhibited such a stress pattern, although possible parallels to the Rhythmic Law in Middle Persian will be discussed below.

Second, the phonological basis of this stress pattern remains unclear. Phonologists now tend to see stress as a form of prominence assigned cumulatively from lower to higher levels of phonological structure (cf. e.g. Nespor, Vogel 1986). In Hayes' (1995) influential account, stress patterns arise from the parsing of phonological units (syllables or moras) into prosodic feet, which are then grouped and assigned prominence according to language-specific parameters. Hayes has argued that all attested stress patterns can be accounted for with an inventory of only three types of prosodic feet: syllabic trochees, moraic trochees, and iambs. In order to motivate Stress III, we must specify how Sogdian parses its syllables into prosodic feet. Once we do so, however, we no longer need Stress III to account for the Rhythmic Law, as shown below.

Third, the traditional analysis of the Rhythmic Law understands stress assignment to work primarily at the level of the stem, i.e. prior to the addition of inflectional endings; hence there are 'heavy' stems and 'light' stems. But this analysis is complicated by the fact that some inflectional endings can make a stem heavy for the purposes of the Rhythmic Law, as we will see below. The fact that the same stem can be either light or heavy (for the purposes of the Rhythmic Law) depending on the ending strongly suggests that the Rhythmic Law is not sensitive to the vowel quantities within a stem but rather the syllable quantities within a prosodic word as a whole.

Fourth, the Rhythmic Law can be seen as a manifestation of a more general phenomenon, namely FSR, which cuts across the distinction between light and heavy stems in Sogdian, and which is widely attested among the neighbouring languages. Postulating Stress III to account for the deletion of final vowels in heavy stems leaves other aspects of this phenomenon unexplained. In my account, the Sogdian Rhythmic Law (and subsequently Stress IV) operates on the prosodic structures produced by these changes.

I propose that Sogdian parses its syllables into moraic trochees, and the resulting structures account for various facets of Sogdian grammar, including the loss of final vowels in heavy stems. A moraic trochee is a foot consisting of two moras with prominence on the first mora (Hayes 1995, 71):

Figure 1 The moraic trochee (foot and moras)

Hence either two light syllables or a single heavy syllable can be parsed into a moraic trochee:

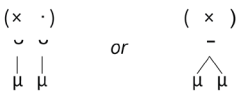

Figure 2 The moraic trochee (foot, syllables, and moras)

Let us assume that Sogdian parses syllables into moraic trochees from right to left. (This analysis may work for left-to-right parsing, but the syllable structure of longer words is not always clear in Sogdian, so I will confine my discussion to the last two to three syllables). Degenerate feet, in this case single light syllables, are not allowed.

Hence we have two parses of the Sogdian words above prior to the operation of the Rhythmic Law:

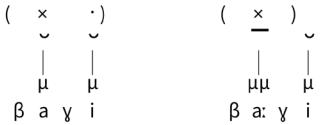

Figure 3 Parsing of /βaɣi/ and */βaːɣ/ into moraic trochees

The Rhythmic Law could thus be rephrased simply as: delete a vowel in an unfooted syllable. No shift in word-level accent is necessary. I leave open whether this applies to all unfooted vowels, or just those in final position. In the former case, we may have an explanation of syncope in Sogdian (the extent of which is difficult to determine because Sogdian is written without vowels) in addition to an explanation of the Rhythmic Law. As noted above, 'ambivalent' stems, which count as either light or heavy depending on the inflectional ending, pose a problem for the traditional analysis of the Rhythmic Law. But they are easily accommodated in this account. Let us start with nominal forms. All of the inflectional endings of the singular begin with vowels, and would not make the preceding syllable heavy. But the endings of the plural all begin with /t/, and hence can make a preceding syllable heavy. One example is /rur/ ⟨rwr⟩ 'plant', which was originally light in the singular (e.g. nom.-acc.sg. /rura/ ⟨rwr'⟩) and heavy in the plural (e.g. nom.-acc.pl. /rurt/ ⟨rwrt⟩ , as opposed to the expected light ending */rurta/ * ⟨rwrt'⟩). In this analysis, the heavy form results from the deletion of the final vowel, which is now in an unfooted syllable. (The light forms are sometimes analogically restored).

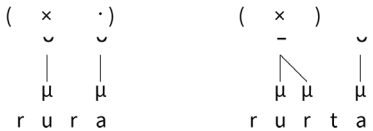

Figure 4 Parsing of /rura/ and */rurta/ into moraic trochees

One might expect all stems ending in a consonant to pattern this way, if a following conjunct consonant really makes the preceding syllable heavy. But that is not the case: light stems ending in 'n' and 'm', like /un/ ⟨wn⟩ 'water' or /rəm/ ⟨rm⟩ 'people', remain light even in the plural (nom.-acc.pl. /unda/ ⟨wnd'⟩ , /rəmta/ ⟨rmt'⟩). I can only aver that

these letters are non-moraic before other consonants, as Sims-Williams did in suggesting they marked the nasalisation of a preceding vowel rather than fully-fledged nasal consonants.

A similar phenomenon occurs in the verbal system. Verbal stems that end in 'r' are light if followed by an inflectional ending beginning with a vowel, and heavy otherwise. The deletion of the final vowel would have taken place in the inherited third person singular **ti* and in the inherited second person plural **ta*, but in both cases (especially in the latter) there has been a tendency to restore the vowel (Gershevitch 1954, §§ 722-45). Here is an example of the stem /βar/ ⟨βr⟩ 'bring'.

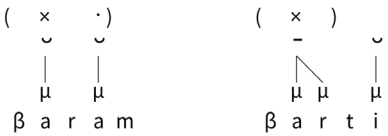

Figure 5 Parsing of /βaram/ and */βarti/ into moraic trochees

This interpretation of the Sogdian Rhythmic Law has parallels in many other languages. In the history of Germanic, high vowels (/i/ and /u/) were lost only in the position targeted by the Sogdian Rhythmic Law, viz. in an unfooted syllable (Dresher, Lahiri 1991; Smith, Ussishkin 2015, 17.3.3.1).

Thus we can see that, in Sogdian, FSR operated on two levels: first, the quantity of all final syllables was reduced across-the-board (except in monosyllables, in observance of word minimality constraints); second, final vowels were targeted for deletion when they were unfooted, that is, at least when they were preceded by a heavy syllable, and perhaps in other cases as well, depending on how exactly syllables were parsed into feet. Sogdian therefore represents an 'intermediate' phase of FSR, when inherited final syllables were already quantitatively reduced, but before they had been eliminated entirely.

Still, even at this intermediate stage, FSR had wrought enormous changes in the inherited system of inflection. Heavy nominal bases ended up with very few inflections: nouns did not distinguish gender, and among cases, only the direct, vocative, and oblique were distinguished. Light bases, by contrast, maintained a two-way gender distinction and a seven-way distinction among cases, if only in the singular.

Besides nominal bases that synchronically ended in a consonant, which admit of a distinction between light and heavy, Sogdian also had nominal bases that synchronically ended in a vowel. These were formed by the addition of the pleonastic suffix **-ka-* or **-kī-* to an

existing base. These 'vocalic stems' maintain the distinction between masculine and feminine only in the direct case of the singular (*-ē* ⟨-y⟩ masc. vs. *-ā* ⟨-'⟩ fem.). In the singular oblique, and in the plural direct and oblique, there is no distinction of gender (Skjærvø 2007, 17).[6]

The use of an extended vocalic declension was grammaticalised in Sogdian. The inherited past passive participle, which was used as such in the earliest documents, is only used as a participle in the later language in the vocalic declension (Gershevitch 1954, § 531: 82). The same applies to the present participle, which is only used as such in the later language in the vocalic declension (132).

4.2 Middle Persian

The final syllables of Old Persian were all lost in Middle Persian. Their loss appears to have been preceded, as one would expect, by a reduction in their quantity. Even later Old Persian inscriptions attest to the loss of distinctions in final syllables. Cantera (2009, 26) notes that, for nouns whose nominative and accusative forms had the same number of syllables, the distinction between these forms was effectively lost by the later fifth century BCE. Korn (2013, 84-5) argues that distinctions of quantity and quality were lost before a final 'm', and *-Vm* had become *-u* by the time Persian words were borrowed into Armenian. Much discussed in connection with the loss of final vowels is the interpretation of the sign ⟨y⟩ at the end of a word in Middle Persian. It had generally been seen as a phonologically insignificant marker of the end of a word. Back then claimed, controversially, that it represented a weak word-final vowel that was lost, in later Old Persian, if the stem was polysyllabic and the preceding syllable contained a long vowel (1978, 41). This is therefore an attempt to apply the Rhythmic Law of Sogdian to Persian.[7] Huyse (2003) revised this account, and concluded instead that the grapheme ⟨y⟩ represented not any final vowel in Old Persian, but specifically the ending */e:h#/ of the genitive singular (from */ahya:/, p. 54), and that this final vowel was lost not in all words at the same time, but at first in (a) polysyllabic words which were (b) neither compounds nor nominal derivatives and where (c) the final vowel */e:h#/ follows a stressed syllable with a long vowel, so long as (d) this long vowel does not represent the end of the word (p. 98). Effectively, as Hintze (2006) notes in her review, this means that monosyllabic stems regularly retain 'y',

6 Sims-Williams (1990, 286) follows Tedesco in deriving the vocalic declension by means of contraction from *-*aki* and then *-*ai*; Gershevitch derived them rather from *-*ak*, as in Middle Persian (with loss of the final vowel).

7 Back 1978, 40-6; cf. MacKenzie 1982; Huyse 2003, 23-5 for a discussion.

while polysyllabic stems retain it only after a long vowel or a short vowel followed by a consonant. There must be a prosodic explanation for this distribution, especially in view of Huyse's observation that in compounds with *bed* 'master', 'y' is written only when the syllable preceding the final member of the compound is prosodically light (p. 80). I will, however, resist trying to state a general prosodic condition for the loss or retention of 'y', given that it was widely extended by analogy to the point where it became, in the late Sassanian period, an orthographic marker of the end of a word.

For the purposes of this paper, it is sufficient to state that the final vowels and vowel + consonant sequences of Old Persian were quantitatively reduced before the Middle Persian stage, resulting in a system wherein every word ended with a short vowel. Whether or not some of those endings were preserved in words with a certain prosodic shape in Middle Persian, the general tendency, evident by the end of the Sassanian period, was the loss of all final vowels.

4.3 Bactrian

The Bactrian data are easy to summarise: "The P[roto-]Ir[anian] [final] vowels are generally lost" (Gholami 2014, 2.7.3). The letter 'o' is typically found at the end of Bactrian words, although it does not have any phonetic significance. Huyse (2003, 61) argues that it originally represented "une voyelle brève réduite" that was later reinterpreted as a word-boundary marker.[8] Gholami (2014, 2.7.3: 61) notes a few Bactrian words that end in vowels other than 'o', without however noting that all of them consist of two light syllables: ιθα 'thus', βαγε 'gods', ταδι 'then', and λιζα 'citadel' (the last from **dizā*). These 'exceptions' clearly indicate that the loss of final vowels was blocked if the resulting form would have been shorter than the minimal prosodic word, i.e. a moraic trochee (two light syllables or one heavy syllable). The vowel is lost even in disyllabic words if the first syllable is heavy, e.g. μαδο /mād/ 'mother' from **mātā* (Gholami 2009, 31). This reminds us of the Sogdian Rhythmic Law: a final syllable is lost if and only if it is preceded by a moraic trochee.

The suffix **-ka* is very common, although as far as I can determine, it occurrence is conditioned lexically rather than by grammatical category (one exception being adjectives from **-ānaka-* > -αγγο, Gholami 2009, 36-7). It is hard for me to say, based on the materials

8 Sims-Williams (*Encyclopedia Iranica* s.v. Bactrian) notes that 'o' also represents an unstressed **a* or **i* in the middle of a word, and that its use at the end of words already ending in a vowel (e.g. ναμωο for Sanskrit *namō*) suggests that it does indeed function as "a word- or morpheme-divider."

I have seen, how this suffix interacted with prosodic principles; in some cases (perhaps most?) a stem-final 'a' was lost by syncope, resulting in forms like ʋαþκο /waːʃk/ 'pupil' < **hāwišta-ka-* and αγγαργο /angaːrg/ 'property' < **ham-kāraka-* (both from Gholami 2009, 36); in other cases (following a conjunct consonant?) it was not, e.g. ανδαγο /andag/ 'borough' < (?) **anta-ka-* (36).

4.4 Khotanese

The following changes are grouped under Kümmel's "Auslautgesetze" (2008, 10). I assume, with Kümmel, that final 'e' and 'o' in Khotanese are short, despite their length not being distinguished in the writing system (cf. also Hitch 2016, 121).

- Inherited final vowels are shortened.
  - */aː#/ becomes /a#/.
- Inherited final consonants are lost.
  - */am#/ becomes /u#/.
  - */ah#/ becomes /ə#/.
  - */ih#/ becomes /ə#/.
  - */aːm#/ becomes /o#/.
  - */aːh#/ becomes /e#/.

These reductions did not apply to monosyllables; cf. the metrically heavy words *o* 'or', *ko* 'if', *co* 'who', etc. (Hitch 2016, 139). Thus the reduction of the quantity of final syllables in Khotanese is exactly parallel to Sogdian, in that it bypassed heavy monosyllables in order to satisfy word minimality constraints.

Long vowels could occur at the end of a word secondarily in Khotanese, i.e. as a result of a contraction of syllables at the end of a word (e.g. *āstai* nom. sg. of the stem *āstaa-* 'bone'). This contraction was usually occasioned by the use of the suffix **-ka-*, which was conditioned, as in the case of Bactrian, lexically rather than by grammatical category.

The reduction of final syllables entailed, as in Sogdian, a loss of distinction between the nominative, accusative, and vocative case forms (Sims-Williams 1990, 284).

Scholars have proposed that Khotanese had a Latin-like stress pattern wherein the penultimate syllable was stressed if it was heavy, and otherwise the antepenultimate syllable was stressed (Kümmel 2008, 4.1.4). This largely (though not entirely) coincides with Maggi's rule that the "stress accent fell on the first heavy syllable from the end of the word but never on the word final syllable" and "[a] word containing only light syllables was stressed on the initial syllable" (Emmerick 2009, 2.2). Hitch noted that "normally the final stem

syllable in O[ld] Kh[otanese] carries primary stress" (2016, 3.3.2.1.5), which was probably meant as a generalisation, not a rule.

Khotanese, somewhat like Sogdian, represents an 'intermediate' phase of the operation of FSR. By the time of the earliest documents, the final syllables had been quantitatively reduced, but never lost. Early Khotanese is therefore the most 'conservative' Middle Iranic language, because final syllables, and hence inflectional endings, are largely retained, albeit reduced in quantity. By the time of the latest documents, however, those syllables were lost entirely (Hitch 2016, 3.2.6; cf. also Emmerick 1987). Bailey (1963, VII-VIII) distinguished between four stages of the language on the basis of this process of loss, illustrated by the inherited genitive plural ending (Proto-Iranian *-ānām*):

- *-ānu* (*/ānam#/ becomes /ānu#/)
- *-āni* (*/ānu#/ becomes /ānə#/)
- *-āṁ* (*/ānə#/ becomes /ān-∅/)
- *-ā* (*/ān#/ becomes /ā-∅/)

Interestingly, he compares the last stage to Apabhramsha (see below), although does not explain it.

4.5 Gandhari

In Gandhari, as in most (if not all) Middle Indic languages, all inherited final consonants other than /m/ were lost, and hence all words ended with a vowel or a vowel followed by /m/ (written as an *anusvāra* in Kharōṣṭhī). The Kharōṣṭhī script in which Gandhari was written does not normally distinguish vowel length. Nevertheless, on the basis of an analysis of the writing of final vowels in post-Aśōkan Gandhari, Fussman (1989, 479) came to the conclusion that at the beginning of the common era Gandhari words ended in one of two vowel sounds:

- ultra-short vowels, resulting from earlier */V/, */V:/, or */Ṽ/ (the last representing a still earlier */Vm/, which, in Fussman's account, yielded a nasalised vowel segment prior to the shortening of final vowels);
- short vowels, resulting from earlier */Ṽ:/ (representing a still earlier */V:m/).

Fussman submitted that qualitative distinctions among the ultra-short vowels were also lost, and that the use of vowel markers such as ⟨-e⟩ , ⟨-o⟩ , ⟨-i⟩ , or ⟨-u⟩ are simply conservative spellings for sounds

that had all become /ə/ in the spoken language.[9] Metrical texts, insofar as they present Gandhari versions of texts that were transmitted in other languages, are very likely to contain such conservative spellings/pronunciations, and hence one important source of information about the length of final vowels is less dispositive than we might like it to be.

A less radical position was taken by Baums (2009, 126-8), who argued that (at least in the British Library fragments 7, 9, 13 and 18) the evidence only supports a merger, in final position, of /e/ and /i/, on the one hand, and of /o/ and /u/, on the other hand: "There is no clear indication that word-final palatal and labial vowels were conflated either with each other or with the neutral vowel" (2009, 128). This would appear to be supported by Brough's observations about the Gandhari *Dharmapāda*: etymological 'ē' and 'ō' are very often written with 'i' and 'u', respectively, but only when they occur in word-final position; in his grammatical sketch he often refers to the "weakness of final syllables" (e.g. 1962, 80, 83).

Loukota offers a different phonological interpretation of Gandhari, according to which there must have been a quantitative contrast between final ⟨-o⟩ (the masculine-neuter direct case ending) and final ⟨-a⟩ (the feminine direct case ending).[10] Evidence for this contrast comes from metrical texts in Gandhari (where a final ⟨-a⟩ often has to be realised as [aː]) as well as the borrowing of Gandhari words into Khotanese and Han Chinese. In Han Chinese, Gandhari words in ⟨-o⟩ are borrowed without the final vowel, whereas Gandhari words in ⟨-a⟩ are borrowed with the final vowel. This contrast could be interpreted in a number of ways, for example as ∅ ~ /a/, /ə/ ~ /aː/, or (following Fussman, albeit with different conditioning factors), ultra-short V ~ short V.

On any of the above interpretations, the status of monosyllables like ⟨ka⟩ is unclear. In other words we cannot know whether the spelling represents /kaː/ or /ka/, or /kə/ or /kă/. In Fussman's and Baums' analyses, as well as one interpretation of Loukota's, the only sound that could occur at the end of a word in Gandhari by the first century or so was a short (or ultra-short) vowel. This probably represents a change from Aśōka's inscriptions at Shabazgarhi and Mansehra. These inscriptions often write *-Vṁ* at the end of a word. On the assumption that 'ṁ' represents a final consonant segment that contributes to syllable weight, rather than a marker of non-moraic

9 "Il est probable qu'en prononciation normale les voyelles ultra-brèves de type a) ne se distinguaient plus les unes des autres et ne jouaient donc plus de rôle phonologique" (Fussmann 1989, 479).

10 I thank Diego Loukota (p.c.) for discussing this with me. This paragraph is my representation of the views he shared with me, which I hope not to have misrepresented. Cf. Loukota 2022 for a detailed discussion of Gandhari loanwords in Khotanese.

nasalisation, and on the assumption that such forms are not borrowings from another dialect, this would mean that Aśōkan Gandhari tolerated heavy syllables at the end of a word, and presumably it would have tolerated long vowels in that position as well. Nevertheless we see already in the Aśōkan inscriptions the use of different endings for the same grammatical category (e.g. *danaṁ*, *dana*, *dane* for the neuter nominative-accusative singular, *jano*, *jana*, *jane* for the masculine nominative singular). Precisely what to make of these alternations is not clear, although it seems significant that they primarily affect those cases that would serve as the subject of a verb (nom. masc. sg. and nom.-acc. neut. sg.; cf. also Caillat 1992, 2.5, 4.4).

The most important grammatical consequence of the reduction of all final syllables was the collapse of the nominative and the accusative into a single form, the direct case. Distinctions between the non-direct cases were retained, because they all added a syllable to the stem. Hence the direct cases were isosyllabic (stem *jana* > direct *jano*), loosely following Cantera (2009) on Middle Persian, whereas the non-direct cases were pleosyllabic (stem *jana* > genitive *janasa*). Gandhari in fact inherited two endings for the locative singular, isosyllabic *-e* (*jane*) and pleosyllabic *-ami*, *-aṁmi*, or *-aspi* (e.g. *janaṁmi*). The inherited ending *-e* was largely used in formulaic phrases (Fussman 1989, 460); a new pleosyllabic ending, *-ae*, came to be used instead.

Since masculine and neuter were only distinguished in the direct cases to begin with, the loss of the distinction between the nominative and accusative would have entailed a loss of the distinction between masculine and neuter stems. There is abundant evidence that historically masculine forms (*-o* and *-e*) were used for erstwhile neuter stems; *-a*, from earlier **-am*, is also often used for the nominative of erstwhile masculine stems, where it might come from either the inherited accusative or the neuter nominative-accusative.

Finally, I will mention that the prevailing interpretation of the Gandhari writing system involves a large number of fricative sounds that are not found in any other Middle Indic language. Single intervocalic consonants were fricativised in Gandhari, which we can tell because the resulting sounds were sometimes written in such a way that distinguished them from their non-fricativised (or voiceless) counterparts:

Table 2 Intervocalic stops in Sanskrit and Gandhari

Sanskrit form	Gandhari form
matam	*mada* [məðə]
katham	*kasa* [kəθə]
mēdhāvī	*meṣ̲avi* [me:za:vĭ]
māsē	*maṣ̲e* [ma:zĕ]
prabhā	*pravha* [prəʋă] or [prəβă]
svakam	*spaya*, *spaga*, *spaa*, *spaka*, etc. [spəjə] (evidently through [spəɣă])

Note that the use of these signs to indicate fricativisation occurs in documents after Aśōka's inscriptions (Konow 1929, lcviii), the significance of which I will return to below (see section 6).

4.6 Apabhramsha

I will leave for section 5 the question of precisely how Apabhramsha fits into this picture in chronological and geographical terms. For now, however, I note that Apabhramsha reduces all final syllables of polysyllabic words (superheavy syllables such as *ām* had already been eliminated in earlier stages of Middle Indic):

- Inherited long vowels are reduced finally. /a:#/, /i:#/, /u:#/, /e:#/ and /o:#/ became /a#/, /i#/, /u#/, /i#/, and /u#/ respectively.
- Inherited vowels followed by a final nasal are reduced. This development is slightly different from the above. Vowels that were followed by a nasal segment are also shortened, with different outcomes (in the case of /a/) depending on whether the vowel is long or short:
  - Old Indic /am#/, /im#/ and /um#/ become /u#/, /i#/ and /u#/ (probably through */Ṽ#/);
  - Old Indic /a:m#/, /i:m#/ and /u:m#/ become /a#/, /i#/ and /u#/ (probably through */Ṽ:#/);

Note that the reduction described above only happens regularly in polysyllabic words. Monosyllables such as *kō* 'who' etc. often retain the length of their final vowel. (Apabhramsha, as a literary language, freely admits forms from Prakrit, another Middle Indic literary language, which obscures to some degree the reduction of final syllables, since Prakrit was not affected by this reduction at all).

There were no contexts in which Old Indic /e:/ and /o:/ were followed by a nasal segment at the end of a word, which is why I did not include them above. However, the regular instrumental singular ending of nominal stems in /-a/ in Apabhramsha is /-ẽ:/ (usually written ⟨-eṁ⟩ and metrically heavy), which seems to derive from the

Old Indic ending /-eːna/. This raises another question: what happens to inherited short vowels in final position in Apabhramsha? Although a thorough investigation would be outside the scope of this paper, we can note a major generalisation that we will return to subsequently (I cite the Prakrit forms as a representative of 'Common Middle Indic'):

- Endings that in Common Middle Indic have the prosodic shape ¯˘(/V:CV#/ or/VC:V#/) are regularly continued in Apabhramsha by endings of the shape ¯ or ˘˘:
  - /aːni#/ (neuter plural nominative-accusative) → /aĩ#/;
  - /eːna#/ (instrumental singular) → /ẽː#/;
  - /eːhĩ#/ (masculine-neuter instrumental plural) → /ehĩ#/ (also locative);
  - /aːhĩ#/ (feminine instrumental plural) → /ahĩ#/;
  - /aːe#/ (feminine instrumental) → /ae#/ (also note the appearance of a new genitive-dative-ablative in /Vhe#/ for feminine stems);
  - /asːa#/ (genitive singular) → /aha#/;
  - /aːnã#/ (genitive plural) → /ahũ#/;
  - /aːmi#/ (first-person plural) → /ami#/;
  - /aːmo(ː)#/ (first-person plural) → /ahũ#/;
  - /anti#/ (third-person plural) → /ahĩ#/.
- Endings that in Common Middle Indic have the prosodic shape ˘˘ (/VCV#/) are regularly continued in Apabhramsha by endings of the same prosodic shape:
  - /asi#/ (second-person singular) → /ahi#/.
  - /aha#/ (second-person plural) → /aha#/.

While the stages in the development of some of these endings are somewhat unclear and a matter of great controversy among linguists – some forms, however, being clearly analogical – we can nevertheless make a generalisation: all terminations were adjusted to match either the template ˘ (if the Common Middle Indic termination was ¯) or the template ˘˘ (if the Common Middle Indic termination was ˘˘ or ¯˘), which I will call 'template A' and 'template B'.

The prosodic form of the word had other implications. The very productive suffix **-ka-* or **-kī-* (Jamison 2009) was added to nominal stems – but not in the manner of a derivational suffix, which is how it had been used in Sanskrit and Prakrit, but more in the manner of an augment used to build certain case suffixes. That is, the nominative-accusative singular of all genders could take the suffix (resulting in the endings *-aü* in the masculine and neuter and *-ia* in the feminine) as well as the locative singular of the masculine and neuter (resulting in the ending *-ae*/*-aï*). These are the 'isosyllabic' cases discussed in connection with Gandhari above, and they are precisely the endings that would have taken template A (˘, deriving from a Common

Middle Indic template ¯). The suffix was not used in the remaining ('pleosyllabic') endings.

The effect of the suffix was to make the endings in template A conform to template B. The underlying motivation for this change is the same avoidance of unfooted vowels that we saw in the Sogdian Rhythmic Law: a final ˘˘ constituted a moraic trochee, whereas ˘ did not, and might end up unfooted, depending on the weight of the preceding syllable. Indeed the distribution of the extended endings corroborates this explanation: they appear "mainly after heavy syllables" (Tieken 1998, 1), i.e. in words that would have ended ¯˘ without the extension, resulting in an unfooted vowel. That is not to say that words in Apabhramsha did not or could not end in the pattern ¯˘. It only means that the use of extended endings was much more likely in words that would have ended in that shape than others.

The prosody of the stem and the endings only partly accounts for the distribution of extended endings in Apabhramsha, since – in a way that once again reminds us of Sogdian – the extended endings tend to be used with only certain grammatical categories, namely adjectives and participles. And finally, as Tieken (1998) has emphasised, Apabhramsha exhibits the synchronic availability of different diachronic stages of linguistic development, which effectively means that poets could use extended or unextended forms based on preference and metrical exigencies. In spite of the opportunity for free variation, however, the regularity with which extended endings were used only in template A endings and largely after heavy syllables suggests that prosodic considerations played an important role in the reorganisation of the inflectional system after FSR.

Finally, regarding the phonetic value of intervocalic stop consonants in Apabhramsha, it would appear superficially that Apabhramsha and Prakrit are exactly the same in this respect: namely, aspirate stops are debuccalised (/VC$^{\text{ɦ}}$V/ becomes /VɦV/) and most non-aspirate stops are generally lost (/VtV/, /VkV/, /VdV/, /VgV/ become /VV/). But this account only covers what I will later call, following Bhayani, the second stage of Apabhramsha, which was strongly influenced by Prakrit. In the first stage, intervocalic stops are written with their voiced equivalents, whether aspirated or not, e.g. *kadhidu* for *kathitam*. Although there is no evidence for a fricative pronunciation of these consonants, these forms arguably represent a stage of the language closer to Gandhari (where *kathitam* would have developed to /kaθiðə/) than to Prakrit (where *kathitam* developed into /kaɦiaN/).

5 Analysis

Several questions can be formulated on the basis of the preliminary results above. One is whether FSR itself needs an explanation, and if so, what kind of explanation it requires. Another is whether the occurrence of FSR in Gandhari and Apabhramsha, alone (at least for a time) among the attested Middle Indic languages, requires an explanation, and if so, what kind.

At least within the study of the Indic languages, the reduction of final syllables has traditionally been seen as a natural continuation of widespread phonetic tendencies in the Indo-European languages. Nearly a century ago, Turner called attention to the "phonetic weakness" of "terminational elements" in the Indic languages. He noted that:

> From the time of Aśoka onwards the documents of Middle Indian testify to the complete disappearance of final consonants and the progressive shortening of final vowels, of which in the majority of the modern languages there is now no trace, except where their identity has been partially preserved by contraction with a preceding vowel. (Turner 1927, 229)

He also noted that "in the ancient Indo-European languages the phonetic elements following the initial consonant of the last syllable of words are subject to quite special alterations and weakenings" (Gauthiot, quoted in Turner 1927, 229).[11] If we stipulate, as a law of historical linguistics, that final elements of words are subject to various kinds of weakening and reduction, then no explanation is necessary for the loss of such elements in the Indic languages between the stage represented by Sanskrit and the stage represented by, for example, Hindi.

Earlier philological studies remained vague about the ultimate explanation for a general tendency toward loss of material at the right edge of a word. More recent research, however, has appealed to the general "articulatory and perceptual properties of speech" as a way of explaining why certain sound changes appear to be very widely attested and "natural" (Blevins 2008). In her "Field Guide" to natural and unnatural sound changes, Blevins lists "final vowel shortening" as a change with a primarily perceptual basis (referring to Myers, Hansen 2006).

11 Cf. also Tagare 1948, 49: "There seems to be a tendency in OIA to pronounce the final syllable weakly, as it was probably unaccented" (I thank an anonymous reviewer for this reference).

I note, however, that FSR in the form in which we encounter it in Middle Indic and Middle Iranic is neither identical to a generic "weakness of terminational elements" nor the specific neutralisation of vowel-length contrasts at the end of a word. As a matter of fact, the examples that Turner adduced from Aśōka's inscriptions, as well as earlier stages in the Indic languages, were examples not of the reduction of quantity or quality of final syllables, but of the lenition of intervocalic consonants in inflectional endings. The specific features of FSR discussed above require a more specific explanation, that is, one in which the prosodic structure of a word interacts with constraints on the preservation and loss of linguistic material.

Auer (1993) introduced a graded typological classification of languages based on whether the syllable or the word is more fundamental in their prosodic phonology. A wide range of phonological and morphological features have now been shown to correlate with each other, placing a language on a continuum between 'word languages' and 'syllable languages'. In a contribution to a volume on this topic, Martin Kümmel (2014b) has argued convincingly that, despite their very close relationship, the Indic and Iranic languages have "drifted" toward different ends of the word/syllable continuum. The Iranic languages have exhibited a number of sound changes that are typically associated with word languages, even from the very earliest evidence. For example, the distinction in the quantity of final vowels was neutralised in all of the Old Iranic languages (Old Persian, Avestan, and Young Avestan; Kümmel 2014b, 209). Another diagnostic feature of word phonology in the Iranic languages is complex onsets and codas relative to the Indic languages. By contrast, many of the changes that separate the Indic languages from Proto-Indo-Iranian are associated with syllable languages, including the reduction of complex onsets and codas and the implementation of a variety of *sandhi* rules both within and between words. Taken together, these features allow us to at least formulate the hypothesis that the changes diagnostic of membership in the Iranic branch of Indo-Iranian implicate the word as a significant prosodic unit, while those that indicate membership in the Indic branch implicate the syllable.

From this perspective, it is very easy to see FSR as a word-related development, given that it is at the right edge of a word that syllables are reduced. The changes implicated in FSR can thus be restated:

1. Neutralise syllable quantity at the end of a prosodic word (evident in Old Persian, Avestan, and Young Avestan; also Khotanese and Apabhramsha, and perhaps Gandhari);
2. Neutralise vowel quality or delete a vowel altogether at the end of a prosodic word:
    a. if the vowel is in an unfooted syllable (Sogdian);
    b. generally (Bactrian, Middle Persian, perhaps Gandhari).

In the pre-print version of his paper, Kümmel (2014a) does not discuss FSR *per se* (although he does note, e.g. the loss of all final syllables in Western Iranic), but he does discuss the reduction and loss of unstressed vowels in Middle Iranic in general, which contrasts very clearly with the tendency toward epenthesis in Indic.

Kümmel (2014a) also noted that some of the syllable-linked developments in the Middle Indic languages did not operate in the far northwest. Specifically, the reduction of heterorganic consonant clusters, which operates almost without exception in all of the other attested Middle Indic languages (Ardhamagadhi, Prakrit, Epigraphic Middle Indic, and with the exception of Sanskrit loanwords, Pali), bypassed the languages of the Northwest, including Gandhari as well as modern Dardic languages.

By this point it should be clear that, on the one hand, FSR is 'natural' in the sense that it is based on the articulatory and perceptual properties of speech, and therefore could, in principle, happen anywhere and at any time; on the other hand, however, FSR is part of a 'conspiracy' of sound changes and typological features associated with languages wherein the word is prominent as a unit of prosodic structure, and therefore, in fact, has been confined to languages where such changes have been able to overcome general structure-preserving constraints.

Now we can return to the distribution of FSR in the Middle Indic and Middle Iranic. As noted above, FSR occurs across-the-board in Middle Iranic and only in those Middle Indic languages that are likely to have originated in the Indo-Iranian contact zone, in Gandhara and along the Indus river. Now we can add that FSR is part of a cluster of sound changes – and typological characteristics, which we can see as the ultimate outcomes of structure-modifying sound changes – that are indeed associated with the Iranic language family as a whole, and that the Middle Indic languages that participate in FSR are, by the same token, less likely to participate in the syllable-related sound changes that affect other Middle Indic languages.

Consider, as a parallel case, the loss of final syllables in French. On the one hand, some of the antecedents of this change – the loss of final consonants in Vulgar Latin, and the subsequent reduction in quantity of all final vowels – are shared by all of the Romance languages. Yet not all Romance languages lost their final syllables outright (Vaissière 1996). In fact, it is only the Gallo-Romance branch in which this change was regular and widespread. In other branches, such as Italic, the loss of final syllables is manifestly related to proximity to the Gallo-Romance branch. This very strongly suggests that the decisive sound change, although 'natural' from one perspective, occurred in one branch of the Romance languages, and then spread, by contact, to neighbouring branches.

Thus I propose that contact with Iranic languages is one of the main reasons why FSR, among other word-related phenomena, is found in the Indic languages of the Northwest. I would go further and claim that, despite the long history of contact between Indic and Iranic speakers in that region, attested by many layers of loanwords, evidence suggests that FSR in Gandhari, at least, coincides with waves of Iranian migration in the region that began in the second/first century BCE.

I would even suggest that the extent of FSR in the languages spoken in those waves of Iranian migration – above all the Saka language – can be reconstructed based on its effects in Indic languages such as Gandhari and Apabhramsha. We have no contemporary documents in the Saka language, although words and names in that language were written in the Kharōṣṭhī and Brāhmī scripts with characteristic modifications (e.g. 'ys' for /z/). In that language, final syllables were likely all light, as they were in the oldest stage of the most closely related attested languages, Khotanese and Tumshuqese. Whether the inflectional endings were adjusted to a prosodic template is hard to say: we have observed that such an adjustment is evident in Gandhari (isosyllabic direct cases and pleosyllabic non-direct cases), and much more regularly in Apabhramsha (with two dominant prosodic templates for the ending of a word); perhaps similar phenomena were involved in Saka phonology. The absence of syncope in Gandhari suggests, too, that the Saka language did not exhibit syncope to the same extent as, for example, Bactrian.

We can now turn to other linguistic features that I have associated with FSR. I noted above that the inherited Indo-Iranian suffix *-ka-* is often used to 'repair' the effects of FSR by building new forms which are sometimes grammaticalised. By contrast, consider the explanation of Jamison (2009, 314) for the spread of the suffix in the Indic languages. She invokes the

> relentless, inexorable progress of sound change, which was steadily eroding the ends of Indo-Aryan words, and morphological change, which was streamlining the old complex series of nominal stem formants in favor of vowel-stems, especially *-a-* and *-ā-* stems. These processes set the stage for the 'real' part of the *-ka-* explosion. If *ka*-forms could always have served as lower register doublets to more dignified higher register forms, as those latter forms became threatened by linguistic erosion, speakers would have been inclined to make themselves clearer and to preserve the physical body of a word by, as it were, promoting the more colloquial *-ka-*forms into standard discourse, and also generating more of them – not to mention finding this a convenient way of avoiding consonant stems and other nasty bits of morphological business.

This is an eloquent account of some of the reasons why semantically weak suffixes can spread throughout a language.[12] But I am not sure that sound change in general is "inexorable" and "steady" (despite this serving as a useful assumption in some recent work in historical linguistics); I am convinced that all sound change is sociolinguistic in origin, and hence its progress depends very much on rhythms of interactions that philologists may not have access to. If I am correct, then we need a finer-grained account of the "*-ka*-explosion". There is of course the underlying sociolinguistic phenomenon that Jamison's article concerns in the first place, namely the use of *-ka*-forms in lower registers of Sanskrit and related languages, and their 'percolation', through lexicalisation, to higher registers. In this respect a close parallel is offered by Khotanese and (apparently) Bactrian, where some lexemes are obligatorily formed with the *-ka-* suffix. Prakrit is a bit different, in that *-ka-* can be used, optionally, with any stem at all, although there are of course preferences among different authors, genres, and periods.

The situation we see in Sogdian – where the presence or absence of *-ka-* is conditioned by grammatical category rather than by lexeme – is rare in most of the Middle Indic languages. I can think of two counterexamples. One was noted by Jamison herself: in the Gandhari of the Niya documents, which was probably in even closer contact with Iranic languages than the Gandhari of Gandhara, the past participle in *-ta-* was used for the conjugated past tense, whereas past participles functioning adnominally took the suffix *-taka-* (Jamison 2009, 317). In Apabhramsha, as we have seen, the suffix *-ka-* was virtually integrated into the inflectional endings of isosyllabic cases. And this, too, was conditioned not just by prosodic factors (being more likely after a heavy syllable) but also by grammatical category: "the long and the extended endings… are indeed found almost exclusively with adjectives and past [participles]" (Tieken 1998, 3). Some categories, such as future participles in *-tavya-*, are "always extended" (15).

Finally we can consider the development of fricative consonants. Here, too, scholars have seen the loss of intervocalic stop consonants in the Middle Indic languages as part of a general (perhaps even "inexorable" and "steady") tendency within this language family. And here, even more in the case of FSR, the change is explicable based on general principles of articulatory economy (i.e. speakers will be as lazy as their addressees allow them to be). As noted above, scholars

12 Although Jamison may be right about *-ka-* allowing speakers to "avoid" consonant stems, my impression is that inherited consonant stems were 'thematicised' in the Middle Indic languages with the simple addition of *-a-* (e.g. *śarad-* to Prakrit *saraa-*) or *-ā-* (e.g. *diś-* to Prakrit *disā-*) rather than with *-ka-*.

sometimes envision this process in several stages: voicing of voiceless stops and fricativisation of voiced stops; fricativisation of the newly-produced voiced stops and total loss of the newly-produced fricatives; repeat (e.g. /VtV/ → /VdV/ → /VðV/ → /VV/). But it must be said that we have no evidence for a fricative pronunciation of any intervocalic stops outside of Gandhari. And it is perfectly possible for intervocalic consonants to be weakened or dropped without an intermediate phase of fricativisation. (Note that intervocalic /t/ is lenited to [ɾ] or [ʔ] in weak positions in English words like *little*, but never fricativised.) In texts written in the Brāhmī script, fricatives could not even be written without introducing new orthographic conventions that distinguished ⟨tt⟩ [t] ~ ⟨t⟩ [d] ~ ⟨d⟩ [ð], as was done for Khotanese (cf. von Hinüber 1981). And even if they were pronounced, the lack of a phonemic distinction between, say, [d] and [ð] would have encouraged a conservative orthographic practice that wrote both sounds with the same sign. All of this raises the question of why fricatives not only did develop in Gandhari but were actually written with newly-developed characters of the Kharōṣṭhī script, generally formed by the addition of a *cauda* sign (Glass 2000, 136). These characters, I repeat, were not used in Aśōka's inscriptions, but only appear in Gandhari documents of the Saka-Kuṣāṇa Age. The most straightforward explanation is that Gandhari adopted fricative sounds from neighbouring Iranic languages.

To support this view we can observe that some of the same letters that are used to represent a fricativised outcome of an inherited stop are also used to represent fricative sounds in Iranic loan words and proper names: 'vh' represents inherited */bʱ/, but also Iranic /f/ (e.g. *vharna*); 'v' and 'vh' both seem to represent Iranic /β/ (e.g. *vaka* or *vhaka*). In other cases, It has also been proposed that some conjunct consonants that look like simple continuations of inherited conjuncts, such as 'ks', are actually used, in some cases at least, to represent Iranian fricative clusters (e.g. *kṣuna* /χʃunə-/; see below on this word). In fact, the fricativisation of stops in Gandhari may have been conditioned by the very same changes in word-level prosodic phonology implicated in FSR. We know that processes of lenition are sensitive to foot structure, thanks especially to the work of Margaret Withgott (1982).

Before moving on to some of the implications of this analysis, I will mention one more possible explanation for the differences observed between the Indic languages. The classification of the Indic languages remains controversial (cf. Ivani, Paudyal, Peterson 2021). Several of the features that appear to distinctively characterise the languages of the Northwest, such as Gandhari and the modern Dardic languages, have been attributed by scholars to the membership of these languages in subgroupings where special developments occur, or where developments that commonly took place elsewhere did not

occur. The idea of an inner/outer divide between the Indic languages is particularly challenging for the analysis proposed here, because on that theory, we would expect at least some features characteristic of the Indic languages of the Northwest to be due not to contact but to their membership in a subgrouping of 'outer' languages that includes, for example, Sinhala – a language as far removed from contact with Iranic speakers as can be imagined.

Claus Peter Zoller, a proponent of the inner/outer thesis, on the one hand, attributes a cluster of phonological features found in the languages of the Northwest to their membership in the outer subgrouping: the weakness of aspiration in Gandhari, and its total loss in some modern Dardic languages as well as Sinhala (2023, 317) and the development -MP- → -P- (345).[13] On the other hand, he acknowledges that certain features of Niya Prakrit (i.e. the Gandhari used in the Niya documents) are suspiciously similar to features of Khotanese. Rather than see these features as evidence of Iranic influence, however, he sees them the other way around, as evidence of the influence of Niya Prakrit on Khotanese (2023, 368).

Zoller clearly believes that the features that set some of the Middle Indic languages apart from common Indic developments are due to membership in the outer group, and resists contact-based explanations wherever possible. I find his arguments unconvincing for several reasons. One is that he does not distinguish between sound changes and the absence of sound changes when discussing diagnostic features of the inner/outer subgrouping. It of course makes sense that languages as widely separated as Gandhari and Sinhala will, in some cases, not participate in sound changes that affect a 'central' group of languages. For example, Gandhari and Sinhala did not exhibit fortition of initial /j/ to /dž/ (Kümmel 2014a). In some cases we can attribute such conservatisms to a weakened influence of syllable-based phonology, which was evidently stronger in the central group. But some explanation is required for sound changes that affect languages widely separated in space. It may be that there are general features of the parent language of the outer languages that made it more likely for them to undergo certain sound changes – for example, a prosodic phonology that was more word-based than the syllable-based phonology of the inner languages – but these would have to be specified and weighed against alternative explanations. Another

13 A few other features he takes to be limited to the Northwest without implicating other outer languages, e.g. the continued preservation of intervocalic stops (Sinhala 2023, 345).

reason for my scepticism is Zoller's arbitrary, imprecise, and sometimes incorrect interpretation of the Middle Indic evidence.[14]

6 The Development of Gandhari

As new Gandhari documents are discovered and studied, we are likely to gain a much clearer picture of how the language developed. But the following represents what I take to be the commonly-accepted account.

Gandhari is first attested in Aśōka inscriptions at Mansehra and Shahbazgarhi in the middle of the third century BCE. For close to two centuries there are very few surviving inscriptions. But then, "around the latter half of the first century BCE, Buddhist inscriptions suddenly become very common in Gandhāra and the surrounding areas" (Salomon 2018, 29). The earlier Gandhari birch-bark scrolls, all containing Buddhist texts, date from around this period as well (although some may be somewhat older). As Salomon noted, the period of Gandhari's use as a "Buddhist literary language" in the Greater Gandhara region coincides with the period of "Indo-Scythian" rule in the northwest (Salomon 2018, 28-31; 2002, 128). Gandhari also came to be used as an administrative language in the Central Asian kingdom of Kroraina, probably because of political, economic, and religious connections between the Tarim Basin and the Greater Gandhara region during the Kuṣāṇa Empire. As a literary language, however, Gandhari was always in competition with, and influenced by, both Sanskrit and other Middle Indic languages. In South Asia it was displaced by Sanskrit by the end of the third century of the common era (Strauch 2012).

I would make two adjustments to this story, one of emphasis and one of fact. Salomon noted that Gandhari "stands apart from all of the other languages" in the Middle Indic family in regard to its phonology (2002, 119), but by this he meant its conservatism rather than its innovations: its preservation of the three-way distinction between sibilants ('ś', 'ṣ' and 's'), and its preservation of consonant clusters with 'r'. But it is important to note that the Gandhari of the Indo-Scythian period is also distinguished from other Middle Indic languages – and, for that matter, from earlier forms of Gandhari – by its innovations,

14 The invocation of Deśya Prakrit as an 'outer' language is one example (*dēśya-* simply refers to Prakrit lexemes that are not obviously derived from Sanskrit equivalents, and they are attested in [Mahārāṣṭrī] Prakrit, a language that Zoller otherwise considers an 'inner' language); cf. 2023, 364. He also attributes the 'inner' features of Gandhari (as opposed to Niya Prakrit) to the influence of Pali, which is unlikely (2023, 368), and interprets the absence of *anusvāra* in Gandhari writing to reflect a sound change -MP- → -P-, rather than what it almost certainly is, the failure to record the nasality of the syllable in writing (345).

and in particular, the quantitative reduction of all final syllables, and consequent upon this, the restructuring of the gender and case system, features I attribute above to contact with Iranic languages.

Among these distinctive features of Gandhari in the Saka-Kuṣāṇa age is the "inconsistency and lack of standardisation in [its] orthography and morphology" (Salomon 2002, 131). Salomon suggests that "[Gandhari] fell out of use" before the developments that would have established it as a standardised literary language took place (131). We can note that Gandhari is not unique among Middle Indic languages in this time: neither Ardhamagadhi, nor inscriptional Middle Indic, nor Pali were, as far as we know, subject to the kind of grammatical description and regulation that Sanskrit was. But Gandhari is uniquely chaotic in its grammar and orthography. Why? I suggest that it had been 'unstandardised' by intensive contact with speakers of Middle Iranic languages.

What do I mean by this? The example of Sanskrit and Pali might lead us to assume literary languages should be relatively uniform over a large space and a long time. But Gandhari's reinvention as a literary language in the first century BCE was premised on its use by a wide variety of people. Among these new users of the language were the political and military elite of the Northwest, many of whom held Iranian names and titles and presumably spoke Iranic languages as well. Gandhari was probably elevated to the status of a literary language by people who were not native Gandhari speakers, which resulted in a prominent and presumably prestigious variety of the language having a phonology that was basically Iranic. (Compare the use of Gandhari in Kroraina, where we presume that it was an administrative language among people whose native languages were forms of Tocharian.) This 'Iranicised' Gandhari presumably coexisted for some time with a 'non-Iranicised' variety; indeed this may be the root of the distinction between spoken and written Gandhari noted by Fussman (1989). But both were subject to a strong and persistent influence from other Middle Indic languages and Sanskrit. All of these influences probably made Gandhari quite heterogenous at exactly the time it was being cultivated as a literary language in the Greater Gandhara region.

I do not mean to say that Iranic influences were absent prior to the first century BCE. Konow (1929, cxiii) noted that the nominative masculine singular ending *-e*, previously considered to betray the influence of eastern Middle Indic languages that continued **-aḥ* as *-ē*, shows the same development of an inherited **-ah* as the neighbouring Iranic languages (Khotanese *-ä* and *-i*). It is quite possible that parallel developments such as these were reinforced by contact. Similarly, Aśōka's inscriptions in the Northwest use Iranic loanwords (see below), probably attesting to the influence of Achaemenid bureaucracy. But Gandhari was a relatively conservative Middle Indic language at

the time of Aśōka. Two centuries later, when it was adopted as a literary language among the Buddhists of the Northwest, it had undergone changes that made it the most innovative of the Middle Indic languages, and those changes were probably due to contact.

7 The Origins of Apabhramsha

In discussing Apabhramsha, we are moving several centuries ahead in time, at least as far as our evidence goes. Although the word *apabhraṁśaḥ* had been in use since at least the time of Patañjali (second century BCE) to describe forms that "fell away from" normative Sanskrit usage, the earliest source to use it as the name of a literary language is evidently Bhāmaha, author of the *Ornament of Literature* (*Kāvyālaṅkāra*), in the sixth or seventh century CE. All that Bhāmaha tells us is that there were compositions in Apabhramsha in the *mātrā* meter.[15]

The next author to mention Apabhramsha after Bhāmaha is Daṇḍin, who identifies it as the literary language of "the Ābhīras and others" in his Mirror of Literature (*Kāvyādarśa*, ca. 700 CE).[16] The association between Apabhramsha and ethnic groups of the Northwest – the Ābhīras and the Gūrjaras – is also found in a Prakrit novel of 779 CE, Uddyōtana's *Kuvalayamālā*, in which a Gūrjara traveler recites a *dōhā* in Apabhramsha.[17]

Bhayani (1998a) distinguished between two "strata" of early Apabhramsha. On the one hand, there was a corpus of literature in the *mātrā* and *raḍḍā* meters, which preserves some archaic phonological features. This corpus is totally lost, but for a few quotations in later grammars and metrical handbooks, but we know the names of several important authors: Gōvinda, Chaïlla, Śuddhaśīla. These appear to have been lyric verses, often with a pastoral character. On the other hand, there is the *sandhibandha*, a longer narrative composition with 'sections' (*sandhi*s) composed in alternating passages, called *khaḍavaka*s, of a carrying meter and a single-verse 'coda' (*ghattā*). The *sandhibandha* appears to have been a formal innovation of a poet named Caturmukha, who lived sometime before the ninth century CE, and it represents the vast majority of surviving Apabhramsha poetry. The archaic features found in the fragments of the earlier *mātrā* literature are not found in *sandhibandha*s, and Bhayani speculated that the change in literary form was linked to a

15 *Ornament of Literature* 1.16*cd*: *saṁskṛtaṁ prākṛtaṁ cānyad apabhraṁśa iti tridhā*; 1.30*ab*: *gāthāślōkamātrādi*.

16 *Mirror of Literature* 1.36*ab*: *ābhīrādigiraḥ kāvyēṣv apabhraṁśa iti smṛtāḥ*.

17 *Kuvalayamālā* § 115: 59, l. 5.

change in language: the latter poems, he said (1998a, 40), "developed under the impact of literary Māhārāṣṭrī" (i.e. Prakrit).

The two key features that distinguish the earlier from the later stratum of Apabhramsha are:

- the preservation of 'r' in consonant clusters;
- the lenition, but not complete elision, of intervocalic stops.

These features are taught by Hēmacandra in one of the earliest grammatical accounts of Apabhramsha, namely in *sūtra*s 8.4.398 and 8.4.396 in his *Siddhahēmacandraśabdānuśāsana* (ca. 1140 CE).[18] Bhayani (1998a, 36-9) notes that the examples given for most of these 'archaic' features are in the *mātrā* meter. He also notes that these features are found in quotations of Apabhramsha prior to Hēmacandra as well. These include a number of *mātrā* verses in the aforementioned *Kuvalayamālā*, and a *dōhā* given as an example in Virahāṅka's metrical handbook, the *Compendium of Mora- and Syllable-Counting Meters* (*Vṛttajātisamuccaya*, perhaps eight century). *R*-preserving forms are also given for Apabhramsha by Rudraṭa in his *Ornament of Literature* (*Kāvyālaṅkāra*, ninth century) and by Namisādhu in his commentary thereon (1068 CE).

Here is one example from the early author Gōvinda, as cited by Hēmacandra (*Siddhahēmacandravyākaraṇa* 8.4.422), in the *mātrā* meter:

ekkamekkaü jaï vi jōēdi
hari suṭṭhu savvāyareṇa
tō vi drēhi jahĩ kahĩ vi rāhī
kō sakkaï saṁvarevi
daḍḍhaṇayaṇa ṇēhēṁ paluṭṭā

Although Hari looks on each
one of them with respect, of course,
he looks at Rādhā whenever he can:
when love draws the cursed eye
somewhere, who can stop it? (Author's transl.)

Here we can see *drēhi*, possibly *drakṣyati* or some other form of the verb 'to see', with a preserved (or intrusive) 'r'. And we can also see, in *jōēdi*, an intervocalic stop consonant that has been lenited (*dyōtayati* to *jōēdi*) but not elided (compare *jōēi* in Svayambhū's version of the same verse in the *Svayambhūchandas*, 4.10.2).

18 8.4.396: *anādau svarād saṁyuktānāṁ kakhatathapaphāṁ gaghadadhababhāḥ* (with examples including *kadhidu* for *kathitam*); 8.4.398: *vādhō rō luk* (with the example *jaï bhaggā pārakkaḍā tō sahi majjhu priēṇa*).

The distinction between earlier and later Apabhramsha was made only in the 1990s by Bhayani. All of the premodern authors we have mentioned so far – Rudraṭa and Namisādhu, Virahāṅka, Svayambhū, and Hēmacandra – operate with a single linguistic category, namely Apabhramsha. But the so-called 'Eastern grammarians' recognised different varieties of Apabhramsha from a relatively early period, and one of them has been speculatively connected to this earlier form. The Eastern grammarians, identified as such by George Grierson (1924), were a group of authors based in Bengal who developed a distinctive approach to Prakrit grammar based on a finer-grained classification of regional varieties.

Kramadīśvara, whose date remains unknown, appears to follow Hēmacandra in teaching the aforementioned archaic characteristics as optional features of standard Apabhramsha (5.1-2), and includes a number of additional substitutions with intrusive 'r' (e.g. *vrāsa-* for *vyāsa-*, 5.5). However, he has a single *sūtra* that says that the 'r' is regularly retained in conjunct consonants in *vrācaṭādau*, which designates some specific varieties of Apabhramsha (5.66). He gives *sarpi* (instead of *sappi*) and *jrũ* and *drũ* (instead of *jō* and *sō*?) as examples. He gives Nāgaraka and Upanāgaraka as further varieties of Apabhramsha, the latter of which is described as a mixture of Apabhramsha and Prakrit (5.67).

Puruṣōttama, another eastern grammarian who was an exact contemporary of Hēmacandra, similarly divides Apabhramsha in general into Nāgaraka, Vrācaḍa, and Upanāgaraka varieties, except here the division is exhaustive: Nāgaraka corresponds to the 'standard' Apabhramsha taught by Hēmacandra. The Vrācaḍa variety is somewhat different from what Kramadīśvara teaches as Vrācaṭa – it converts all sibilants to 'ś', for example – but it also has the retention of both consonantal and vocalic 'r' (18.3). Mārkaṇḍēya's treatment in his *Sum-Total of Prakrit* (*Prākṛtasarvasva*, 1558-69 CE) is essentially the same, except he makes the interesting comment that Vrācaḍa "comes from Sindh" (18.1, commentary). Rāmaśarman says the same (*Wish-Granting Tree of Prakrit* [*Prākṛtakalpataru*], 3.2.1-2).

As Bubeník (1998, 28-9) noted, the preservation of 'r' in conjunct consonants is indeed a feature of the Indic languages of the Northwest, including Sindh. Together with Apabhramsha's associations with the Ābhīras, whom the Mahābhārata places in Sindh, this circumstance lends plausibility to the theory that Apabhramsha originated in the Northwest of the subcontinent, and should reflect linguistic developments specific to that region. And as noted above, FSR is a characteristic of Apabhramsha and Gandhari alone among the Middle Indic languages. This account would make the 'earlier' stratum of Apabhramsha, at least, similar to Gandhari in that some of its characteristic developments might plausibly be attributed to influence from Iranic languages. Although this conclusion appears

to follow rather straightforwardly from the data that Bhayani had gathered, he himself remained committed to the earlier view that "the Apabhraṁśa was but a colloquialised form of literary Prakrit" (Bhayani 1998b, 32), and emphasised how little we actually know about the origins and early history of the language (Bhayani 1998b, 33; 1989, 2).

8 Lexical Evidence

As an appendix to the argument above, I would like to corroborate my hypothesis about language contact by revisiting the much-discussed question of Iranic loanwords in Indic languages. Mayrhofer controversially, but for our purposes conveniently, divided his *Etymologisches Wörterbuch des Altindoarischen* into two parts, one for the older (1992; 1996) and one for the younger language (2001). By 'older language' he understands "those lexemes that are first attested in Vedic literature, or in any case in the ancient grammarians such as Pāṇini and Patañjali"; the 'younger language' includes "lexemes whose first attestation does not appear prior to the Epics and Lawbooks".[19] The 'dividing line' between these two phases is effectively what I have been calling the Saka-Kuṣāṇa Age. Hence, as a general rule, the Iranic words cited as evidence in the first two volumes of Mayrhofer's *Wörterbuch* serve to establish the lemma as inherited from Proto-Indo-Iranian. By contrast, the Iranic words cited as evidence in the third volume serve to establish the lemma as borrowed from an Iranic language. My survey of the Iranic words cited in the third volume revealed, unsurprisingly, a number of semantic clusters: words related to writing; words related to governance and administration; and military terms. Before discussing each of these clusters below, I also wanted to mention names for plants and other realia, which can often not be connected directly to an Iranic word but appear to be *Kulturwörter* of wide extension (e.g. *karkētana-*, *kuñcikā-*, *dāḍima-*, *mātuluṅga-*, *maśaka-*, *rājāvarta-*, *samūra-*, *stavaraka-*, *hiṅgu-*, *hispittha-*). Bailey noted that many words that are peculiar to Buddhist Sanskrit, and hence discussed in Edgerton's *Buddhist Hybrid Sanskrit Dictionary*, are loanwords, "which have been, as it would seem, introduced by the Iranian-speaking 'Indo-Scythians' of northwestern India in the period from the second century

19 Mayrhofer (1992, X): "Gemeint sind in der esteren Gruppe [i.e. die Lemmata der ‚älteren Sprache', AO] jene Lexeme, die in der vedischen Literatur – oder allenfalls bei alten Grammatikern wie Pāṇini oder Patañjali – erstmals belegt sind; solche Wörter, deren Erstbeleg nicht vor den Epen oder den Rechtsbüchern erscheint, werden in der Abteilung ‚jüngere Sprache' behandelt". Cf. also Mayrhofer 1983, 150.

B.C. to the fourth century A.D." (1955, 14); some of these words will be discussed below.

A particularly interesting feature of the contact situation between Iranic and Indic in the Indo-Iranian contact zone is that the languages were already quite closely related, and in some cases it would not have been immediately obvious whether a word was Iranic or Indic. One example is the word *kṣaṇa-*/*kṣuna-* in Gandhari. Both forms appear to refer to a particular time. The former is identical to Sanskrit *kṣaṇa-*, and the latter is identical to Khotanese *kṣuṇa-*. The Khotanese word, like the Gandhari word, is probably borrowed from Bactrian χϸονο, referring to a regnal year, which Tremblay (2005, 436) in turn takes to be a borrowing from Greek χρόνος.[20] It is difficult to know whether Gandhari speakers had a clear sense of the distinction.

8.1 Words Related to Writing

In the realm of writing, almost all of the core Indic vocabulary comes from Iranic loanwords, as Falk (2010a) summarises.

- *lipi-* 'writing' (Chatterji 1960, 129; Mayrhofer 1956-80, 3.103; 2001, 443-4; Falk 2010a, 212): from Old Persian *dipi*, which itself is a borrowing (probably from a word for writing originating in Sumerian *dub*, and found in Elamite *tuppi* and Akkadian *ṭuppu*). In Middle Persian and Sogdian, this word was continued by *nipīk*/*nibīg*, which yielded modern Persian *nivē* (Henning 1957). The variation between Gandhari *dipi-* and Sanskrit *lipi-*/*libi-* suggests, as Henning noted, that the word was borrowed into Indian languages from an eastern Iranian dialect that had /ðipi/. We now know that the Bactrian form was λιβο. Hence we do not even have to invoke the analogical influence of the Sanskrit verbal root *lip* 'smear' to account for the forms starting with 'l'.
- *pustaka-* 'book' (Mayrhofer 1956-80, 2.319; 2001, 331; Falk 2010a, 212): borrowed from a Middle Iranic word **pōstaka-* (itself from *pavasta-*, attested in Old Persian) meaning 'hide'. The Middle Iranian word, apparently meaning 'manuscript' or 'document', travelled widely: Sogdian (*pwstk*), Khotanese (*pūstia*), Tocharian (*postak*), Parthian (*pwstg*), and Bactrian (πωσταγο). The word is attested in Gandhari as *postaga-*, probably as a direct loanword from Iranian, and borrowed into Prakrit as *potthaya-*. Sanskrit *pustaka-*, which appears for the first time in 'classical' literature, might be a re-Sanskritisation of the Prakrit form.

20 Bailey (1979, 69) thought that it was inherited from Proto-Indo-Iranian.

- *mudrā-* 'seal' (Chatterji 1960, 128; Mayrhofer 1956-80, 2.654; 2001, 409): found throughout the Iranian languages in this meaning (Bactrian μολρογο and variants 'sealed document', Persian *muhr* 'signet ring', Khotanese *mūra-* 'coin', etc.). The word appears to come from the Old Persian word for Egypt (*mudrāya*), but Mayrhofer notes that Egypt actually borrowed the sealed letter from Western Asia. In any case, with Gandhari *mudra-*, Sanskrit *mudrā-*, and Pali/Prakrit *muddā-* we are evidently dealing with a loanword from an Iranian language.
- *divira-* 'scribe': from Persian **dipīra-* (see above; in Achaemenid documents it is only attested in Elamite *tup-pi-ra*), whence also Modern Persian *dibīr*. Used in Gandhari in the form *tipira*. The parallel Indian and Iranian terms *divīrapati-* and *dabīrbad* were used to designate a 'chief secretary' (Bailey 1949, 127-8; Falk 2010a, 213; von Hinüber 1989, 46).
- *nipista-*/*nipesita-* '(made to be) written': found in the Gandhari version of Aśōka's fourth Rock Edict at Shāhbāzgarhī. From the widely used Persian word *nipišta-* 'written, inscribed' (Falk 2010a, 209).

As Chatterji (1960) noted, many of these words appear to belong to what he designated as the first period of loanwords, viz. the time in which Gandhara was an Achaemenid territory.

8.2 Words Related to Governance

"In the second and third centuries AD, in the times of the Arsacid and Sasanian dynasties, a multitude of Iranian terms came to India, mostly denoting officials" (Falk 2010a, 212). Many of loanwords are limited, in their attestation, to either Gandhari or Sanskrit inscriptions from the Saka-Kuṣāṇa Age. One well-known example is *erzuna-* 'prince', used once in a Gandhari inscription (Takht-i-Bahi, CKI 53), equivalent to Khotanese *alysānai* (Falk 2010b, 76). Another is *ṣāhi-* 'king' and *ṣāhānuṣāhi-* 'king of kings', found in the royal titulature of the Kuṣāṇas, and clearly a continuation of Old Persian *xšāyaθiya-* and *xšāyaθiya- xšāyaθiyānām*. But the title adopted by earlier Saka rulers, *rājātirāja-* (subsequently appearing as *rājādhirāja-*, which became a standard feature of royal titulature in India), appears to be a calque of the same phrase. Two such loanwords, and a set of titles ending in *-pati-*, call for some further comment.

8.2.1 *kṣatrapa-* 'governor'

The Iranian title **xšaθra-pā-* or **xšaθra-pāvan-* 'protecting the realm' was used by regional governors (satraps) in the Achaemenid empire. The forms, presumably Median, are starred because an original **-θr-* is presupposed by the Greek (σατράπης) and Indian borrowings, whereas the Old Persian form, attested in Darius's Behistun inscription, is *xšaçapāvan-*. Its first attestations in India are all in coins and inscriptions of rulers who took this title, mostly of Iranian background, starting in the first century BCE. The title *mahākṣatrapa-* came into use around the same time. It originally designated a subordinate official, as we can see in the plate of Vasa-Abdagases (Falk 1996, 395) of Azes year 9 (ca. 48 BCE), where Vasa-Abdagases is called a *mahakṣatrava-* and Azes is called *maharaja-*. The title later came to be used for independent rulers, around Mathura, Ujjain, and Bharuch (Salomon 1974).[21]

The Gandhari pronunciation of this word would have been almost identical to the Iranian title, and its transparent etymology permitted a straightforward Sanskritisation to *kṣatrapa-* (Chatterji 1960, 129). The Sanskrit word had a relatively long life in coins and inscriptions, being in use as an official title until the end of the Kārdamaka dynasty in the later fourth century (Falk 2010b, 74). However, as Salomon (1974, 15) points out, it is never used in Sanskrit, Prakrit, or Pali literature.

8.2.2 *bhaṭāraka-* 'lord'

Falk (2010b, 75) has suggested a compelling alternative to the prevailing etymology of the word spelled as either *bhaṭṭāraka-* or *bhaṭāraka-* in Sanskrit. Rather than see it as a Middle Indic pronunciation of the Old Indic word *bhartṛ-* (Mayrhofer 1956-10, 464; 2001, 362; Sircar 1966, 52), he sees it as a continuation of an Iranian title *fratara-*, influenced by a folk etymology from *bhartṛ-*. The latter would regularly give *bhaṭṭa-*, which is well attested; the *-āraka-* has remained unexplained.

The Persian word *fratara-* or *fraθara-*, etymologically meaning 'prior', is used as a title in Achaemenid documents in Aramaic (Skjærvø 1997, 102), referring to "the administrative head of a district or province in Egypt" (Wiesehöfer 2012) or "under-satrap" (Wiesehöfer 1991, 306 *apud* Falk 2010b, 78). It appears on the legends of a series of coins

21 Bailey (1949, 127) suggested that an Iranian **xšaθra-pāvan* also underlies the title *cojhbo* (now transcribed *cozbo*) in the Niya documents; it is now taken to be from **čazdahwant* (Tremblay 2005, 429).

of the kings of Persis who broke free of Seleucid rule in the early second century BCE. Skjærvø argued that, in the phrase found on the legends (prtrk' zy 'lhy') "*frataraka* is not exactly the title of the dynasty, but an epithet stating priority of the king among others of divine descent" (1997, 102). When Persis was incorporated into the Parthian empire later in the second century BCE, the king, Mithridates I, allowed the local kings of Persis a degree of autonomy.[22]

As far as I know, the word is not used prior to the first century BCE in Sanskrit, Pali, or Ardhamagadhi texts. Falk (2010b, 75) noted that the earliest use of the title in India might be on a Greek legend of a coin issued by Higaraka, who ruled around the middle of the first century BCE. The legend reads BAPTAP///, which Falk interpreted as *bartar*[*akos*], corresponding to an unattested Gandhari **vhartaraka-*, in turn adapted from *frataraka* with metathesis of the 'r', likely influenced by popular etymology from *bhartṛ-*. The form **vhartaraka-* might have subsequently developed into *bhaṭaraka-*. The length of the vowel in Brāhmī writing (*bhaṭāraka-*) may reflect a stress accent in the Iranian source word.

In the first century CE, *bhaṭaraka-* is found in Gandhari inscriptions as a title, much like *kṣatrapa-*. The copper plates of Helagupta (CKI 564), dated to Azes 121 (i.e. 73/74 CE), seem to refer to Yodavharṇa as a *bhaṭaraka* (so Falk 2014, 21; Salomon 2020 takes it as a proper name). The word *bhaṭaraka* occurs on a sandstone reliquary (CKI 536) dated to Azes 147 (i.e. 99/100 CE), and Baums (2012, 238) takes it there as a title. It remained in use as a title and honorific into Kuṣāṇa times. In the Spinwam inscription (CKI 244), dated to year 39, presumably of the Kuṣāṇa era (hence ca. 166 or 266 CE; cf. Falk 2009, 29), the Kuṣāṇa king is called *bhaṭarakasami* (adopting Falk's readings). A Brāhmī inscription dated to the 45th year of Huviṣka (ca. 195 CE or so), now in the Chhattrapati Shivaji Museum in Mumbai, uses the word *bhaṭārikā* (Lüders 1961, § 180 = 205). The inscription records the donation of an image of Śākyamuni in the Rōśikavihāra at Āḷikā by Khvasicā.[23] The image is dedicated to the good health of a number of people, among whom *bhaṭārikā* figures, either referring to a separate person ("his mistress", as per Lüders), or modifying one of the other people as a title ("the mother of Śamaṇikā, the *bhaṭārikā*").

When we come to the Central Asian kingdom of Niya, around the third century CE, *bhaṭaraka-* was used both for lower officials (*cozbo*s and *soṭhaṁga*s) as well as king Aṁgoka (in the Endere inscription,

22 Cf. Wiesehöfer 2012; 2013 and Wiesehöfer 1994, 105-8 for the title.

23 Khvasicā is probably a Saka name, involving the diminuitive suffix *ica*, attested in Khotanese as *īca* (Degener 1989, 128), perhaps after a word meaning 'first' from **fravišta-*, like Khotanese *hvaṣṭa-* (Bailey 1979, 505).

CKI 363, Salomon 1999a) and his successor Mahiri (Boyer, Rapson, Senart 1920, nos. 415, 573). A number of documents make clear that *bhaṭaraga-* is a complementary term to *daza-* 'slave' or 'servant': *taha na dharma bhaṭaragasa tanu dazasa ṛnena giṁnidavya* (Boyer, Rapson, Senart 1920, no. 24, translated by Burrow 1940 as "such is not the law, that the property of the master should be taken for the debt of the slave"); [*da*]*zajaṁna bhaṭaragasa* (Boyer, Rapson, Senart 1920, no. 31; Burrow 1940: "slave people and the […] people of the master").

By the fourth century or so, the title *bhaṭāraka-* (more often *bhaṭṭāraka-*) had been fully absorbed into Indian royal titulature, where it characterises the Guptas (all *paramabhaṭṭāraka-*) and their queens (all *bhaṭṭārikā-*); it is also found in contemporary inscriptions at Chilas (von Hinüber 1989, 53-4). By this time it had perhaps already spread to Indonesia, where it is always a title of a god, and spelled with the more conservative spelling *bhaṭāraka*.

8.2.3 Words Ending in *-pati*

Around the turn of the common era, we find many words ending in *-pati-* as designations for military, political, or bureaucratic positions. On the one hand, this could be explained by reference to earlier Indic lexical patterns (Vedic *vākpáti-*, *vācáspáti-*, *bṛ́haspáti-*, *bráhmaṇaspáti-*) or indeed lexical patterns of the protolanguages (Vedic *viśpáti-* = Avestan *vispaiti-* < Proto-Indo-Iranian **wić-pati-*; Vedic *dámpati-* = Avestan *dəṇg paiti-* = Greek δεσπότης < Proto-Indo-European **dms-poti-*). On the other hand, titles ending in *-pati-* and its cognates were (and remain) widespread throughout the Iranian world, and many such words were borrowed directly into Indian languages at this time. Here are a few examples:

- *gañjapati-* 'treasurer': from an Eastern Iranian form **ganza-pati-*, like its synonym *gañjavara-* (from **ganza-bara-*, cf. Persian *ganjwar*). Used in "hybrid" Sanskrit inscriptions (Damsteegt 1978, 255), Gandhari (Bailey 1949, 127) and Kashmiri Sanskrit.
- *bakanapati-* 'temple attendant': from **baγanə-pati-*, 'master of the gods'. Used in "hybrid" Sanskrit inscriptions from the Kuṣāṇa period (Damsteegt 1978, 255).
- *haysārapati-* 'chiliarch': a military title attested from the Upper Indus Valley (from **hazāra-pati-*; von Hinüber 1986, 149; Falk 2010b, 78).
- *navhapati-* 'clan-master': used as a title by the kings of Oḍi, from **nāfa-pati-*. The first word, though cognate with Sanskrit *nābhi-* 'navel', refers to a clan or family in Iranic languages (Sodgian and Middle Persian *nāf*). The whole compound was loaned into Armenian as *nahapet* (Bailey 1980, 25; Falk 2010b, 75).

- *divīrapati-* 'chief secretary': found in inscriptions of the fifth century from the Upper Indus and in the colophons of the Gilgit manuscripts (von Hinüber 1986, 149), as well as in Kashmiri Sanskrit (the *Rājataraṅgiṇī*). Based on the word *divīra* 'scribe' or 'secretary' discussed above; compare Persian *dabīrbad*.
- *haṁmārapati-* 'chief accountant': in inscriptions of the Upper Indus and in the colophons of the Gilgit manuscripts (von Hinüber 1986, 150).
- *naśćīrapati-* 'hunt-master' (Mayrhofer 2001, 285 compares Parthian ⟨nhšyrpty⟩ /naxcir-pati/ 'Jagdmeister' and Middle and New Persian *naxčīr* 'Jagd'); attested in Kuṣāṇa-era inscriptions.

These words are clearly loans from Iranic languages, given the first element. I pass over a number of other words ending in *-pati-* found in Kuṣāṇa-era inscriptions, since their interpretation and etymology is not secure, but these are likely loans from Iranic languages as well (*kharāsalērapati-*, *manapākapati-*; Falk 2010b, 78).

The word *dānapati-* 'master of giving' i.e. 'patron' is attested in Sanskrit from the *Mahābhārata* onwards. Lüders (1961, 95-6) suggested that this word has an Iranian equivalent attested in Brāhmī inscriptions as *hōramurṇḍaga-* (with much variation) and in Kharōṣṭhī inscriptions as *horamurta*, if this word means 'master of gifts' (cf. Khotanese *hōra-* 'gift'). There are some problems with this interpretation, as Lüders himself admitted. But I am inclined to see this, and a number of Sanskrit words ending in *-pati-*, as calques of Iranian titles.[24] In other words, the use of the element *pati* in compounds referring to the 'chief' person in a particular role or office corresponds with, and is probably influenced by, the use of the etymologically-identical element *bed* (*vel sim.*) in Iranic languages. This tendency would of course have been reinforced by earlier Sanskrit compounds ending in *pati*, such as *sēnāpati-* (first attested in the *Aitarēyabrāhmaṇa*).

One problematic example is the word *sthapáti-* 'architect', which is first attested in the Atharvavēda. It presents certain problems of derivation as a Sanskrit word (namely the use of *stha-* as the initial element of a compound), and hence Mayrhofer (1996, 764) suggests that it may be an "Umformung eines [iran.?] LW", probably on the basis of other Iranian loanwords in *-pati-*. If Mayrhofer is right, this would indicate that the process of borrowing such words from neighbouring Iranic languages had already begun prior to the Saka-Kuṣāṇa Age.

24 I do not include *dārapati-*, which von Hinüber (1986, 149) suggested to read in several inscriptions from the Upper Indus, since he subsequently changed the reading to *dānapati-* (1989, 56). Still, the context of the word in these inscriptions (alongside Iranian terms such as *divīra-*) suggests it comes from the sphere of Iranian influence.

8.3 Military Terms

When scanning the third volume of Mayrhofer's *Etymologisches Wörterbuch des Altindoarischen* I noted a relatively large number of words that pertain to military equipment and roles, which I will merely list here, with the corresponding pages in Mayrhofer (2001); the references to cognates are his; I have tried to find the earliest attestations.

- *aśvavāra-* 'horse-rider' (p. 18) (Old Persian *asabāra*, East Middle Iranic **aspa-βāra-*). On this word, first attested in label inscriptions at Bharhut (second century BCE, Chatterji 1960, 129), cf. especially Morgenstierne 1974, 275-6.
- *khōla-* 'helmet' (p. 148) (Bāṇa's *Kādambarī* and *Harṣacarita*, seventh century) (Old Persian *xaudā-*, Avestan *xaoδa-*, East Middle Iranic **xōla-*).
- *tīrī-* 'arrow' (p. 248) (Halāyudha's *Abhidhānacintāmaṇi*, tenth century) (Persian *tīr*, from Old Persian *tigra-*).
- *druṇā-* 'bow' (p. 274) (Halāyudha) (Middle Persian *drōn*, Khotanese *durna-*).
- *nipaka-* 'pledge' or 'hostage' (p. 291) (*Divyāvadāna*, third century) (Sogdian ⟨npq⟩ , Khotanese *nvi-*); cf. Bailey 1955, 18; the Kashmiri Sanskrit word *nīvī*, used in the sense of 'hostage' in the *Rājataraṅgiṇī*, is probably also related.
- *paryāṇa-* 'saddle' (pp. 307-8) (Varāhamihira, sixth century) (Sogdian *pyrδn*); Bailey 1955, 14.
- *padāti(ka)-* (Mayrhofer 1996, 79): attested already in the *Taittirīyabrāhmaṇa*, which evidently made Mayrhofer reluctant to consider it an Iranic loanword outright despite the close parallel with Middle Persian *payādag* and Persian *piyāda*; Morgenstierne has "no doubt" that the word was borrowed from Persian (1974, 275: 8).
- *padika-* 'footsoldier' (p. 303) in the *Amarakōśa*: derivable from Sanskrit *pada-*, but Mayrhofer mentions the possibility that it might come from Iranian **padik*, continued by Persian *payg* 'footsoldier' (see above).
- *lastaka-* 'bow-grip' (p. 441) in the *Amarakōśa* ("Veilliecht iran. *δasta(k) *'Griff', *'Handstelle'")
- *vārabāṇa-* 'cuirass' (p. 467) in the *Amarakōśa* ("Iran. **varo-pāna* 'Brustschutz'").

8.4 Names

There are of course many Iranian names attested in Indian inscriptions and manuscripts during the Saka-Kuṣāṇa Age, and in certain areas of the Indo-Iranian border regions (e.g. Gilgit) they continued to appear alongside Indian names well into the first millennium. There is no point in listing the names here, but I did want to make two points about the impact of Iranic languages on the Indian onomasticon.

First, in a multilingual and multiethnic culture, it might not have been obvious which elements are 'Indic' and which elements are 'Iranic' (Morgenstierne's 1974, 271). This is especially true of cognate elements, such as the words derived from Proto-Indo-Iranian **priHa-*: Indic *priya-* and Iranic *friya-* must have sounded very similar. Hence it is unsurprising to have names such as Indrafriya (*iṁtavhria* in CKI 60) in Gandhara, where we might have expected Indrapriya. But even elements such as *spāla-*, the Eastern Iranic word for 'army', were essentially treated as Indic words, the exact equivalent of *sēnā-*, in names such as Suśpāla (= Suṣēṇa; Falk 2006, 396) and Dharmaspāla (= Dharmasēna; von Hinüber 1986, 151). The emergence of Sanskrit as a 'cosmopolitan' language in the period immediately following the Saka-Kuṣāṇa Age might explain the disappearance of Iranic elements from the onomasticon, both because the formal study of the Sanskrit language would have made their foreignness more apparent, and because names were increasingly regulated by the norms of both Sanskrit grammar and the various *dharmaśāstra*s. Consider, for example, the names of the Kārdamaka kings of Ujjayinī: the first few are completely Iranic (Zamōtika, Caṣṭana), and then they switch over to Indic names (Jayadāman, Rudradāman), with a few ambiguously Iranic elements (Dāmazāda = Dāmajāta; Tandon 2009).

Second, it appears that many Iranic names were given an *interpretatio indica*, and appear in this form in Sanskrit texts. I found all of these in Mayrhofer's *Etymologisches Wörterbuch*; more could certainly be found.

Parṇadatta, governor of Saurāṣṭra mentioned in Skandagupta's Junāgaṛh inscription, is probably a Sanskritisation of the common Iranic name Farnadāta (Charpentier 1931; Mayrhofer 2001, 306). *Parṇa-* stands in for *farna-* also, evidently, in Ṛtuparṇa = *Ṛtafarnah, a king of Kośala (Mayrhofer 2001, 38). As in the case of Parṇadatta, Indic *datta-* probably corresponds to Iranic *dāta-* in the name of a character in the *Mahābhārata,* Bhagadatta = *Bagadāta (Mayrhofer 2001, 360), since the theophoric element more naturally represents Iranic *baga-* 'god' than Indic *bhaga-*, which might suggest something untoward. Similarly, in the *Mahābhārata*, the king of Sauvīra in the Lower Indus Valley is called Dattāmitra, recognizable as the common Iranic name Dātamiθra (compare Mithradates, Miθradāta; Mayrhofer 2001, 259), the equation being more likely given the importance of

Mithra as an Iranian god. Mayrhofer also noted (281) that the name of Sahadēva's son, Dhruvāśva, might represent an attested Iranic name Druwasp (Λροοασπο in Bactrian and *druuaspa* in Avestan). Finally, Mēghākṣa, the Persian king allied to Malayakētu in the *Mudrārākṣasa*, probably contains the element **xšaya-* 'rule' (Mayrhofer 2001, 414, who connects the first element to Pamir *mēγ* 'name of a mythical horse').

9 Conclusions

I expect that my argument about the linguistic influence of Middle Iranic languages on the Indic languages of the Northwest will be sufficiently clear by now. I will conclude, then, with three features of the analysis offered here that might have broader historical or methodological implications.

First, contact is an important feature in the history of language, and a major focus of linguistic research in the present day. But the study of historical languages, that is, languages to which we have access only through written records, approaches linguistic change by default through the model of accumulated changes over time that differentiate one speech community from another, i.e. the *Stammbaum* model. The study of contact in historical languages, and especially in the historical languages of South Asia, is much less developed than the study of neogrammarian sound change, and it often figures in 'marginal' cases where words or forms cannot be explained through neogrammarian sound change; there is a whole genre of Indological research attempting to identify words in Sanskrit (and rarely other languages) that have come from Dravidian, Munda, or other 'substrate' languages.

There are good reasons for this imbalance, of course. We have neither the plenum of evidence that would tell us, for example, about whether and how the linguistic practices of communities in contact differ from each other. Nor do we have the kind of evidence that would tell us, unambiguously, how certain forms were pronounced, and indeed as we have seen with the Kharōṣṭhī script, there are a number of different phonological interpretations of the orthography. Nevertheless, the linguistic circumstances prevalent in South Asia – where 'linkages' of related languages are likely to be found, and where areal phenomena have already been documented – should invite us to consider other features of the historical languages as possibly resulting from contact-induced change. One parallel case is the change of 's' to 'h' in Greek, Armenian, and Iranian (Parpola 2002, 82).

Second, philology – here narrowly understood as the study of historical languages – can and should avail itself of new concepts and methods. By 'new' I do not mean to suggest that prosodic phonology,

which I have used in this paper, is new, but only that its application to the study of historical languages is still limited. In fact prosodic phonology in particular can be an important tool for detecting contact phenomena, precisely because the 'same' language can be pronounced with different prosodic features by different communities of speakers. Kümmel's work (2014) correlating the features of Indic and Iranic languages with the parameters of syllable and word languages, respectively, is exemplary in this respect.

Third and finally, Apabhramsha figures in this analysis as the 'bridge' whereby sound changes produced by contact in a specific region at a specific time continued to be represented in the literary record of India in the later first millennium. That is, as Gandhari was superceded as a literary language by Sanskrit, and as Middle Iranic languages receded from the linguistic horizons of India, Apabhramsha remained as the single surviving Indic language with an Iranic phonology. Now the *Sprachwirklichkeit* of Apabhramsha is a large and complex issue. Most philologists have seen Apabhramsha as a somewhat crystallised or frozen form of a popular language (Bubeník 1998), and some have seen the 'underlying' popular language(s) as very widely distributed vernaculars that were the antecedent to the modern vernaculars of North India (cf. Ghosal 1956). Whatever spoken language(s) to which Apabhramsha was connected might, in any analysis, have served as the 'vector' by which a number of important sound changes entered the linguistic communities of North India. The loss of final vowels is of course found in all of the modern North Indian vernaculars, and for that reason it might seem attractive to analyse it as the result of tendencies or pressures internal to the Indic languages. But it was not inevitable of course. If we require an explanation for it, the long-term influence of FSR from Iranic languages could be considered.

Bibliography

Primary Texts

Gandhari documents and inscriptions are cited by their CKI or CKD number on `http://gandhari.org` (maintained by Stefan Baums and Andrew Glass).

Bhāmaha (6th c. CE?). "Ornament of Literature (Kāvyālaṅkāra)". Trivedî, K.P. (ed.) (1909), *The Pratâparudrayośobhûshaṇa of Vidyânâtha. With the Commentary, Ratnâpaṇa, of Kumârasvâmin, Son of Mallinâtha, and With a Critical Notice of Manuscripts, Introduction, Critical and Explanatory Notes and an Appendix Containing the Kâvyâlaṅkâra of Bhâmaha*. Bombay: Government Central Press.

Daṇḍin (ca. 700 CE). "Mirror of Literature (Kāvyādarśa)". Thakur, A.; Jha, U. (eds) (1957), *Kavyalakṣaṇa of Daṇḍin (Also Known as Kāvyādarśa) with Commentary Called Ratnaśrī of Ratnaśrījñāna*. Darbhanga: Mithila Institute of Post-Graduate Studies and Research in Sanskrit Learning.

Hēmacandra (12th c. CE). "Siddhahēmacandraśabdānuśāsana". Pischel, R. (Hrsg.) (1877), *Hemacandra's Grammatik der Prâkritsprachen (Siddhahemacandram Adhyâya VIII), mit kritischen und erläuternden Anmerkungen*. Halle: Waisenhaus.

Kramadīśvara (s.d.). "Distilled Essence (Saṁkṣiptasāra)". Banerjee, S.R. (ed.) (1980), *Prākṛtādhyāya (The Prākṛta Book of the Samkṣiptasāra) by Kramadīśvara*. Ahmedabad: Prakrit Text Society. Prakrit Text Series 22.

Mārkaṇḍēya (1560s CE). "Sum-Total of Prakrit (Prākɼtasarvasva)". Acharya, K.C. (ed.) (1968), *Mārkaṇḍeya's Prākṛta-Sarvasva*. Ahmedabad: Prakrit Text Society. Prakrit Text Society Series 11.

Puruṣōttamadēva (12th c. CE). "Prakrit Grammar (Prākɼtānuśāsana)". Nitti-Dolci, L. (éd.) (1938), *Le Prākṛtānuśāsana de Puruṣottama*. Paris: Société Asiatique.

Rāmaśarman (s.d.). "Wish-Granting Tree of Prakrit (Prākɼtakalpataru)". Ghosh, M. (ed.) (1954), *Rāmaśarman's Prākṛta-kalpataru*. Calcutta: The Asiatic Society. Bibliotheca Indica 278.

Uddyōtana (778 CE). "Kuvalayamālā". Upadhye, A.N. (ed.) (1959), *Uddyotanasūri's Kuvalayamālā (A Unique Campū in Prakrit)*. Part I, *Kuvalayamālā, Prākrit-Text & Various Readings*. Bombay: Bhāratīya Vidyā Bhavana. Singhi Jain Series 45.

Secondary Literature

Auer, P. (1993). *Is a Rhythm-Based Typology Possible? A Study of the Role of Prosody in Phonological Typology*. Konstanz: Universität Konstanz. KontRI Working Paper 21. `http://paul.igl.uni-freiburg.de/auer/userfiles/downloads/Phonotypo_Kontri1.pdf`.

Back, M. (1978). *Die sassanidischen Staatsinschriften*. Téhéran: Bibliothèque Pahlavi.

Bailey, H.W. (1946). "Gāndhārī". *Bulletin of the School of Oriental and African Studies*, 11(4), 764-7. `https://doi.org/10.1017/S0041977X00089801`.

Bailey, H.W. (1949). "Indo-Iranica II". *Bulletin of the School of Oriental and African Studies*, 13(1), 121-39. `https://doi.org/10.1017/S0041977X0008188X`.

Bailey, H.W. (1955). "Buddhist Sanskrit". *The Journal of the Royal Asiatic Society of Great Britain and Ireland*, 87(1-2), 13-24. https://doi.org/10.1017/S0035869X00106975.

Bailey, H.W. (1963). *Khotanese Texts V*. Cambridge: Cambridge University Press.

Bailey, H.W. (1979). *Dictionary of Khotan Saka*. Cambridge; New York: Cambridge University Press.

Bailey, H.W. (1980). "A Kharoṣṭrī Inscription of Seṇavarma, King of Oḍi". *The Journal of the Royal Asiatic Society of Great Britain and Ireland*, 112(1), 21-9. https://doi.org/10.1017/S0035869X00135877.

Bashir, E. (ed.) (2016). "Contact and Convergence: The Northwest". Hock, H.H.; Bashir, E. (eds), *The Languages and Linguistics of South Asia: A Comprehensive Guide*. Berlin: De Gruyter Mouton, 264-99.

Bashir, E. (2006). "Indo-Iranian Frontier Languages". *Encyclopædia Iranica*. https://iranicaonline.org/articles/indo-iranian-frontier-languages-and-the-influence-of-persia.

Baums, S. (2009). *A Gāndhārī Commentary on Early Buddhist Verses: British Library Kharoṣṭhī Fragments 7, 9, 13 and 18* [PhD dissertation]. Seattle: University of Washington.

Baums, S. (2012). "Catalog and Revised Texts and Translations of Gandharan Reliquary Inscriptions". Jongeward, D.; Errington, E.; Salomon, R.; Baums, S. (eds), *Gandharan Buddhist Reliquaries*. Seattle: Early Buddhist Manuscripts Project, 200-51.

Bhayani, H.C. (1998a). "On Early Apabhraṁśa". *Indological Studies: Literary and Performing Arts, Prakrit and Apabhraṁśa Studies*, vol. 2. Ahmedabad: Parshva Publication, 35-43.

Bhayani, H.C. (1998b). "The Rise of Apabhraṁśa as a Literary Language". *Indological Studies: Literary and Performing Arts, Prakrit and Apabhraṁśa Studies*, vol. 2. Ahmedabad: Parshva Publication, 26-34.

Bhayani, H.C. (1989). *Apabhraṃśa Language and Literature: A Short Introduction*. Delhi: Bhogilal Leherchand Institute of Indology.

Blevins, J. (2008). "Natural and Unnatural Sound Patterns: A Pocket Field Guide". Willems, K.; De Cuypere, L. (eds), *Naturalness and Iconicity in Language*, 121-48.

Boyer, A.M.; Rapson, E.J.; Senart, E. (1920). *Kharoṣṭhī Inscriptions Discovered by Sir Aurel Stein in Chinese Turkestan. Part I, Text of Inscriptions Discovered at the Niya Site*. Oxford: Clarendon Press.

Brough, J. (1962). *The Gāndhārī Dharmapada*. London: Oxford University Press.

Bubeník, V. (1998). *A Historical Syntax of Late Middle Indo-Aryan (Apabhraṃśa)*. Amsterdam; Philadelphia: John Benjamins.

Burrow, T. (1940). *A Translation of the Kharoṣṭhi Documents from Chinese Turkestan*. London: The Asiatic Society.

Caillat, C. (1992). "Connections Between Asokan (Shahbazgarhi) and Niya Prakrit?". *Indo-Iranian Journal*, 35(2-3), 109-19.

Cantera, A. (2009). "On the History of the Middle Persian Nominal Inflection". Sundermann, W.; Hintze, A.; de Blois, F. (eds), *Exegisti monumenta: Festschrift in Honour of Nicholas Sims-Williams*. Wiesbaden: Harrassowitz, 17-30.

Charpentier, J. (1931). "Pherendates–Parṇadatta". *Journal of the Royal Asiatic Society*, 63(1), 140-1. https://doi.org/10.1017/S0035869X00066296.

Chatterji, S.K. (1960). "Some Iranian and Turki Loans in Sanskrit". Dil, A.S. (ed.), *Shahidullah Presentation Volume*. Lahore: Linguistic Research Group of Pakistan, 123-40.

Damsteegt, T. (1978). "Epigraphical Hybrid Sanskrit: Its Rise, Characteristics and Relationship to Buddhist Hybrid Sanskrit". *Orientalia Rheno-Traiectina*, vol. 23. Leiden: Brill.

Degener, A. (1989). *Khotanische Suffixe.* Wiesbaden: Steiner.

Dresher, B.E.; Lahiri, A. (1991). "The Germanic Foot: Metrical Coherence in Old English". *Linguistic Inquiry*, 22(2), 251-86.

Emeneau, M.B. (1956). "India as a Linguistic Area". *Language*, 32(1), 3-16.

Emeneau, M.B. (1980). "The Indian Linguistic Area Revisited". Dil, A.S. (ed.), *Language and Linguistic Area: Essays by Murray B. Emeneau*. Stanford: Stanford University Press, 197-249.

Emmerick, R.E. (1987). "The Transition from Old to Late Khotanese". *Transition Periods in Iranian History = Actes du Symposium de Fribourg-en-Brisgau* (22-24 mai 1985). Leuven: Association pour l'avancement des études iraniennes, 33-42.

Emmerick, R.E. (2009). "Khotanese and Tumshuqese". Windfuhr, G. (ed.), *The Iranian Languages*. Milton Park: Routledge, 377-415.

Falk, H. (1996). "Three Inscribed Buddhist Monastic Utensils from Gandhāra". *Zeitschrift der Deutschen Morgenländischen Gesellschaft*, 156, 393-412.

Falk, H. (2006). *Aśokan Sites and Artefacts: A Source-Book with Bibliography*. Mainz: Von Zabern.

Falk, H. (2009). "The Pious Donation of Wells in Gandhara". Mevissen, G.; Banerji, A. (eds), *Prajñādhara: Essays on Asian, History, Epigraphy and Culture in Honour of Gouriswar Bhattacharya*. New Delhi: Kaveri Books, 23-6.

Falk, H. (2010a). "Foreign Terms in Sanskrit Pertaining to Writing". de Voogt, A.; Finkel, I. (eds), *The Idea of Writing: Play and Complexity*. Leiden: Brill, 207-17.

Falk, H. (2010b). "Names and Titles from Kuṣāṇa Times to the Hūṇas". Alram, M.; Klimburg-Salter, D.; Inaba, M.; Pfisterer, M. (eds), *Coins, Art and Chronology II: The First Millennium C.E. in the Indo-Iranian Borderlands*. Vienna: Österreichische Akademie der Wissenschaften, 73-90.

Falk, H. (2014). "The First-Century Copper-Plates of Helagupta from Gandhāra hailing Maitreya". *Annual Report of the International Research Institute for Advanced Buddhology at Soka University*, 17, 3-26.

François, A. (2015). "Trees, Waves and Linkages: Models of Language Diversification". Bowern, C.; Evans, B. (eds), *The Routledge Handbook of Historical Linguistics*. Abingdon: Routledge, 161-89. https://doi.org/10.4324/9781315794013.

Fussman, G. (1989). "Gāndhārī écrite, Gāndhārī parlée". Caillat, C. (éd.), *Dialectes dans les littératures indo-aryennes*. Paris: Collège de France, 433-98.

Gershevitch, I. (1954). *A Grammar of Manichean Sogdian*. Oxford: Blackwell.

Gholami, S. (2014). "Historical phonology of Bactrian". *Selected Features of Bactrian Grammar*. Wiesbaden: Harrassowitz, 31-71.

Gholami, S. (2009). *Selected Features of Bactrian Grammar* [PhD Dissertation], Georg-August-Universität Göttingen: Göttingen.

Ghosal, S.N. (1956). "Controversy Over the Significance of Apabhraṃśa and a Compromise Between the Views of Jacobi and Grierson". *Journal of the Asiatic Society. Letters*. 12, 23-37.

Glass, A. (2000). *A Preliminary Study of Kharoṣṭhī Manuscript Paleography* [MA Thesis]. Seattle: University of Washington.

Grierson, G. (1924). *The Prakrit Dhātv-ādeśas According to the Western and the Eastern Schools of the Prakrit Grammarians*. Calcutta: Asiatic Society of Bengal.

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

Henning, W.B. (1957). "The Inscriptions of Tang-i Azao". *Bulletin of the School of Oriental and African Studies*, 20, 335-42.

Hintze, A. (2006). Review of *Le Y final dans les inscriptions moyen-perses et la 'loi rythmique' proto-moyen-perse*, by Huyse, P. *Kratylos*, 51, 161-7. http://dx.doi.org/10.29091/kratylos/2006/1/13.

von Hinüber, O. (1981). "Die Paiśācī und die Entstehung der Sakischen Orthographie". Bruhn, K.; Wezler, A. (Hrsgg), *Studien zum Jainismus und Buddhismus: Gedenkschrift für Ludwig Alsdorf*. Wiesbaden: Franz Steiner, 121-7.

von Hinüber, O. (1986). "Zu einigen iranischen Namen und Titeln aus Brāhmī-Inschriften am oberen Indus". *Studia Grammatica Iranica: Festschrift für Helmut Humbach*. München: R. Kitzinger, 147-62.

von Hinüber, O. (1989). "Brāhmī Inscriptions on the History and Culture of the Upper Indus Valley". Jettmar, K. (ed.), *Antiquities of Northern Pakistan: Reports and Studies (Band 1): Rock Inscriptions in the Indus Valley*. Mainz: Philipp von Zabern, 41-72.

Hitch, D. (2016). *The Old Khotanese Metanalysis* [PhD dissertation]. Cambridge (MA): Harvard University.

Huyse, P. (2003). *Le y final dans les inscriptiosn moyen-perses et la 'loi rythmique' proto-moyen-perse*. Paris: Association pour l'avancement des études iraniennes.

Ivani, J.K.; Paudyal, N.; Peterson, J. (2021). "Indo-Aryan – A House Divided? Evidence for the East-West Indo-Aryan Divide and its Significance for the Study of Northern South Asia". *Journal of South Asian Languages and Linguistics*, 7(2), 185-202. https://doi.org/10.1515/jsall-2021-2029.

Jamison, S. (2009). "Sociolinguistic Remarks on the Indo-Iranian *-ka-Suffix: A Marker of Colloquial Register". *Indo-Iranian Journal*, 52, 311-29. http://dx.doi.org/10.1163/001972409x12562030836615.

Kielhorn, F. (1905-06). "Junagadh Inscription of Rudradaman". *Epigraphia Indica*, 8, 36-49.

Konow, S. (1929). *Kharoshṭhī Inscriptions, With the Exception of Those of Aśoka*. Calcutta: Government of India.

Korn, A. (2013). "Final Troubles: Armenian Stem Classes and the Word-end in Late Old Persian". Tokhtasev, S.; Luria, P. (eds), *Commentationes Iranicae: Vladimiro f. Aaron Livschits nonagenario donum natalicum*. Petersburg: Nestor-Historia. 74-91.

Korn, A. (2019). "Isoglosses and Subdivisions of Iranian". *Journal of Historical Linguistics*, 9, 239-81. https://doi.org/10.1075/jhl.17010.kor.

Kümmel, M. (2008). *Khotansakisch* [Handout]. https://www.academia.edu/1748443/Introduction_to_Khotan_Saka_in_German_.

Kümmel, M. (2014a). "Syllable- and Word-related Developments in Earlier Indo-Iranian." Pre-print version of Kümmel 2014b. https://www.academia.edu/379001/Syllable_and_word_related_developments_in_earlier_Indo_Iranian.

Kümmel, M. (2014b). "Syllable- and Word-Related Developments in Earlier Indo-Iranian". Reina, J.C.; Szczepaniak, R. (eds), *Syllable and Word Languages*. Berlin: De Gruyter, 204-21.

Kümmel, M. (2018). "Areal Developments in the History of Iranic: West vs. East". https://ia801207.us.archive.org/15/items/dokumen.tips_areal-developments-in-the-history-of-iranic-west-vs-east/

dokumen.tips_areal-developments-in-the-history-of-iranic-west-vs-east.pdf.

Loukota, D. (2022). "'Ne Hāḍe Vajrropamä Vaśärä': Indic Loanwords in the Khotanese Book of Zambasta and the Chronology of the Spread of Buddhism to Khotan". *Buddhist Road: Dynamics in Buddhist Networks in Eastern Central Asia 6th-14th Centuries*, 1(7), 3-32.

Lévi, S. (1903). "Notes on the Indo-Scythians". *Indian Antiquary*, 32, 381-9, 417-26.

Lüders, H. (1961). *Mathurā Inscriptions*. Göttingen: Vandenhoeck & Ruprecht.

MacKenzie, D.N. (1982). Review of *Die sassanidischen Staatsinschriften*, by Back, M. *Indogermanische Forschungen*, 87, 280-98.

Masica, C. (1976). *South Asia: Defining a Linguistic Area*. Chicago: University of Chicago Press.

Mayrhofer, M. (1956-80). *Kurzgefasstes etymologisches Wörterbuch des Altindoarischen*. Heidelberg: Carl Winter.

Mayrhofer, M. (1983). "Überlegungen zu einem neuen etymologischen Wörterbuch des Altindoarischen". Bammesberger, A. (Hrsg.), *Das etymologische Wörterbuch: Fragen der Konzeption und Gestaltung*. Regensburg: Friedrich Pustet, 147-54.

Mayrhofer, M. (1992, 1996, 2001). *Etymologisches Wörterbuch des Altindoarischen*, 3 Bde. Heidelberg: Carl Winter.

Morgenstierne, G. (1974). "Early Iranic Influence upon Indo-Aryan". *Commémoration Cyrus. Hommage Universel*, vol. 1. Téhéran; Liège: Bibliothèque Pahlavi, 271-9.

Morgenstierne, G. (1975). "Ancient Contacts Between N.E. Iranian and Indo-Aryan?". *Mélanges linguistiques offerts à Émile Benveniste*. Paris: Société Linguistique de Paris, 431-4.

Müller, M. (1883). *India: What Can It Teach Us?*. London: Lovell's Library.

Myers, S.; Hansen, B.D. (2006). "The Origin of Vowel Length Neutralization in Final Position: Evidence from Finnish Speakers". *Natural Language and Linguistic Theory*, 25(1), 157-93.

Nespor, M.; Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.

Novák, Ľ. (2013). *Problem of Archaism and Innovation in the Eastern Iranian Languages* [PhD dissertation]. Prague: Charles University.

Parpola, A. (2002). "From the dialects of Old Indo-Aryan to Proto-Indo-Aryan and Proto-Iranian". Sims-Williams, N. (ed.), *Indo-Iranian Languages and Peoples*. Oxford: Oxford University Press, 43-102.

Salomon, R. (1974). "The Kṣatrapas and Mahākṣatrapas of India". *Wiener Zeitschrift für die Kunde Sudasiens*, 18, 5-25.

Salomon, R. (1999a). "A Stone Inscription in Central Asian Gāndhārī from Endere (Xinjiang)". *Bulletin of the Asia Institute*, 13, 1-13.

Salomon, R. (1999b). *Ancient Buddhist Scrolls from Gandhāra*. London: The British Library.

Salomon, R. (2002). "Gāndhārī and the Other Indo-Aryan Languages in the Light of Newly Discovered Kharoṣṭhī Manuscripts". Sims-Williams, N. (ed.), *Indo-Iranian Languages and Peoples*. Oxford: Oxford University Press, 119-34. Proceedings of the British Academy 136. https://doi.org/10.5871/bacad/9780197262856.003.0005.

Salomon, R. (2018). *The Buddhist Literature of Ancient Gandhara*. Somerville (MA): Wisdom.

Salomon, R. (2020). "The Copper Plates of Helagupta: A New Edition and Study". *Indo-Iranian Journal*, 63, 3-69. http://dx.doi.org/10.1163/15728536-06301006.

Sims-Williams, N. (1984). "The Sogdian 'Rhythmic Law'". *Middle Iranian Studies: Proceedings of the International Symposium Organized by the Katholieke Universiteit Leuven from the 17th to the 20th of May 1982*. Leuven: Peters, 203-15.

Sims-Williams, N. (1990). "Chotano-Sogdica II: Aspects of the Development of Nominal Morphology in Khotanese and Sogdian". Gnoli, G.; Panaino, A. (eds), *Proceedings of the First European Conference of Iranian Studies*, vol. 1. Rome: Istituto italiano per il Medio ed Estremo Oriente, 275-96.

Sims-Williams, N. (1996). "The Sogdian Manuscripts in Brāhmī Script as Evidence for Sogdian Phonology". Emmerick, R.E. (ed.), *Turfan, Khotan und Dunhuang: Vortrage der Tagung Annemarie v. Gabain und die Turfanforschung', veranstaltet von der Berlin-Brandenburgischen Akademie der Wissenschaften*. Berlin: Akademie Verlag, 307-15.

Sims-Williams, N. (2011). "Bactrian Language". *Encyclopædia Iranica*, 3(4), 344-9. https://www.iranicaonline.org/articles/bactrian-language.

Sircar, D.C. (1966). *Indian Epigraphical Glossary*. Delhi: Motilal Banarsidass.

Skjærvø, P.O. (1997). "The Joy of the Cup: A Pre-Sasanian Middle Persian Inscription on a Silver Bowl". *Bulletin of the Asia Institute*, 11, 93-104.

Skjærvø, P.O. (2007). "An Introduction to Manichean Sogdian". https://web.archive.org/web/20150106094817/http://www.fas.harvard.edu/~iranian/Sogdian/sogdiancomplete.pdf.

Smith, L.C.; Ussishkin, A. (2015). "The Role of Prosodic Templates in Diachrony". Honeybone, P.; Salmons, J. (eds), *The Oxford Handbook of Historical Phonology*. Oxford: Oxford University Press, 262-86. https://doi.org/10.1093/oxfordhb/9780199232819.013.028.

Strauch, I. (2012). "The Character of the Indian Kharoṣṭhī Script and the 'Sanskrit Revolution': A Writing System Between Identity and Assimilation". de Voogt, A.; Quack, J.F. (eds), *The Idea of Writing: Writing Across Borders*. Leiden: Brill, 131-68.

Tagare, G.V. (1948). *Historical Grammar of Apabhraṁśa*. Poona: Deccan College Post-Graduate and Research Institute.

Tandon, P. (2009). "The Western Kshatrapa Dāmazāda". *The Numismatic Chronicle*, 169, 173-87.

Tieken, H. (1998). "The Distribution of the Variant Endings -u, -aü, and -ā in the Apabhraṃśa Verses in Hemacandra's Prākrit Grammar". *Indo-Iranian Journal*, 41(1), 1-31.

Tikkanen, B. (2008). "Some Areal Phonological Isoglosses in the Transit Zone between South Asia and Central Asia". Ud-din, I. (ed.), *Proceedings of the Third International Hindu-Kush Cultural Conference*. Oxford: Oxford University Press, 250-62.

Tremblay, X. (2005). "Irano-Tocharica and Tochario-Iranica". *Bulletin of the School of Oriental and African Studies*, 68, 421-49. https://doi.org/10.1017/S0041977X05000248.

Turner, R.L. (1927). "The Phonetic Weakness of Terminational Elements in Indo-Aryan". *The Journal of the Royal Asiatic Society of Great Britain and Ireland*, 59(2), 227-39. https://doi.org/10.1017/S0035869X00057439.

Vaissière, J. (1996). "From Latin to Modern French: On Diachronic Changes and Synchronic Variations". *AIPUK. Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel,* 31, 61-74.

Wiesehöfer, J. (1991). "PRTRK, RB μYL', SGN und MR". Sancisi-Weerdenburg, H.; Kuhrt, A. (eds), *Asia Minor and Egypt: Old Cultures in a New Empire = Proceedings of the Groningen 1988 Achaemenid History Workshop*. Leiden: Nederlands Instituut voor het Nabije Oosten, 305-9.

Wiesehöfer, J. (1994). *Die'dunklen Jahrhunderte' des Persis: Untersuchungen zu Geschichte und Kultur von Fārs in frühhellenisticher Zeit (330-140 v. Chr.)*. München: Beck.

Wiesehöfer, J. [2000] (2012). "Frataraka". *Encyclopædia Iranica*. https://iranicaonline.org/articles/frataraka.

Wiesehöfer, J. (2013). "Fratarakā and Seleucids". Potts, D.T. (ed.), *The Oxford Handbook of Ancient Iran*. Oxford: Oxford University Press, 718-27.

Withgott, M. (1982). *Segmental Evidence for Phonological Constituents* [PhD Dissertation]. Austin: The University of Texas at Austin.

Zoller, C.P. (2018). "'Pagan Christmas: Winter feast of the Kalasha of the Hindu Kush' and the true frontiers of 'Greater Peristan'". *Acta Orientalia*, 79, 163-337.

Zoller, C.P. (2023). *Indo-Aryan and the Linguistic History and Prehistory of North India*. Wiesbaden: Harrassowitz. Neuindischen Studien 20. https://doi.org/10.5617/ao.7672.

e-ISSN 2785-5953

Bhasha
Vol. 3 – Num. 1 – April 2024

The Sanskrit Paradigm of Tamil Grammar: Embrace and Resistance

E. Annamalai
Central Institute of Indian Languages, India; The University of Chicago, USA

Abstract This paper is about the historical relationship between the grammars of Sanskrit and Tamil and the debates around it. The debates are between the grammarians and commentators on Tamil grammar, not of Sanskrit. It is thus limited to one linguistic community. This is a result of the asymmetrical bilingualism of the Tamil-speaking and the Sanskrit-speaking grammarians. Their many other relationships – religious, literary, etc. – are also asymmetrical in the sub-continental perspective. The regional perspective of Tamil about the relationship of Tamil to Sanskrit is different but is not constant and varies according to the period and its language ideology.

Keywords Sanskrit grammar paradigm. Tamil grammar tradition. Compound formation. Universal grammar. Language autonomy.

Edizioni Ca'Foscari

Peer review

Submitted 2023-03-30
Accepted 2023-09-23
Published 2024-03-27

Open access

Citation Annamalai, E. (2024). "The Sanskrit Paradigm of Tamil Grammar: Embrace and Resistance". *Bhasha*, 3(1), [1-16] 107-122.

DOI 10.30687/bhasha/2785-5953/2024/01/002

There are political and social factors underpinning the relationship between Tamil and Sanskrit and their communities. The debates over the grammatical description of Tamil are defined in a certain way. This paper does not go into these underlying factors but limits itself to cultural, ideological and intellectual factors. The cultural factor comes under the rubric of tradition – tradition that is emphasized by historical continuity and distinctive identity of a linguistic practice, which in the present case is the writing of grammars. This emphasis on tradition remains throughout the entire history even when it could be demonstrated that the distinctive identity of Tamil tradition changed to become closer to the Sanskrit's. The two traditions that are codified and given generic names – the Sanskrit grammatical tradition and the Tamil grammatical tradition. Tradition is called *marapu* in Tamil, whose meaning is expansive and includes convention and lineage, but the tradition obviously is not homogeneous or monolithic historically and even at the given period.

The authoritative head of the Tamil grammatical tradition is Tolkāppiyar, the author of the first known grammatical treatise in Tamil, written in the milieu of earliest literary texts called Sangam poetry of the period before and after Common Era. Most probably he was a Jain (S. Vaiyapuri Pillai quoted in Zvelebil 1973). His grammar is the *Ur*-text of the Tamil tradition and is based on the interpretation of texts, primarily literary texts. The authoritative head of the Sanskrit grammatical tradition is Pāṇini, a Brahmin who generated the *Ur* text of that tradition; its purview is the codification of the language of the later Vedic texts. Tamil grammarians from the earliest times take Tolkāppiyar as the fountainhead of their tradition and as the primary authority of Tamil grammar, if the mythological figure Agastya is kept out of count. The Tamil grammarians take Pāṇini to be the fountainhead and the authority of the Sanskrit grammatical tradition, if the passing reference to the grammar of Indra is ignored. Tolkāppiyar represents the Tamil grammatical tradition of about two millennia, and Pāṇini represents the Sanskrit grammatical tradition during the same period in the Tamil country. If one visualizes a cultural war between these two grammatical traditions, it is a war of ideas attributed to the two above authorities. But historically and culturally the defence of ideas is not visualized as a war until the modern period with its nationalistic interpretation of Tamil history, including the history of grammatical science. Even during this modern period, war is understood as the question of who took ideas from whom and what ideas were imposed on Tamil grammars, particularly those found in the commentaries on *Tolkāppiyam*. The 'imposed' ones are viewed as contrary to the intellectual tradition of Tamil and are

rejected so that the alleged 'purity' of the tradition is maintained.[1]

The intellectual tradition is understood and defined in terms of texts and the relationship between them. Moreover, they are defined by shared and improved grammatical analyses, or by the absence or presence of their deviations. The chronologically earlier and authoritative text is called *mutal nūl* 'primary or original (scientific) text', while the secondary and later texts are called *vaḻi nūl* 'descendent text' and *cārpu nūl* 'dependent text'. The former of the two stays close to the original except for making changes to accommodate historical changes in the language and some new ideas that are considered supplementary. Expanded and condensed versions of the original also belong to the same category of *vaḻi nūl*. The other type is called dependent as it shares many analyses with the original but substitutes new analyses in some cases. It is not an independent text despite this partial deviation from the original.

Cankaranamaccivāyar, an eighteenth-century commentator on the thirteenth-century *Nannūl* (Dhamotharan 1999), gives the analogy of a son to *vaḻi nūl*, and of a son-in-law, who does not inherit from the father-in-law, to *cārpu nūl*. Both these categories of texts stay within the tradition, which this analogy views as an extended family. The cultural emphasis is on staying within and perpetuating the tradition or genealogy, and so there is no significant discussion of texts that break the tradition.

Mayilainātar, the earlier commentator on *Nannūl*, mentions another type of texts referred to as *etir nūl* 'counter text'. This would qualify texts that break away from the tradition and start a new one. No Tamil grammatical text has ever been designated as such to the best of my knowledge. One could argue that the texts I will discuss below, such as Vīracōḻiyam and Ilakkaṇakkottu, may be classified as such counter texts, but they are not. This shows that the tradition within Tamil grammar is inclusive.

Sivañāna Munivar (of Mutal Cūttira virutti, which pegs his ideas in his commentary of the first sutra of *Tolkāppiyam*), a grammarian, litterateur and a Saivite philosopher of the eighteenth century, gives arguments for deciding the relationship between the two texts, *Tolkāppiyam* and *Nannūl*, with more than a millennium between them. These two texts are universally considered to be solid parts of the Tamil tradition and to belong to the same lineage in spite of the fact that they have different concepts of the grammar. For *Tolkāppiyam*, poetics is part of the grammar, indeed the largest section in this text,

1 Such views can be seen in many interpretations of the sutras of *Tolkāppiyam* by Tamil scholars in the modern period such as Ialkkuvanar, who translated this work into English with critical notes. Similar views can be seen in some publications in Tamil relating to *Tolkāppiyam* by scholars of similar persuasion.

but it is altogether dropped from the grammar by *Nannūl*. This is not just a matter of condensation for *Nannūl*, but is an ideological shift. It is believed to have made *Tolkāppiyam* simpler for students and to have incorporated the linguistic changes in the Tamil of his time.

Let me give two illustrative examples from Munivar about the textual relationship between these two texts, one from phonology and another from morphology. These are examples considered to be stepping out of the legacy (Annamalai 2018).

Tolkāppiyam divides letters/phonemes into two categories, viz., primary (*mutal eḻuttu*) and secondary (*cārpu eḻuttu*). The latter category is determined by their occurrence in specific syllabic structures and sequences having *sandhi* effects. Their phonetic feature is length reduction of the phonemes, and this synchronic phonetic alternation is contextual. The first category (primary letters) has thirty phonetic manifestations of letters and the second category has three. *Nannūl*, on the other hand, puts ten phonetic manifestations of letters in the second category. It uses a different criterion of phonetic alternation for counting, that is, morae measures (*māttirai*) of letters/phonemes in the basic unit of prosody, called *acai* 'syllable' (it does not treat prosody to be a part of its grammar). This is a significant difference that extends the phenomenon of ordinary language to versified language for the analysis of phonetic reduction and elongation.

Another is an example from morphology. The first person (and the second person) pronouns are treated as human (*uyar tiṇai*) in *Tolkāppiyam*, but as common to human and non-human (*viravu-t tiṇai*) in *Nannūl*. The former's criterion is the verbal ending (it is pronounced in predicates that are participial nouns: *naan paṭittavan* / *paṭittavaḷ* 'I am an educated person'; the gender male or female is marked in the predicate for human subjects, which is *nān* (I) in this case. *Nannūl*'s criterion is referential, where the referent of the first (and the second) person could be any animate being, inclusive of humans and animals. This difference in the criterion to determine gender of the first (and second) person is theoretical and therefore significant.

Though differences such as the above are significant, they do not show that these two texts belong to different legacies. Munivar brings in the concept of error (*vaḻu*) to argue that the differing analysis in the later text is an error and so does not count. His motivation is to protect the tradition from disintegration and to keep the reproduction of knowledge within the tradition's framework (Annamalai 2018). His view tells us that breaking a tradition by unrestrained novel analyses is perceived to be a cultural issue, in fact a problem, in knowledge production.

The meaning of content dependence, intellectual descent and counter-analyses with reference to texts mentioned above is not understood by mere empirical and quantitative facts but rather by the conceptual framing of the grammatical problem and its solution.

Admission of a counter text, given this criterion, is rather an exception than a norm in the Tamil grammatical tradition.

The analysis of the grammar of Sanskrit by Pāṇini (and others) was an attractive alternative for Tamil grammarians, allowing them to rethink Tamil grammar. There were multiple reasons for this attraction, tied to the religious, political, and intellectual power the Sanskrit scholars and linguists constructed around that language. The response of the Tamil grammarians to the analytical power of Sanskrit was either to embrace this mode of analysis or to resist it. Those who embraced it were called the followers of Sanskrit textualism (*vaṭanūlār matam* 'school of thought based on Sanskrit [Shastric] texts' [S-School]) and those who resisted were called the followers of Tamil textualism (*tamiḻ nūlār matam* 'school of thought based on Tamil texts' [T-School]). *Yāpparunkala virutti*, a detailed commentary on Prosody of the eleventh century, but the idea of two schools of thought is widespread. The interesting point is that neither school of thought questions *Tolkāppiyam* as the *Ur* text of Tamil grammar of all times.

The grammatical theory and analysis taken from Sanskrit grammarians was also argued, from this perspective, to be in line with the Tamil grammatical tradition enunciated in *Tolkāppiyam*. That is, there was no need to reject *Tolkāppiyam* and justify a counter text. What was required was a reinterpretation of the text of *Tolkāppiyam* for the new analytical model. The commentators on *Tolkāppiyam* differ about the interpretations of this text, but all insist that theirs do not deviate from what the text intended to mean by its author. Their text-based arguments (the way the sutra is worded etc.) to validate their interpretation as true are interesting in themselves, but they are a subject for a different paper.

Resistance to accepting a Sanskrit analysis to describe Tamil is in most cases based on the view that such an analysis is not the view of *Tolkāppiyam* as stated in its sutras. The opposite is true for those later grammarians who embraced analysis from Sanskrit grammar, as mentioned above. When they argue that their new analysis was the actual intent of the sutra of *Tolkappiyam*, they take the position that the medieval commentators have misinterpreted the sutra.

The embrace of Sanskrit models may be selective or total. Selection is guided by the principle of maintaining the nature of the Tamil language, which boils down to maintaining the Tamil tradition in analysing Tamil grammar. This is called, subjectively, by some modern scholars as paying attention to the Tamil 'genius' (Thirugnanasambandam 1994). The selective embrace is a feature of earlier times in history, i.e. of earlier commentators and grammarians. The preference in historically later times is to be liberal with selection, coming closer to total adoption. We will see latter in the paper how the idea of Sanskrit grammar and of Tamil grammar changed during the latter times.

Let me first illustrate the selective adaptation of Sanskrit analysis of compound nouns of Tamil in *Tolkāppiyam*. Analysis of compound nouns is a preferred topic in the traditional grammars of all major Indian languages. Tamil is no exception.

D'Avella (2012) makes an incisive comparative study of the analysis of compound nouns in *Aṣṭādhyāyī* and *Tolkāppiyam*. Two grammatical concepts are operative in compound formation viz., semantic integration of the meanings of the constituent words (*ekārthabhāva* in Sanskrit) and formal condensation (by dropping the morphemes that identify the grammatical relation between the constituent words – (*lopa* in Sanskrit). Tamil has a homophonous root with both these meanings, *toku* (*tokukka*, 'collect together') and *toku* (*tokka*, 'elide'). The wording of the sutras on compounds in *Tolkāppiyam* suggests that it gives primacy to semantic integration; interpreting sutra 2.1.1 of *Aṣṭādhyāyī* of Pāṇini, Patañjali, as cited by D'Avella (2012, 10), says that *samartha* in the sutra "indicates that when we apply an operation to words a single meaning should emerge from the constituting parts". Pāṇini and Tolkāppiyar, living in different periods, concur. *Tolkāppiyam* (*Collatikāram* 414) states: *ellāt tokaiyum oru connaṭaiya* (all compounds exhibit the behaviour of a single word). He further says that the integrated meaning of the compounds derives from the claim that they have the conjunct meaning of their analytical equivalents. The analytical equivalents of the compounds may or may not have inflection in the words that compose them. Hence, elision will be superfluous to define some types of compounds. If there is an inflection such as a case suffix in the analytical equivalents, this suffix is elided in the compounds to give them the nature of a unitary word defined by the grammatical behaviour of compounds, as in, for example, the plural formation. Furthermore, elision is not universally true for all types of compounds. Nevertheless, elision is considered to be the defining feature of compounds by some grammarians (including the commentators) both in Sanskrit and Tamil (D'Avella 2012, 7). The debate between these two theoretical stances about compounds took place across language boundaries and across century divisions. The grammatical tradition of Tamil contributed to the refinement of the theory. D'Avella concludes his paper that is focused on Sanskrit and Tamil thus:

> The processes of borrowing and adopting were often nuanced both at the lexical as well as the conceptual levels […]. The mix of these strategies result from the complex linguistic reality Tamil occupied and the efforts Tamil grammarians made to maintain a distinct identity for their grammatical tradition. (7)

The epistemological interaction between Tamil and Sanskrit was at meta-grammatical and grammatical levels, according to him.

Let me illustrate another case of similar interaction from the analysis of compounds itself. Though compounds received much attention by Pāṇini and Tolkāppiyar, their interests are different. The former's lies in the construction of larger lexical units built on phrasal structures; the ambiguous semantic interpretations of such units is of secondary interest to him. The latter's interest is in accounting for the semantics of the compounds, as they are used abundantly in literary texts such as the ancient Sangam texts. In the words of D'Avella (2012, 12):

> the linguistic element which has been lost in compounding need not always be specified exactly because the *Tolkāppiyam* is not interested in *deriving* compounds, as is the *Aṣṭādhyāyī* but rather offering a means to analyze and talk about the words that we find in literary language.

Though there was shared interest in grammatical phenomena common to Sanskrit and Tamil,

> not all the compounds in Sanskrit have an equivalent in Tamil, namely the *avyayībhāva* 'adverbial compound' and the *dvigu* 'numerical compound'. Similarly, the *viṉaiyiṉ tokai* and *uvamat-tokai* have no equivalent among the major types of Sanskrit compounds, although Pāṇini and subsequent commentators do recognize simile compounds as a subtype of *karmadhāraya*s. The *viṉaiyiṉ tokai*, which consists of a bare verbal root followed by a noun, is completely foreign to the Sanskrit language. (D'avella 2012, 10)

D'Avella further reasons that

> the addition of the verbal compound to Tamil grammar not only demonstrates a willingness to deviate from the general Sanskrit categories but also reflects the author's (Tolkāppiyar's) attention to syntactic structures specific to Tamil poetry. The use of a bare verbal stem instead of a relative participle is ubiquitous in Caṅkam poetry.

It may not even be seen as an addition to Tamil grammar from the point of view of the Sanskrit grammar, but could be seen as independently conceived, empirically grounded conclusion in the Tamil tradition. This is in spite of the fact that the Sangam texts have evidence that the bare verb of this compound may have other syntactic elements – Subject, Object etc. – just like the verbal predicate of a sentence (Lehmann 1994; Wilden 2018). It suggests that, in the Tamil conceptualization of the compound, fusion into one word is paramount (as shown above) even when it allows one constituent of the

compound to remain in relation with the other words of the sentence which it is part of.

The above citations from D'Avella show that there were grammatical ideas flowing probably bi-directionally and that Tamil grammar and the *Ur* text of its tradition took cognizance of the facts of the Tamil language and the purpose of its grammar when it began the tradition. To cite D'Avella's (2012) over-all view about the Tamil grammars of the early period (i.e. before the medieval period of the second millennium),

> it would be [...] extreme to see a slavish adherence to the preexisting Sanskritic models. Instead, I envision the author(s) of the *Tolkāppiyam* as turning to Sanskrit texts (or ideas derived therefrom) to create a unique system of linguistic analysis well suited for Tamil, albeit a few mismatches remain.

While such an interactive approach continued to be followed in the description of Tamil grammar by the commentators on *Tolkāppiyam*, who individually varied in adopting and adapting the ideas of Panini's grammar of Sanskrit and its elaborations and refutations, there was a fundamental retooling of Tamil grammar in the pre-modern period (around the eighteenth century) by grammarians patronized by Saiva mutts in the Kaveri delta, which were headed by non-Brahmin pontiffs and were engaged in codifying and promoting Saiva texts in Tamil as well as the Tamil language; their work included translations into Tamil from Sanskrit. This effort however started much earlier, at the dawn of the second millennium in the same region under the Chola dynasty by Vīracōḻiyam (VC), an eleventh-century Buddhist grammar. It was part of the role of Buddhist thinkers of its time to generate knowledge in Sanskrit and spread it in the languages of the regions where they were preaching. As D'Avella (2021) has demonstrated non-Paninian grammatical texts played a greater role in creating new grammatical models; the text *Prayogamukhi*, which was important to the wandering Buddhist textual community, was the primary source for VC.

VC makes many references to Sanskrit texts (without naming them but calling them generically as *vaṭanūl* 'northern (Shastric) texts'. Its goal might be to apply the Sanskrit grammatical model to Tamil. It creates a new model and a new metalanguage (using the Sanskrit technical terms without calquing) for the description of Tamil. It does not criticize or reject *Tolkāppiyam* openly when deviating from it. It keeps the conceptual framework of the five-fold division of grammar of *Tolkāppiyam* that includes poetics, but the VC's theories are derived from works in Sanskrit. As D'Avella notes, this development was motivated by the perceived superiority of Sanskrit theories or knowledge as well by the changing empirical realities of the Tamil language

and literature, some of which exhibit the influence of Sanskrit.

A good example of modelling Tamil grammar (Meenakshisundaran 1974) after the grammar of Sanskrit is the treatment of cases. *Tolkāppiyam* analyses case as the morphology of nouns that relates them syntactically with the predicate verb of the sentence. It is a relation of a set of verbs that are grouped semantically as they govern the nouns of particular declensions, i.e., nouns with particular case markers. The first case (*mutalām vēṟṟumai*) or the nominative or Subject case (*eḻuvāy vēṟṟumai*), which does not have a case marker, is different from other cases; it is not governed by any set of lexical verbs but by sentence types indicated by specific predicates such as verbal indicative, nominal indicative, interrogative and imperative. The case theory of *Tolkāppiyam* is based on the simple idea that the meaning of predicates determines certain types of nouns (like the fact that the transitivity of the verb determines the occurrence of the Object noun).

The types of nouns so determined are expressive of a certain associations of the verbal action. They are: the noun as the object, as the recipient of the object, as the location of the action, as the source from which the action emanates, as being similar to another object. As the case nouns have different morphologies and their corresponding associate statuses are different, this phenomenon of nouns is given the name *vēṟṟumai* (difference). This idea is similar to Panini's idea of *vibhakti* (difference).

Tolkāppiyam calls what I have termed inadequately 'associate' by the term *mutal*, as in *vinai mutal* (the antecedent of predicate), which could be understood as the 'antecedent' of predicates. They are antecedents in the sense that the associates object, recipient etc. exist independently, and the predicate assigns them to nouns to construct the structure of a sentence. *Tolkāppiyam* calls this assignment 'case meaning' (*vēṟṟumaipporuḷ*) and the marker that indicates the assignment 'the case morph or suffix' (*vēṟṟumai urupu*). Panini adds another layer to *vibhakti*s or case-morphs and calls it *kāraka*, which are extra-syntactic in their generation and are the logical antecedents of an action. The grammar maps the nominal syntactic units of a sentence with the logical antecedents of the action of the predicate.

Tolkāppiyar's description of cases is a tool for interpreting a literary text such as the ancient Sangam texts, which have a preponderance of elided cases in a sentence. The way to recover the right case is from the meaning of the predicate, which governs the case. For the same purpose, *Tolkāppiyam* is also concerned with the appearance of one case in place of the anticipated and legitimate another case dependent on a specified meaning of the predicate but without altering the case meaning. This is called *vēṟṟumai mayakkam* (alternation of cases). That is, one case marker alternates with another for a predication in the same meaning. The predicate *kuttu* (poke) may

govern an object case or the location case: poked the eye/poked in the eye; the predicate *veṭṭu* (cut down) may do the same: cut down the branch/cut down (the tree) at the branch. These are cases of predicates overlapping in their governance of cases.

Karaka theory is concerned with the absence of one-to-one correspondence between *karaka* and *vibhakti* endings. A 'mismatch' may, along with other ways, have a resolution in the morphology of the verbal predicate. In a sentence in Tamil or Sanskrit that is equivalent to the English sentence 'the book sold well', the book's logical relationship with 'sell' is that of an Object, not of a Subject, as its case indicates. *Karaka* in Sanskrit has a dual effect on the noun and on the verb. *Tolkāppiyam* does not deal with the verbal effect in the chapter on cases but in the chapter on verbs. It says that it is the property of some verbs like 'sell' (not its morphology) to allow the use of Object as Subject by convention, not by the grammar (*Tolkāppiyam Collatikāram* 246. செயப்படுபொருளைச் செய்தது போலத் தொழிற்படக் கிளத்தலும் வழக்கியல் மரபே *ceyappaṭuporuḷaic ceytatu pōlat toḻiṟpaṭak kiḷattalum vaḻakkiyal marapē* 'There is a convention in language use to express the Object (what is done) as the doer (Subject) that gives action'. The commentator Cenāvaraiyar takes this to be an error of convention (மரபு வழு, i.e. a usage that is not sanctioned by a general rule of grammar and so it needs a special rule), and this sutra legitimizes the usage. The corresponding sentences in English are 'the book sold well, this rice cooked well'.

In essence, *Tolkāppiyam* doesn't have a use for a theory of *karaka* to be added to the description of case for his purpose of the grammar, which is to aid interpretation of literary texts. *Karaka*, on the other hand, is a theory to explain people's common 'mismatches' between the case of the noun and the semantic role of it in the proposition, which is a commonplace in language use, as the above sutra of *Tolkāppiyam* says.[2]

The absence of the description of *karaka* in *Tolkāppiyam* is a challenge to the desire of VC to approximate the grammar of Tamil to that of Sanskrit, though the purposes of these grammars are different. VC, in the very first sutra of the chapter on *vēṟṟumai* (case), takes the chapter to be inclusive of *vibhakti* and *karaka* and goes on to say that there are eight *vēṟṟumai* following *Tolkāppiyam* and six *karaka*s following Panini. He takes the latter's view that a *karaka* has morphological effects on the noun as well as on the verb and concludes that the person-number-gender suffixes of the verbs are reflected in the nouns that are in the nominative case and that they are the markers of the nominative case. This is totally different from the analysis in

2 Matilal 1991 attributes this observation of the commonness of the mismatch to some scholars.

Tolkāppiyam that the nominative case is not morphologically marked and it is simply the noun itself. This difference between the nominative (i.e., Subject) case in Tamil and Sanskrit is obliterated by the misapplication of analysis of *karaka*'s effect on the predicate verb. VC and its commentator go on to sub-divide the six *karaka*s, which reflect the post-Paninian discussion of *karaka*s by Sanskrit grammarians, and to explain the lack of one-to-one match between the *karakas* and the case marked nouns. To have the nominative case marked morphologically is contrary to the Tamil system, which has it unmarked as it is governed by all predicates of all meanings, unlike the other cases.

Cēnāvaraiyar, a thirteenth-century commentator of *Tolkāppiyam* steeped in the Sanskrit grammatical tradition, is subtler than VC in merging the *kāraka* theory with the case theory of *Tolkāppiyam* (Tol Col 112). He does not alter the sutras of cases in *Tolkāppiyam* but takes a sutra at the end of the chapter on *vēṟṟumai mayakkam*, which is a list of antecedents, *mutal*, mentioned above. It is called *toḻil mutalnilai* (the standing of the antecedents of the verb), which he takes to mean *karaka*. His logic for getting this meaning is this: antecedent (*mutal*) gives case meaning (*vēṟṟumaip poruḷ*), which is the same as the reason/rationale for the case (*vēṟṟumaik kāraṇam*), and *kāraṇam* is synonymous with *karaka*. It must be noted that Cenāvaraiyar does not assign any marker for the nominative case and does not deviate from the Tamil tradition in this respect.

The eighteenth-century grammars Pirayōkavivēkam and Ilakkaṇakkottu take the route of VC, though they take their cue from Cēnāvaraiyar (to extend *vēṟṟumaip poruḷ* of *Tolkāppiyam* to the *karaka* meaning of Panini). But they go farther than VC in giving multiple case markers for the nominative case. Ilakkaṇakkottu, using his self-claimed prerogative of bringing up hard-to-solve rules of grammars (*ariya viti*) for a solution, adds a sutra to the Tamil-oriented grammatical texts: "while there are three morphological markers of the nominative case in Sanskrit, Tamil could desire to have more, or less, of this number" (வடமொழி எழுவாய்க்கு உருபு மூன்று; அவைதாம் விரியவும் தொகவும் விரும்பும் என்ப, *vaṭamoḻi eḻuvāykku urupu mūṉṟu; avaitām viriyavum tokavum virumpum eṉpa*, 140). He thus goes beyond reinterpreting the Tamil grammatical texts to alter them in order to approximate the Tamil grammar with the Sanskrit grammar. He justifies his move in the following way: தமிழிற்கு இன்றாகிய வடமொழி இலக்கணம் தமிழில் வருதலானும் எழுவாய் உருபும் வடமொழியிற் கண்டு இன்று கொண்டுவந்தது என்றலும் ஒன்று *tamiḻiṟku iṉṟākiya vaṭamoḻi ilakkaṇam tamiḻil varutalāṉum eḻuvāy urupum vaṭamoḻiyiṉ kaṇṭu iṉṟu koṇṭuvantatu eṉṟalum oṉṟu* (Ilakkaṇakkottu, 140). "As Sanskrit grammatical features not found in the Tamil grammar do occur in Tamil language, it can be said that the case marker of the nominative found in Sanskrit is brought to Tamil in the current times". His argument is to take the features of Sanskritized Tamil as evidence for including

them in the Tamil grammar.

A new grammar rule is added to the grammar of Tamil when a grammatical feature is not mentioned in the Tamil grammar but the Sanskrit grammar has it; this is to claim that the Tamil language has that grammatical feature also. The empirical evidence provided, however, can be interpreted not as a case marker at all. Nevertheless, the axiomatic statement of Ilakkaṇakkottu (92) "keep in mind that the two languages Sanskrit and Tamil have the same grammar" increase the motivation to ignore the nuances of empiricism (வடமொழி தமிழ்மொழி எனும் இரு மொழியினும் இலக்கணம் ஒன்றே என்றே எண்ணுக *vaṭamoḻi tamiḻmoḻi eṉum iru moḻiyiṉum ilakkaṇam oṉṟē eṉṟē eṇṇuka*).

The centuries-long interaction between grammarians in the history of the North and the South described here did not take place merely within the grammar of a single language but also between the grammarians of two different languages, Sanskrit and Tamil. It was between the grammarian of Tamil as the first language and the grammarian of Sanskrit as a second language (not the grammarian of Sanskrit as the first language). The grammarians of Tamil, almost all of them bilingual in Sanskrit, were exposed to Sanskrit grammatical descriptions to different degrees, and some were well-trained in the Sanskrit tradition. They, at the very least, participated directly or indirectly in pan-Indian epistemological practices and theory building, and they absorbed pan-Indian trends while contributing to them as well.

Among the Tamil grammarians, there are two broad groups, as mentioned above; one group added new ideas to Tamil grammatical concepts, primarily, but not exclusively, from the Sanskrit grammatical system; the other group intended to bring Tamil grammatical concepts closer, if not identical, to those in Sanskrit. Within the second group, the dominant view was that the Sanskrit language was the *Ur* language for all the languages of the sub-continent and so its grammar is applicable to all other languages. A variant formulation of this view is to attribute primacy to Sanskrit texts as the embodiment of universal knowledge or theory. The commentator of VC (Sutra 60) Peruntēvan (twelth century, a student of the author of the work), trying to explain the absence of providing etymologies of Tamil words in the text he is writing the commentary on, states this: தமிழ்ச்சொற்கெல்லாம் வடநூலே தாயாகி நிகழ்கின்றமையின் அங்குள்ள வழக்கெல்லாம் இங்கும் பெறும் *tamiḻccoṟkellām vaṭanūlē tāyāki nikaḻkiṉṟamaiyiṉ aṅkuḷḷa vaḻakkellām iṅkum peṟum* (As the treatises in Sanskrit are the mother /source (to explain) all the words in Tamil, all the explanatory practices there will apply here also). Note that the phrase is வடநூலே தாய் (Sanskrit treatises are the mother source), and not வடமொழியே தாய் (the Sanskrit language is the mother of Tamil words). Peruntēvan, it could be argued, is not thinking here of a genetic or historical relation between languages but is claiming an

epistemological relation of inquiry. Any analysis or theory propounded in Sanskrit Shastric texts has universal application. Hence there is no need to develop a treatise on Tamil etymology. This reveals the appeal of the knowledge produced in Sanskrit. One can see a similarity to the position with reference to English in the modern period. If this logic about the knowledge in Sanskrit is taken to its logical end, there is no need for a separate grammar based on different principles, and thus there can't be a Tamil grammatical tradition. To state it more precisely, there is an ethnic Tamil grammatical tradition but it is absorbed into a universal grammatical system discovered and expressed in Sanskrit.

The intellectual conflict between these two groups is not expressed through open clashes or condemnation in the Tamil context. *Tolkāppiyam* did not lose it status as the *Ur* text of the Tamil transition and there was no attempt to discredit it or to dislodge its status. This is in spite of the fact that S-School changed the Tamil grammar as initiated by *Tolkāppiyam* with new concepts taken from Sanskrit grammarians such as Pāṇini, but not only from him. This is an epistemological war based on the ideology of language order (Ollett 2017), which is ultimately about submersion of the identity and multiplicity of intellectual traditions relating to the study of grammar. But it was fought under the camouflage of reinterpretation of Tamil by searching for the universal truth of language through a single grammar for Tamil and Sanskrit.

Acknowledgement

This is a revised version of the paper presented in June 2023 at the Seminar on Grammar Wars sponsored by Israel Institute of Advanced Studies and conducted at Hebrew University of Jerusalem. I owe any clarity of the paper to the comments of David Shulman, Whitney Cox. Andrew Ollett, Victor D'Avella on the earlier drafts of the paper and the questions asked by the participants of the seminar. I express my gratitude to the anonymous first reviewer, whose extensive comments helped to make the paper accessible to readers not familiar with the Tamil grammatical tradition.

The research underlying this paper was supported by the NEEM Project of the above University funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 786083 – NEEM).

Bibliography

Primary Sources

Cēṉāvaraiyar (1938). *Tolkāppiyam with Cēṉāvaraiyam*. Ed. by Piramaśrī Ci. Kaṇēcaiyar. Cuṉṉākam: Tirumakaḷ Aḻuttakam

Ilakkuvanār, Singaravel (1963). *Tholkāppiyam (in English) with Critical Studies*. Madurai: Kuṟaḷ Neṟi Publishing House.

Munivar, Civagnāna (1968). *Tolkāppiyac cūttira virutti* தொல்காப்பியச் சூத்திர விருத்தி விளக்கக் குறிப்புகளுடன். Thiruvavadudurai: Thiruvavadudurai Atiinam.

Muṉivar, Pavaṇanti (1946). *Naṉṉūl*. With the commentary of Mayilai Nātar. Ed. by S. Kaliyāṇa Cuntar. Ceṉṉai: Kapīr Accukkūtam.

Muṉivar, Pavaṇanti (1999). *Naṉṉūl* of Pavaṇanti Muṉivar பவணந்தி முனிவர் இயற்றிய நன்னூல் மூலமும் சங்கரநமச்சிவாயர் செய்து சிவஞான முனிவரால் திருத்தப்பட்ட புத்தம் புத்துரை என்னும் விருத்தியுரையும் (*Nannūl* Text written by Pavaṇanti Munivar. Expansive Commentary written by Cankaranamaccivāyar and improved with a very new Commentary by Civagnana Munivar). Ed. by A. Dhamodharan. Chennai: International Institute of Indian Studies.

Puttamittiraṉ (1942). *Vīracōḻiyam*. With the commentary by Peruntēvaṉār. Ed. by Kā. Ra. Kōvintarāja Mutaliyār. Ceṉṉai: Pavāṇantar Kaḻakam.

Tecikar, Cāmināta (1973). *Ilakkaṇak Kottu*. Ed. by Ti. Vē. Kōpālaiyar. Tañcai: Caracuvati Malkāl Nūlnilayam. Tañcai Caracuvati Makāl Veḷiyīṭu 146.

Tīkṣitar, Cuppiramaṇiyaw (1973). *Pirayōka Vivēkam*. Ed. by Ti. Vē. Kōpālaiyar. Tañcai: Caracuvati Makāl Nūlnilayam. Tañcai Caracuvati Makāl Veḷiyīṭu 147.

Secondary Sources

Annamalai, E. (2018). "Restoring Knowledge by Integration and Refutation: Sivañana Munivar's Conceptualization of Tradition" [Unpublished]. *The Tenth NETamil Workshop = Tracing School Formations and Scholarly Networks*. Puducherry: EFEO, Pondicherry Centre.

D'Avella, V. (2012). "Translating and Transforming Grammar: Compounding in the *Tolkāppiyam*" [Unpublished]. *40th Annual Conference on South Asia* (Masidon, 20-23 October 2011). Madison.

D'Avella, V. (2020). "The Theorization of Tamil Compounds". *Beitrage zur Geschichte de Sparachwissenschafft*, 30.

D'Avella, V. (2021). "The *Vīracōḻiyam*: A Tamil Grammar Through the Eyes of Sanskrit". Ciotti, G.; McCann, E. (eds), *Linguistic and Textual Aspects of Multilingualism in South India and Sri Lanka*. Pondicherry: École Française d'Extrême-Orient ; Institut Français de Pondichéry, 327-410.

Lehmann, T. (1994). *Grammatik des Alttamil*. Stuttgart: Steiner Verlag.

Matilal, Bimal K. (1991). "Bhavanada on 'What is Karaka?'". Deshpand, M.M.; Bhate, S. (eds), *Paninian Studies: Professor S.D. Joshi Felicitation Volume*. Ann Arbor: University of Michigan Press, 263-82.

Meenakshisundaran, T.P. (1974). *Foreign Models in Tamil Grammar*. Trivandrum: Dravidian Linguistics Association.

Ollett, A. (2017). *Language of the Snakes. Prakrit, Sanskrit, and the Language Order of Premodern India*. Berkeley: The University of California Press.

Thirugnanasambandham, P. (1992). *Sanskrit Tamil Contact*. Thiruvananthapuram: International School of Dravidian Linguistics.

Wilden, E. (2018). *Grammar of Old Tamil for Students*. Pondicherry: École française d'Extrême-Orient (EFEO); Institut français de Pondichéry (IFP). Collection Indologie 137, NETamil Series 3.

Zvelebil, K. (1973). *The Smile of Murugan: On Tamil Literature of South India*. Leiden: Brill.

e-ISSN 2785-5953

Bhasha
Vol. 3 – Num. 1 – April 2024

‘To Be Smeared’ or ‘To Be Attached’? An Investigation of Sanskrit *lipyate kāmaiḥ* and Pāli *lippati kāmesu* in Light of Their Chinese Translations

Francesco Barchi
Ludwig-Maximilians-Universität München (LMU), Germany

Abstract In Buddhist translated literature, the Chinese expression *bù rǎn yú yù* 不染於欲 (not attached to desires) is apparently used as variant of the passive construction *bù wéi yù suǒ rǎn* 不為欲所染 (not tainted by desires) to translate the same Sanskrit source expression *na lipyate kāmaiḥ* (not being smeared by desires). An Indic parallel closer to *bù rǎn yú yù* 不染於欲, namely *na lippati kāmesu* (not being attached to desires), is found in Pāli and in some Hybrid sources. This paper argues that the Sanskrit and Pāli forms can be traced back to a common archetype akin to the Pāli form and that the *-ya*-present *lipyate* was originally used as a class IV intransitive present. Owing to use of the historical instrumental suffix *-ehi* as a generalised oblique plural ending in Middle Indo-Aryan, the form *lipyate kāmehi* (< **lipyate kāmeṣu*) was eventually reanalysed as a present passive. The two variants found in Chinese translations bear witness to the semantic and grammatical ambiguity underlying the Indic source expression.

Keywords Chinese Buddhist translations. Pāli. Buddhist Sanskrit. Passive constructions. Oblique plural.

Summary 1 Introduction. – 2 Passive Constructions in Indo-Aryan and Chinese. – 2.1 Passive Constructions in Indo-Aryan. – 2.2 Passive Constructions in Chinese. – 3 *Bù rǎn yú yù* 不染於欲: A Passive Construction?. – 4 A Broader Focus. – 5 The Elusive Meaning of *lipyate*/*lippati*. – 5.1 The Indo-Aryan Root *lip-*: Meaning and Case-Marking. – 5.2 Diachronic Development: From Locative to Instrumental. – 6 How Was the Expression Understood by Translators?. – 7 Conclusion.

Edizioni Ca'Foscari

Peer review

Submitted 2023-09-24
Accepted 2024-02-24
Published

Citation Barchi, F. (2024). “‘To Be Smeared’ or ‘To Be Attached’?”. *Bhasha*, 3(1), [1-46] 123-168.

DOI 10.30687/bhasha/2785-5953/2024/01/003

1 Introduction

For the scholar who wishes to research Buddhist Chinese, the language of Buddhist translated literature presents a series of challenges pertaining to every domain of the linguistic system. As regards the grammatical dimension, a common problem one has to face is that some grammatical features of the Indic source text might emerge in the Chinese translation. Using a metaphorical expression borrowed from translation studies (Teich 2003, 22), the linguistic shape of the Indic source text tends to 'shine through' the translated *sūtra* making the language of translated texts appear different from coeval literature composed in standard literary Chinese.[1] The distortive influence of the Indic source text is particularly evident in early translations, often characterized by a hyper-overt rendition of the original text into an obscure variety of Chinese almost incomprehensible without resorting to the Indic parallels.[2]

A philological approach to the linguistic investigation of translated *sūtras* generally allows one to avoid the possible pitfalls in the grammatical analysis of this typology of texts: in most cases, the comparison of the Chinese translations with the extant Indic parallels, when available, provides the key to the exact grammatical interpretation of the Chinese texts. Notwithstanding, in certain cases the mere comparison with the parallels does not suffice for the correct analysis. The study of the Chinese Buddhist translations often requires a deeper philological and linguistic analysis of a specific expression or passage, unfolding the various diachronic and textual layers underlying the use of a certain grammatical feature. In this paper, I intend to use the Chinese expression *bù rǎn yú yù* 不染於欲 and its Indic, Chinese and Tibetan parallels as a case study to illustrate the linguistic and philological factors underlying the grammatical analysis of the Chinese Buddhist translations. I argue that the alternation between *bù rǎn yú yù* 不染於欲 and the passive construction *bù wéi yù suǒ rǎn* 不為欲所染 found in synchrony in Chinese translations

This article is an adaptation of a paper presented at LMU University of Munich on 22 October 2021 as a part of the workshop *Buddhism and Language: A Twofold Perspective: The Role of Language in Buddhist Teachings and the Role of Buddhist Sources in Linguistic Research*. I thank the participants of the workshop for their helpful feedback, in particular Stefan Baums and Niels Schoubben. I am indebted to Zhang Yiren and Benedikt Peschl for valuable remarks on a draft version of the article. I also want to thank Kelsey Martini for 'polishing' my English.

1 For an introduction to language contact through-translation, see Kranich 2009; 2014; Kranich, Becher, Höder 2011; Baumgarten, Özçetin 2008; Becher, House, Kranich 2009. On Chinese Buddhist translations as a locus of grammatical interference, see Barchi (forthcoming).

2 See Zacchetti 2007 for a detailed discussion of Ān Shìgāo's 安世高 (fl. ca. 148-80, being the first translator mentioned in Chinese historical sources) translation technique.

to translate the same source form is motivated by the semantic and morpho-syntactic ambiguity of the Indic expression, reflected in the opposition between the two patterns *na lippati kāmesu* and *na lipyate kāmaiḥ/kamehi* found in the extant Indic parallels.

The paper is organized as follows. In § 2, I briefly introduce passive constructions in Indo-Aryan (2.1) and Chinese (2.2). In § 3, I introduce the expression *bù rǎn yú yù* 不染於欲 and present the use of the verb *rǎn* 染 in Chinese. In § 4, I discuss the Chinese occurrences of *bù rǎn yú yù* 不染於欲 attested in the Taishō Canon and compare them to their Indic parallels. In § 5, I describe the meaning and use of the verb *lipyate/lippati* in Old and Middle Indo-Aryan (5.1); I suggest that *lipyate/lippati* served as a class IV intransitive present and that the instrumental plural *kāmaiḥ/kamehi* found in Buddhist Sanskrit ultimately continues a Middle Indo-Aryan oblique plural form used as a locative (5.2). In § 6, I argue that the Chinese and Tibetan translators were aware of the semantic and grammatical ambiguity of *lipyate kāmaiḥ/kāmehi* and, therefore, specific translation strategies were used to convey the intransitive or passive meaning. In § 7 I summarise the results.

2 Passive Constructions in Indo-Aryan and Chinese

In the preceding section, 'passivity' was mentioned in reference to both Indo-Aryan and Chinese. At the outset, it seems appropriate to provide a definition of 'passive construction' in their respective domains and with respect to the relevant chronological framework. Despite containing features typical of constructions falling within the 'passive continuum', both Indo-Aryan and Chinese passives present distinct features.[3] We might want to start the discussion with Indo-Aryan, as the situation is more straightforward than in Chinese.

2.1 Passive Constructions in Indo-Aryan

In Old Indo-Aryan (OIA, i.e. Vedic),[4] finite passives are typically expressed within all the tense systems by means of 'characterized formations', distinguishing passives from bare middles (Kulikov 2006, 63). Early Vedic presents three different formations typically employed in passives: the 'passive aorist' in *-i* and *-ran*, the 'stative' in

3 For a typological account of the notion of passivity see Abraham 2006; Comrie 1988; Haspelmath 1990; Kazenin 2001; Kulikov 2011; Shibatani 1985; 2004 *inter alia*.

4 For a periodization of Indo-Aryan, see Dahl 2016, 69 fn. 7; Masica 1993, 51-3; Bubeník 1996, x.

-e and *-re* (Kümmel 1996) and the present passives with the accented suffix in *-yá-*. Since the latter is the only formation still productive in Early Middle Indo-Aryan (MIA), I will not deal with the passive aorist and the stative.[5] The creation of a passive paradigm through the suffix *-yá-*, a specialization of the Proto-Indo-European intransitive suffix *-i̯e/o-* is one of the main innovations of the Indo-Aryan verbal system (Fortson 2010, § 5.32, § 10.18). The suffix is used in the present system as a whole, including, therefore, three tenses – present, imperfect, future – and four categories of the *modus irrealis* – injunctive, subjunctive (disappearing in Early MIA), imperative and optative (Kulikov 2006, 69; Gotō 2013, § 3.7.5). Present passives are built by attaching the accented suffix *-yá-* to the root in the zero grade. The suffix *-ya-* is also used to build a class of intransitive verbs, traditionally called class IV; as a norm, the verbs belonging to this class are also built by attaching the suffix to the root in the zero grade, but the accent is placed on the root and not on the suffix. Present passives are inflected with middle endings, whereas *-ya-*presents can take both active and middle endings (Kulikov 1998a, 144; 2012, 4; Hock 2022).

Old Indo-Aryan also inherited the formation of a category of verbal adjectives built through the suffixes *-tá-* and *-ná-* (*-tá-* and *-ná-* in OIA) from Indo-Iranian, itself inherited from Proto-Indo-European *-tó-* and *-nó-*, indicating a completed action (Szemerényi 1996, § 9.6.14; Fortson 2010, § 5.61; Gotō 2013, § 3.8.3).[6] As is the case with *-yá-*passives, the *-tá-/-ná-* suffix is attached to the root in the zero grade. The *-ta* participle can serve as the verbal head of a clause, in particular when accompanied by a copular verb, with the copula typically not appearing in the third person present (Macdonell 1916, § 208; Jamison 1990; Dahl 2016, 73). Another type of verbal adjective, generally called 'gerundive' (Gotō 2013, § 3.8.4), is also relevant to Indo-Aryan passive constructions. In Vedic the main gerundive formation is built with the suffix *-ya-*, gerundives in *-enya-*, *-ā́yya-*, and *-tva-* are also attested. The other two widespread gerundive suffixes of Classical Sanskrit – i.e. *-tavyà-* and *-anī́ya-* – are attested but at first only marginally employed from Early Middle Vedic onwards finding greater attestation in later texts (Delbrück 1888, 396-402; Jamison 1984, 610; Gotō 2013, 141). The morpho-syntactic status of the gerundive is comparable to that of the *-ta* participle: as a verbal adjective it has a strong nominal character, but it can also be used as the verbal head of a clause accompanied by a copula.

5 Only few traces of the passive aorist in *-i* are preserved in Pāli (von Hinüber 2001,§ 462; Geiger 1916, § 177; Oberlies 2019, § 93).

6 See Luraghi, Inglese, Kölligan 2021 for a survey of the inflectional and derivational processes, as well as the periphrastic formations, underlying the passive voice encompassing all the branches of the Indo-European language family.

The three formations (finite passive, *-ta* participle and gerundive) share the fact that with transitive verbs they show O-orientation with verbal agreement between the nominative patient and the verb, or the verbal head in case of the verbal adjectives, whereas the agent, if expressed, is demoted to the oblique case, i.e. the instrumental, but also the genitive for the *-ta* participle and the genitive and the dative for the gerundive (at least in Early Vedic, see Hock 1986). The OIA O-oriented constructions can thus be exemplified with the following examples (after Hock 1986, 15):

(1)

a. **Present passive**
devadattena *kaṭaḥ* *kriyate*
Devadatta:INS mat:NOM make:PRS.PASS.3SG
'By Devadatta a mat is being made'.*

b. ***-ta* participle**
devadattena *kaṭaḥ* *kṛtaḥ*
Devadatta:INS mat:NOM made:NOM
'By Devadatta a mat has been made'.

c. **Gerundive**
devadattena *kaṭaḥ* *kartavyaḥ*
Devadatta:INS mat:NOM make:GRND.NOM
'By Devadatta a mat is to be made'.

* The glosses used in this paper generally follow the Leipzig Glossing Rules (https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf). In distinction from them, in Sanskrit glosses, an equal sign '=' indicates unresolved sandhi.

In Pāli, the present passive is still a productive category; however, in distinction from OIA, the opposition with the present active is only shown by the stem and not also by the ending, with the bare middle surviving only in some forms and its function partially taken over by the causative and passive (von Hinüber 2001, §§ 414-15; Geiger 1916, § 176; Oberlies 2019, 318-20). Historical forms resulting from the development that the *-yá-* suffix underwent during the passage from OIA to MIA are largely preserved. Formally, there is no morphological distinction between such historical passives and inherited class IV presents: with roots ending in consonant, passives and class IV presents both show assimilation of the semivowel *-y-* of the suffix to the preceding consonant (Geiger 1916, § 136), e.g. *vuccati* 'is being said' < OIA *ucyate* and *kuppati* 'shakes' < OIA *kupyate*. New passive stems are also formed by the addition of the suffix *-ĭ̆ya-*; in certain cases, 'double passives' are even formed by adding the passive suffix to a passive stem (von Hinüber 2001, § 458; Geiger 1916, § 175;

Oberlies 2019, § 92). The active/passive opposition by means of mere stem alternation in Pāli is exemplified in (2).

(2)

a. **Present active**

yo	*jānanto*	*pāpakammaṃ*	***karoti***
REL:NOM	know:PTCPL.PRS.NOM	evil.action:ACC	do:PRS.3SG

'One who commits evil actions consciously'. (Mil 84.13-14)

b. **Present passive**

sace	*kho*	*pana*	*karoto*	***karīyati***	*pāpaṃ*
if	indeed	but	do:PTCPL.PRS.GEN	do:PASS.PRS.3SG	evil:NOM

'If bad things happen (lit. are done) to one who does [bad things]'. (AN I.3.65.17)

As regards the other two O-oriented constructions, Pāli largely employs both the verbal adjective in *-(i)ta-/-na-* and the gerundive. The verbal adjective is preserved to a great extent in historical forms, even though the connection with the present stem has often been made opaque by phonological changes (von Hinüber 2001, §§ 492-4; Geiger 1916, §§ 197-8; Oberlies 2019, §§ 107-11). The gerundive presents a series of suffixes, both continuing the OIA suffixes and abstracted from the inherited historical forms (von Hinüber 2001, §§ 495-6; Geiger 1916, §§ 197-8; Oberlies 2019, §§ 100-5). The case syntax of the three Pāli O-oriented constructions is substantially the same as OIA (von Hinüber 2022, §§ 113, 234).

The situation in Gāndhārī is akin to Pāli: inherited present passive forms occur along with innovative forms based on the present stem and the productive suffix *-iya-* (Baums 2009, 231). The majority of *-ta* participles and gerundives are continuants of the OIA forms, but innovative forms based on the present stem are also attested (234-6).[7]

7 The Gāndhārī variety used as the administrative language of the kingdom of Kroraina in the southeastern region of the Tarim Basin in the third to fourth centuries CE, generally known as Niya Prakrit, exhibits the systematic use of an extended form of the past passive participle in *-taka-* instead of the *-ta* participle (Burrow 1937, §§ 93, 110-15). The origin of such formation presumably lies in the use of the inherited *-ta* participle as the basis of an innovative A-oriented periphrastic past construction and the consequent functional ambiguity of the inherited OIA past participle, see Jamison 2000; Barchi, Peschl 2022.

2.2 Passive Constructions in Chinese

The 'passive' definition has been applied to a wide range of constructions in Archaic and Middle Chinese[8] with different behavioural features and distinct diachronic development (Pān 1982, 247-55; Peyraube 1989; Pulleyblank 1995, 35-8; Wáng 2014, 405-22; Wèi 1994; Yáng, Hé 2001, 668-97 *inter alia*). Here, I focus only on those constructions characterized by the overt presence of grammatical markers. Therefore, I do not take into consideration those alternations of verb orientation either unmarked (such as in labile verbs) or motivated by phonological and morphological variation of the verb (see Xu 2006, 62-76 for an overview). I do not address those verbs, such as *kě* 可 (to be able), *zú* 足 (to be sufficient), *nán* 難 (to be difficult) and *yì* 易 (to be easy), typically entailing 'patient subjects' (Wáng 2014, 406).

At least three different constructions with overt marking are commonly described as passives, namely the *jiàn* 見 construction, the *wéi* 為 construction and a type of construction occurring with one of a set of 'transitive inactive verbs' with the meaning of 'to undergo' or 'to receive' (Haspelmath 1990, 40) such as *zāo* 遭, *méng* 蒙, *shòu* 受 and *bèi* 被, the latter representing the source for the Mandarin passive construction (Li, Thompson 1981, 492). Each of these can be divided into different subtypes depending on various parameters, such as the presence of an overt agent and the use of other additional markers. All the constructions originated through the grammaticalization of original verbs, even though the synchronic status of these verbs in each stage of development is much disputed. It appears that one of the conditions that allowed the grammaticalization of these transitive verbs into passive markers was their 'inward semantic meaning' (Zeng 2020, 278), in that they entail the transmission of force from a patient to an agent (cf. Chao 2011, 711).

We might want to start with the earliest attested construction (Yáng, Hé 2001, 668), namely the *jiàn* 見 construction. The verb *jiàn* possesses a full lexical meaning denoting visual perception, but it is semantically and pragmatically not neutral, being non-volitional, uncontrollable and unintentional, which are traits typically associated with 'passive' experiencers (Zeng 2020, 118; cf. Peyraube 1989, 341). The non-volitional and uncontrollable nature of *jiàn*, the association with a 'passive' experiencer, and the 'inward orientation' of the verb were presumably the basis for a semantic extension of the verb from visual perception to a more general meaning of 'to experience' and 'to encounter', see (3).

8 In this paper, I follow Aldridge's (2013a, 40) periodization of Chinese. I provide a tentative chronology of the quoted sources example by example (based on Zeng 2020, 7).

(3) 盆成括見殺.
Pén Chéngkuò *jiàn* *shā*
Pen Chengkuo encounter kill
'Pen Chengkuo was killed (lit. encountered killing)'. (*Mèngzǐ*, Jìn xīn II, fourth-third c. BCE)*

* Before continuing the discussion, I would like to introduce the Chinese corpus used in this paper. All the examples from Chinese Buddhist texts are quoted from the Taishō Edition of the Chinese Buddhist canon (Taishō Shinshū Daizokyō, hereafter = T, ed. Takakusu, Watanabe 1924-32) as contained in the CBETA electronic corpus 《大正新脩大藏經》中華電子佛典協會電子資料庫 (`https://cbetaonline.dila.edu.tw/zh/`). Non-Buddhist Chinese examples are quoted from *Zhōngguó zhéxué shū diànzǐhuà jìhuà* 中國哲學書電子化計劃 (Chinese Text Project; ed. Sturgeon 2011).

The second type of construction is built with the dynamic copula *wéi* 為 (cf. Peyraube, Wiebusch 1994), originally expressing a change of state (Wèi 1994; Zeng, Anderl 2019) see (4a). The *wéi* construction can also occur with the nominalizer *suǒ* 所 marking the element taken by the copula, see (4b).[9]

(4)

a. 止, 將為三軍獲.
zhǐ, *jiāng* *wéi* *sān* *jūn* *huò*
stop will COP three army capture
'[If] you stop, [you] will be captured by the three armies (lit. become what is captured by the three army)'. (*Zuǒzhuàn*, Xiāng Gōng 18, fifth-fourth c. BCE)

b. 負石自投於河, 為魚鱉所食.
fù *shí* *zì* *tóu* *yú* *hé,*
carry stone self throw in river
wéi *yú* *biē* *suǒ* *shí*
become fish turtle NMLZ eat
'Carrying a stone, [he] jumped into the river, becoming the food of fishes and turtles (lit. what is eaten by fishes and turtles)'. (*Zhuāngzǐ*, Dào Zhí, fourth-third c. BCE)

9 The introduction of the nominalizer *suǒ* to mark the constituent following *wéi* has been linked by Aldridge (2013b, 66) to the loss of affixional morphology marking embedded nominalization in Late Archaic Chinese. In other words, the use of *suǒ* in the *wéi* construction would have arisen to overtly mark that the constituent following *wéi* was nominal. Despite not being substantiated with direct evidence, Aldridge's proposal is certainly intriguing, because it provides a functional explanation for the emergence of the use of *suǒ* in the *wéi* construction.

The third type of construction occurs with verbs with the meaning of 'to undergo' or 'to suffer' such as *bèi* 被, see (5a). The original nominal status of the object of the verb *bèi* in Archaic Chinese can be observed by the presence in certain instances of the genitive marker *zhī* 之 between the object of the verb and the agent of the action-noun serving as the object of the verb, namely *zèn* 譖 (object) and *zhòng kǒu* 眾口 (agent) in (5b), clearly marking an adnominal relationship (Zeng 2020, 172 ff.).

(5)

a. 今兄弟被侵.
jīn *xiōngdì* *bèi* *qīn*
now brothers suffer attack
'[If his] brothers have now to undergo an attack..'. (*Hán Fēizǐ*, Wǔ dù, fifth-third c. BCE)

b. 被眾口之.
bèi *zhòng* *kǒu* *zhī* *zèn*
suffer many mouth gen slander
'[He] suffered slander from a large number of people (lit. slander of a large number of people)'. (*Hán Fēizǐ*, Jiānjié shìchén)

Although deriving from verbs with different meanings and having different paths to grammaticalization, the various passive constructions (*jiàn*, *wéi*, transitive inactive verbs) present a significant degree of syntactic convergence in Middle Chinese. For instance, a common phenomenon which represents an innovative feature at this stage of development is the presence of postverbal material after the verbal forms serving as the object of the passive verbs (Ān 2009, 135; Liú 1992, 319 ff.; Peyraube 1989, 354; Wáng 2014, 415), see the following examples in (6) quoted from Buddhist texts.[10]

(6)

a. 必見毀辱神廟.
bì ***jiàn*** *huǐrǔ* ***shénmiào***
certainly suffer revile temple
'[I] will certainly **suffer** the reviling of **the temple**'. (T 200, 254a24-5)*

* *Zhuànjí bǎi yuán jīng* 撰集百緣經 (T 200), being a Chinese translation of the *Avadānaśataka*. The Taishō Canon attributes it to *Zhī Qiān* 支謙 (fl. 223-53), but this ascription is suspect. It probably represents a later translation (sixth c. CE), see Demoto 1995.

10 The presence of postverbal complements after the putative action nouns is hardly compatible with analysing the objects of the passive verbs as nouns. Against this background, a process of reanalysis of the construction can be envisaged, postulating the

b. 如我昔為歌利王割截身.

rú	*wǒ*	*xī*	***wéi***	*Gēlì*	*wáng*	*gējié*	***shēntǐ***
like	I	formerly	COP	Kaliṅga	king	cut	body

'Like when in a past time the king Kaliṅga cut my body [into pieces] (lit. to me the King Kaliṅga cut **the body**)'. (T 235, 750b14-5)*

* *Jīngāng bānrě boluómì jīng* 金剛般若波羅蜜經 (T 235), Kumārajīva's (Ch. Jiūmóluóshí 鳩摩羅什, 344-413) transl. of the *Vajracchedikā Prajñāpāramitā*, translated in 403 CE.

c. 如彼愚人被他打頭.

rǔ	*bǐ*	*yúrén*	***bèi***	*tā*	*dǎ*	***tóu***
like	that	foolish.person	suffer	other	hit	head

'Like that foolish man getting hit in the head by other people (lit. **suffering** other people hitting **the head**)'. (T 209, 543b18)*

* *Bǎiyú jīng* 百喻經 (T 209), translated into Chinese by Guṇavṛddhi (Ch. *Qiúnàpídì* 求那毘地) in 492 CE.

A last type of construction needs to be introduced before proceeding with the discussion. It has been commonly argued that another passive construction attested in Archaic Chinese features the use of the preposition *yú* 於 to mark the agent of the passive verb (Pān 1982, 247; Peyraube 1989, 336; Wáng 2014, 407), see (7).

(7) 勞心者治人, 勞力者治於人.

láo	*xīn*	*zhě*	*zhì*	*rén,*	*láo*
work	mind	NMLZ	govern	people	work
lì	*zhě*	*zhì*	***yú***	***rén***	
strength	NMLZ	govern	by	people	

'Those who labour with their minds rule others, those who labour with their strength are ruled **by others**'. (*Mèngzǐ*, Téng Wén Gōng I)

The passive interpretation of *yú* in such instances has been disputed. As a matter of fact, the preposition *yú* does not only introduce agents, but most commonly locative complements (Pulleyblank 1986) as in (4b) above and in (8), as well as a rather wide range of other complements, including different types of undergoers (patient, recipient, benefactive, see Méi 2018, 296; Zeng 2020, 257 ff.).

reinterpretation of the object of the passive verbs from action nouns or nominalized verb-phrases (see the use of *suǒ*) into sentential objects (Aldridge 2013b; Anderl 2017, 692). From this perspective, considering the object of the passive verbs as embedded clauses would explain the presence of postverbal complements in the construction, as the verbal element serves as the verb of an embedded clause. A similar syntactic analysis has been also proposed for the Mandarin passive construction, see Hashimoto 1988.

(8) 王立於沼上.
wáng *lì* *yú* *zhǎo* *shàng*
king stand at pond above
'The king was standing above the pond'. (*Mèngzǐ*, Liáng Huì Wáng I)

As concluded by Zeng (2020, 265), the passive interpretation of examples such as (7) mainly relies on the context of the sentence rather than on the use of *yú*, which in Late Archaic Chinese appears to have developed into an oblique marker "indicating an 'indirect/loose/marginalized' relationship or some other additional information (e.g., location, agent, object of comparison)" (see also Méi 2018, 298-9 for similar considerations). In any case, the use of *yú* to mark agents in the so-called 'passive *yú* construction' is hardly attested in post-Qin sources (second c. BCE).

3 *Bù rǎn yú yù* 不染於欲: A Passive Construction?

In the middle of the second chapter of Kumārajīva's translation of the *Larger Prajñāpāramitā* (*Móhē bānrě bōluómì jīng* 摩訶般若波羅蜜經, T 223), one encounters the following passage:

(9) 是菩薩摩訶薩不染於欲.
shì *púsà* *móhēsà* *bù* *rǎn* *yú* *yù*
that bodhisattva mahāsattva NEG taint in/by desire
'That *bodhisattva mahāsattva* is not tainted in/by (?) desire'. (T 223, 221b20-1)

At first glance, the grammatical function of *yú* in (9) with respect to the semantic role introduced by it (location vs. agent) is not very clear. Thus, one can compare the passage with its Sanskrit parallel (I am using here the Nepalese recension of the *Pañcaviṃśatisāhasrikā Prajñāpāramitā*) and with the other Chinese translations of the *sūtra*, see (10).[11]

11 There are six Chinese translations of the *Larger Prajñāpāramitā*. The two earliest ones were temporally produced close to each other (late third c. CE) and were made by Mokṣala (Ch. Wúchāluó 無叉羅) in 291 CE and Dharmarakṣa (Ch. Zhú Fǎhù 竺法護) in 286 CE, namely *Fàngguāng bānrě jīng* 放光般若經 (T 221) and *Guāngzàn jīng* 光讚經 (T 222). The *sūtra* was translated a third time in 404 by Kumārajīva. The remaining three translations are by Xuánzàng (translated during 660-63 CE) and reflect the later subdivision of the text in the three versions (*Śatasāhasrikā*, *Pañcaviṃśatisāhasrikā*, *Aṣṭadaśasāhasrikā*), being the first three divisions (Ch. *huì* 會 or *fēn* 分) of Xuánzàng's monumental work in 600 fascicles, i.e. *Dà bānrě bōluómìduō jīng* 大般若波羅蜜多經 (T 220, vols 5-7, see Zacchetti 2015, 189). The example in the main text is quoted from Xuánzàng's translation of the *Pañcaviṃśatisāhasrikā* (i.e. the second division, found in vol. 7).

(10)

a.	*na*	*punar*	*bodhisattvo*
	NEG	but	bodhisattva:NOM
	mahāsattvaḥ	***kāmaguṇair***	***lipyate***
	great.being:NOM	desire.quality:INS.PL	smear:PRS.PASS.3SG

'A bodhisattva great being is not **smeared by the qualities of desire**'. (Pañca 37.8-9)*

* The Gilgit manuscript version of *Larger Prajñāpāramitā* (Zacchetti 2005, 387.17r7-9) reads: [*bodhisattvā mahāsattvā*] *na ca taiḥ* [*paṃcabhiḥ kāmaguṇaiḥ*] *sārdhaṃ saṃvasanti na lipyaṃte*.

b. 無所沾污.

wú suǒ zhānwū

NEG NMLZ smear

'There is no smearing'. (T 221, 4c14, translated by Mokṣala)

c. 其菩薩摩訶薩不為五欲之所沾污.

qí	*púsà*	*móhēsà*	***bù***	***wéi***
that	bodhisattva	mahāsattva	NEG	COP
wǔ	***yù***	***zhī***	***suo***	***zhānwū***
five	desire	GEN	NMLZ	smear

'That *bodhisattva mahāsattva* is **not smeared by the five desires**'. (T 222, 152a10-11, transl. by Dharmarakṣa)

d. 不為五欲之所染污.

bù	*wéi*	*wǔ*	*yù*	*zhī*	*suǒ*	*rǎnwū*
NEG	COP	five	desire	GEN	NMLZ	smear

'[That *bodhisattva mahāsattva*] is not smeared by the five desires'. (T 220, 11b15-16, transl. by Xuánzàng)

As one can see from the comparison with the parallels, the Sanskrit text reads a finite present passive (i.e. *lipyate*) with a non-animate instrumental agent (i.e. *kāmaguṇair*); Dharmarakṣa's and Xuánzàng's translations both present the passive construction *wéi* 為 A *zhī suǒ* 之所 V. In the light of the grammatical formations in the Sanskrit and Chinese parallels, one might conclude that the use of *yú* in Kumārajīva's translation represents an instance of the Archaic Chinese 'passive' construction with the agent introduced by the preposition.

As introduced in § 2.2, however, this type of construction died out at a relatively early stage and therefore it is very unusual to find it in Middle Chinese. As a matter of fact, in pre-Qin texts, the syntagm *rǎn yú* 染於 mainly occurs in *Mòzǐ* 墨子 (fifth-third c. BCE), where it is used with the meaning of 'to dye in', as in (11a), and, by means of

semantic extension, with the sense of 'to be under the influence of', with *yú* introducing the person under whose influence one is, as in (11b).

(11)

a. 染於蒼則蒼, 染於黃則黃.

rǎn	***yú***	***cāng***	*zé*	*cāng*,	*rǎn*	*yú*	*huáng*	*zé*	*huáng*
dye	in	blue	CONJ	blue	dye	in	yellow	CONJ	yellow

'What is **dyed in blue** becomes blue, what is dyed in yellow becomes yellow'. (*Mòzǐ*, Suǒ rǎn)

b. 舜染於許由、伯陽.

Shùn	***rǎn***	***yú***	*Xǔ Yóu*	*Bó Yáng*
Shun	be_under_influence	in	Xu You	Bo Yang

'Shun **came under the influences** of Xu You and Bo Yang'. (*Mòzǐ*, Suǒ rǎn)

In post-Qin sources, the syntagm *rǎn yú* 染於 is frequently employed and occurs almost only in Buddhist texts. See for instance example (12a) quoted from Kumārajīva's translation of the *Vimalakīrtinirdeśa* (*Wéimójié suǒshuō jīng* 維摩詰所說經, T 475). As also shown from the comparison with the Sanskrit parallel in (12b) (cf. Sk. LOC.SG. *dharme*), in this case the preposition *yú* clearly introduces a locative noun-phrase, while the meaning of the verb seems to have a 'to cling on, to be attached' semantic nuance (cf. *rǎnzhuó* 染著 'clinging').[12]

(12)

a. 若染於法乃至涅槃, 是則染著, 非求法也.

ruò	***rǎn***	***yú***	***fǎ***	*nǎizhì*	*nièpán*,		
if	be_attached	to	dharma	so_much_as	nirvāṇa		
shì	*zé*	*rǎnzhuó*	*fēi*	*qiú*	*fǎ*	*yě*	
that	CONJ	clinging	NEG	seek	dharma	FIN	

'If one **is attached to a dharma**, so much as the *nirvāṇa*, that is clinging, it is not seeking the Dharma'. (T 475, 546a16-17)

12 Xuánzàng's translation of the parallel (T 476, 570b16-17) also shows the use of *yú* to introduce a locative complement, but in this case it is fronted before the verb: *ruò* ***yú zhū fǎ*** *nǎizhì nièpán shǎoyǒu* ***tānrǎn***, *shì qiú tānrǎn, fēi wèi qiú fǎ* 若於諸法乃至涅槃少有貪染, 是求貪染, 非謂求法 (If one had the faintest attachment **to a dharma**, as much as the *nirvāṇa*, that would be seeking attachment, it would not be seeking the Dharma). The Tibetan translation of the passage (quoted from SGBSL 2004, 222) reads: *de la gang dag chos gang la chags na tha na mya ngan las 'das pa la yang rung ste, de dag ni chos 'dod pa ma yin gyi, de dag ni 'dod chags kyi rdul 'dod pa'o* (He who is attached to anything, even to liberation, is not interested in the Dharma but is interested in the taint of desire; transl. by Thurman 1976, 50).

b.	*tatra*	*ye*	***kvacid***		***dharme***	***rakṣante***
	there	REL:NOM.PL	INDF:LOC		dharma:LOC	heed:PRS.3PL
	'ntaśo	*nirvāṇe*	*'pi,*	*na*	*te*	*dharmārthikāḥ,*
	so_much_as	liberation:LOC	also	NEG	3PL:NOM	dharma.wanting:NOM.PL
	rajo'rthikās		*te*			
	taint.wanting:NOM.PL		3PL:NOM			

'In which case those who **are passionate*** **about any dharma whatsoever**, even about final release, are not those who want the Dharma, they are those who want the stain of passion'. (Vikn 5.3, transl. by Gómez, Harrison 2022, 62)

* Both the anonymous reviewers aptly pointed out that Sk. *rakṣante*, lit. 'heed', appears to be out of context at least, and that the Chinese (染) and Tibetan (*gnas*, see also example 29c below) translations rather point to such a verb as **rajyante*, which also forms a good word-play on the *rajas* in the closely following compound *rajorthikās* (see Huáng 2011, 170 fn. 4 for the same conclusion). Thus, one is probably dealing with a scribal error (*-kṣ-* < *-jy-*) here.

Another aspect to consider is that the verb *rǎn* 染 also occurs with a meaning akin to that seen in (12a), i.e. 'to cling on', 'to be attached' (even 'to long for' in this case), without the use of the preposition *yú* to introduce the object of attachment, see for instance (13a) and (13b). Note also that both constructions make use of the passive *wéi* construction as well, along with the active use of *rǎn*; (13b) even presents the passive and active uses of *rǎn* one after the other.

(13)

a. 為樂受觸, 不染欲樂.

wéi	*lè*	*shòu*	*chù,*	*bù*	***rǎn***	***yùlè***
COP	pleasant	feeling	touch	NEG	be_attached	pleasure

'[When he is] touched by a pleasant feeling, he does not become **attached to pleasure**'. (T 99, 120a27-8)*

* The passage is quoted from *sūtra* no. 470 of the Chinese *Saṃyuktāgama* (*Zá āhán jīng* 雜阿含經) translated by Guṇabhadra (Ch. *Qiúnàbátuó* 求那跋陀) mid-fifth c. CE. The *sūtra* is very close in terms of content to the *Sallattenasutta* of the *Saṃyutta Nikāya*, but the Pāli version lacks a precise parallel of the passage quoted in (13a). The closest parallel passage (SN IV.36.6.10, 209.18-19) reads: *so dukkhāya vedanāya phuṭṭho samāno kāmasukhaṁ nābhinandati* (While being touched by a painful feeling, he does not long for sensual pleasure).

b. 復為欲所染, 染欲著欲.
fù *wéi* *yù* *suǒ* *rǎn,* ***rǎn*** ***yù*** *zhuó* *yù*
moreover COP desire NMLZ taint be_attached desire cling_on desire
Moreover, [they] are tainted by desire, [they] **are attached to desire**, [they] cling on desire'. (T 26, 796a10-11)*

* The passage is quoted from *sūtra* no. 213 of the Chinese *Madhyamāgama* (*Zhōng āhán jīng* 中阿含經, T 26) transl. by Gautama Saṃghadeva (Ch. *Qùtán Sēngqiétípó* 瞿曇僧伽提婆) at the end of the fifth c. CE. The *sūtra* is close in terms of content to the *Dhammacetiyasutta* of the *Majjhima Nikāya*, but the Pāli version lacks a precise parallel of the passage quoted in (13b). The closest parallel passage (MN II.4.9, 120.14-15) reads *pañcahi kāmaguṇehi samappitā samaṅgībhūtā parivārenti* (They amuse themselves supplied and provided with the five qualities of desire).

As shown by these examples, the functional distinction expressed by *yú* with the verb *rǎn* appears to be quite labile: the forms *rǎn* 染 / *rǎn yú* 染於 are synonymous, both taking a location/goal object which can be optionally introduced by the preposition *yú* (cf. Zeng 2020, 269-72). In light of this, one should note that 'metrical' reasons could also underlie the use of *yú* in this context: Chinese translations often show a strong preference for specific patterns in terms of the number of characters – notably a preference for a four or five-character pattern (Zürcher 1977, 178) – which results into a highly 'rhythmized' text. It is probable that the tetra-syllabic form *bù rǎn yú yù* 不染於欲 suited better certain prosodic contexts than the trisyllabic equivalent *bù rǎn yù* 不染欲. As also noted by Méi (2018, 347), one can conclude that in Middle Chinese *yú* 於 did not have a strong grammatical connotation, but, at least in these examples, mainly served as a prosodic filler.

This concise survey has illustrated how the passive interpretation of *bù rǎn yú yù* 不染於欲 in (9) appears to be very dubious, despite the comparative evidence provided by the parallels. In the following sections, it will be discussed how such an erratic translation could find a possible explanation by broadening the scope of the comparative material taken into consideration and looking to a larger set of Indic parallels.

4 A Broader Focus

By fortunate chance, there are only a few occurrences of the expression *bù rǎn yú yù* 不染於欲 in the Taishō Canon, so it is possible to look at their various Indic parallels quite in detail. Besides the instance found in T 223 quoted in (9), there are other seven occurrences found in the following six texts:

1. T 26, *Zhōng āhán jīng* 中阿含經;
2. T 212, *Chūyào jīng* 出曜經;
3. T 221, *Fàngguāng bānrě jīng* 放光般若經;
4. T 309, *Zuìshèng wèn púsà shízhù chúgòu duànjié jīng* 最勝問菩薩十住除垢斷結經;
5. T 814, *Fó shuō xiàngyè jīng* 佛說象腋經;
6. T 1509, *Dà zhìdù lùn* 大智度論.

For the present discussion, I will focus on the parallels from the first two *sūtras* (T 26 and T 212), as they provide the most interesting insights into the Chinese expression and its relationship with the Indic source forms. As it will be illustrated later, the instances of *bù rǎn yú yù* 不染於欲 in T 26 and T 212 have a number of parallels attested in a group of related texts which have come down to us in various Indic languages, transmitted under the name of *Dharmapada* (*Pāli Dhammapada*) and *Udānavarga* (Nattier 2023, 216-17 for an overview). These texts represent different sectarian modifications of a collection of verses inherited from the earliest Buddhist tradition (Brough 1962, 34-41; Lenz 2003, 11-14). In the following, I will refer collectively to this group of texts as the '*Dharmapada-Udānavarga* texts'. I will comment briefly on the remaining instances of *bù rǎn yú yù* before proceeding to the analysis of T 26 and T 212.

The *Dà zhìdù lùn* 大智度論 (T 1509) is a commentary on *the Larger Prajñāpāramitā* translated by Kumārajīva (see Zacchetti 2021) and thus, not surprisingly, also contains the same expression of the root text translated by the same author. Besides the instance found in T 814 (783b14-15),[13] the remaining two instances are connected to the seventh *bhūmi* of the *bodhisattva* path. In T 221 (translated by Mokṣala) the expression belongs to a list of 40 *dharmas* (20+20) that a *bodhisattva* on the seventh *bhūmi* must avoid (first twenties) and do (remaining twenties) in order to pass to the next stage. More precisely, T 221 (27c11) reads *bù rǎn yú yù shì* 不染於欲事 (not attached to sensual matters) as the last *dharma* of the second group of twenty *dharmas*. As regards the other Chinese translations of the *Larger Prajñāpāramitā*, the expression corresponds to T 222 (196c20, Dharmarakṣa) *wú suǒ*

13 The Taishō Canon ascribes the translation to Dharmamitra (Ch. *Tánmómìduō* 曇摩蜜多, d. 442 CE), but the attribution is dubious, cf. Silk 2010, 376 fn. 23.

rǎnwū 無所染污 (without tainting) (twentieth *dharma* of the first group), to T 223 (257b18, Kumārajīva) *bù rǎn ài* 不染愛 (not attached to desire) (twentieth *dharma* of the second group) and to T 220 (83b25, Xuánzàng) *yīng yuánmǎn wú suǒ àirǎn* 應圓滿無所愛染 (perfectly [and] completely without the taint of desire) (nineteenth *dharma* of the second group). Although all the Chinese parallels seem to agree in listing the very same element towards the end of the second group of dharmas (in T 222 it is however placed at the end of the first), no trace of a parallel element appears to exist in the parallel passage of Nepalese recension of the *Pañcaviṃśatisāhasrikā* (Pañca 216.8 ff.). Lamotte (1980, 2430) provides the Sanskrit reading *akliṣṭo 'nunayaḥ* (unafflicted affection)for the last dharma of the second group in the *sūtra* quotation of his translation of the *Dà zhìdù lùn*. This reading seems indeed to be a genuine parallel of the *dharma* in the Chinese translations, but I ignore whether Lamotte used a Sanskrit parallel from a different version of the *Larger Prajñāpāramitā*,[14] or even from a different text (cf. Lamotte 1980, x), or he back-translated it into Sanskrit from the Chinese form. Even though the Indic source expression underlying the Chinese translations of this *dharma* is not entirely clear, it is interesting to see that the four translations show a certain degree of polarization between the two meanings 'to be tainted' and 'to be attached' already seen in the case of (9) and (10). Note also that in this case Kumārajīva employs the form *bù rǎn ài* 不染愛 (not attached to desire) without the use of the preposition *yú*. The last occurrence of *bù rǎn yú yù* appears in T 309 (978a18) authored by Zhú Fóniàn 竺佛念 (Nattier 2010; Lin, Radich 2021) in the same context of the enumeration of the dharmas to be performed in the seventh *bhūmi*.

The first occurrence of the expression among the two other *sūtras* is in a verse passage of *sūtra* no. 28 of the Chinese *Madhyamāgama* as given in (14a). Other Chinese parallels of the same verse passage are found in the two Chinese translations of the *Saṃyuktāgama*, namely

14 The label *Larger Prajñāpāramitā* denotes what Zacchetti (2005, 36; 2021, 23) called a 'textual family', with the sense of "a group of texts that share a number of common features in structure, content, wording, etc. They exhibit a family resemblance, so to speak, fluid and not always easy to define, but significant enough to set them apart from other texts […] as a distinct group" (2005, 36). The prototype from which the texts belonging to this family stemmed was probably rather fluid in the earliest phase (third-fifth c.), ranging from 17,000 to 22,000 stanzas (Zacchetti 2015, 185). The version represented by the Gilgit *Larger Prajñāpāramitā* bears witnesses in terms of size to this stage of textual development, besides showing a close relationship with the recension of text commented in the *Dà zhìdù lùn* (Zacchetti 2021, 82 ff.); unfortunately, the Gilgit *Larger Prajñāpāramitā* remains largely unedited. In later times, the text saw a process of development and expansion, with the canonical subdivision in the three versions in 100,000 stanzas (*Śatasāhasrikā*), in 25,000 (*Pañcaviṃśatisāhasrikā*) and in 18,000 stanzas (*Aṣṭādaśasāhasrikā*) reflected in Xuánzàng's translations. As a very large number of Sanskrit fragment manuscripts of the *Larger Prajñāpāramitā* have survived (cf. Zacchetti 2005, 17-19 fnn. 53-4), I am not able here to look at all the possible parallels.

in *sūtra* no. 592 of *Zá āhán jīng* 雜阿含經 (T 99), see (14b), and *sūtra* no. 186 of *Biéyì zá āhán jīng* 別譯雜阿含經 (T 100, translated between the second half of fourth c. and the first half of the fifth c. CE by an unknown translator), see (14c). Parallels of the passage are also found in the two Chinese translations of the *Udānavarga,* namely *Chūyào jīng* 出曜經 (T 212, translated by Zhú Fóniàn 竺佛念 in the late fourth c. CE), see (14d), and *Fǎjí yàosòng jīng* 法集要頌經 (T 213, translated by Tiānxīzāi 天息災 in the late tenth c. CE), see (14e).

(14)

a. 如梵志滅度, 以不染於欲, 捨離一切願, 逮得至安隱.
rú *fànzhì* *mièdù,* *yǐ* *bù* ***rǎn*** ***yú*** ***yù***
like brāhmaṇa extinguish CONJ NEG be_attached to desire
shělí *yīqiè* *yuàn* *dàidé,* *zhì* *ānyǐn*
be_free all aspiration reach arrive tranquillity
'Like a *brāhmaṇa** who is calmed, because not **attached to desire**, getting rid of all the aspirations, he attains tranquillity'. (T 26, 460b16-17)

* *Fànzhì* 梵志 lit. 'Brahmā-mind'. See Karashima 2016 for this folk-etymology-based translation of *brāhmaṇa*.

b. 婆羅門涅槃, 是則常安樂, 愛欲所不染, 解脫永無餘.
póluómén *nièpán,* *shì* *zé* *cháng* *ānlè,*
brāhmaṇa extinguish that CONJ always ease
àiyù ***suǒ*** *bù* ***rǎn,*** *jiětuō* *yǒng* *wú* *yú*
desire NMLZ NEG taint liberate ever NEG remainder
'A *brāhmaṇa* who is calmed is always at ease, not **tainted by desire**, completely liberated forever'. (T 99, 158a27-28)

c. 一切事安樂, 婆羅門涅槃, 無為欲所污, 解脫於諸有.
yīqiè *shì* *ānlè,* *póluómén* *nièpán,*
all matter ease brāhmaṇa extinguish
wú ***wéi*** ***yù*** ***suǒ*** ***wū,*** *jiětuō* *yú* *zhū* *yǒu*
NEG COP desire NMLZ stain liberate at INDF exist
'At ease in every matter is a *brāhmaṇa* who is calmed; not **stained by desire**, he is freed in every matter'. (T 100, 441a7-8)

d. 一切得善眠, 梵志取滅度, 不為欲所染, 盡脫於諸處.
yīqiè *dé* *shàn* *mián,* *fànzhì* *qǔ* *mièdù,*
all be_able well sleep brāhmaṇa seize extinction
bù ***wéi*** ***yù*** ***suǒ*** ***rǎn,*** *jìn* *tuō* *yú* *zhū* *chù*
NEG COP desire NMLZ taint completely liberate at INDF aspect
'Able to sleep well in every circumstance is a *brāhmaṇa* who has seized extinction, not **tainted by desire** and completely liberated in every regard'. (T 212, 756c8-9=757a4-5)

e-ISSN 2785-5953

e. 一切得安隱, 梵志取滅度, 不為欲所染, 盡脫於諸處.

yīqiè	*dé*	*ānyǐn,*	*fànzhì*		*qǔ*	*mièdù,*			
all	attain	tranquillity	brāhmaṇa		size	extinction			
bù	***wéi***	***yù***	***suǒ***	***rǎn,***	*jìn*	*tuō*	*yú*	*zhū*	*chù*
NEG	COP	desire	NMLZ	taint	completely	liberate	at	INDF	aspect

'Always attains tranquillity a *brāhmaṇa* who has seized extinction, not **tainted by desire** and completely liberated in every matter'. (T 213, 794c23-4)

Precise parallels of the passage exist both in Pāli, see the passage from the *Cullavagga* given in (15a), and in Buddhist Sanskrit, see example (15b) from the *Saṃghabhedavastu,* example (15c) from the Sanskrit *Udānavarga* and example (15d) from the *Udānavarga* from Subaši.

(15)

a.

sabbadā	*ve*	*sukhaṃ*		*seti,*
always	truly	at_ease		rest:PRS.3SG
brāhmaṇo	*parinibbuto,*			
brahman:NOM	completely_calmed:NOM			
yo	*na*	***lippati***		***kāmesu,***
REL:NOM	NEG	be_attached:PRS.3SG		desire:LOC.PL
sītibhūto	*nirūpadhi*			
dispassionate:NOM	controlled:NOM			

'Always rests at ease a brahman who is completely calmed, one who is not **attached to sensual pleasures,** dispassionate and controlled'. (Culv 6.4.4 = MN II.5.8)*

* Pāli parallels are found also in SN I.10.8.15 and AN I.3.34, 138.3-4.

b.

sarvathā	*vai*	*sukhaṃ*	*śete,*	
in_everyway	truly	at_ease	rest:PRS.3SG	
brāhmaṇaḥ	*parinirvṛtaḥ,*			
brahman:NOM	completely_calmed:NOM			
yo	*na*	***lipyate***		***kāmebhir***,
REL:NOM	NEG	smear:PRS.PASS.3SG		desire:INS.PL
vipramukto	*nirāsravaḥ*			
liberated:NOM	sinless:NOM			

'In every circumstance rests at ease a brahman who is completely calmed, one who is not **smeared by sensual pleasures**, liberated and without sins'. (Ud 30.28).

c.

sarvathā	*vai*	*sukhaṃ*	*śete,*
in_everyway	truly	at_ease	rest:PRS.3SG
brāhmaṇaḥ		*parinirvṛtaḥ,*	

brahman:NOM		completely_calmed:NOM		
lipyate		*yo*	*na*	***kāmair,***
smear:PRS.PASS.3SG		REL:NOM	NEG	desire:INS.PL
hi	*vipramukto*			*nirupadhiḥ*
for	liberated:NOM			controlled:NOM

'In every circumstance rests at ease a brahman who is completely calmed, one who is not **smeared by sensual pleasures**, liberated and controlled'. (Saṅghabh, 169.16-17)

d.

sarvvato	*vai*	*sukhaṃ*	*śeti,*	
in_everyway	truly	at_ease	rest:PRS.3SG	
brāhmaṇā		*parinirvṛtaḥ,*		
brahman:NOM		completely_calmed:NOM		
yo	*na*	***lipyati***		***kāmehi,***
REL:NOM	NEG	smear:PRS.PASS.3SG		desire:INS.PL
vippramuktan		*niropadhiḥ*		
liberated:NOM		controlled:NOM		

'In every circumstance rests at ease a brahman who is completely calmed, one who is not **smeared by sensual pleasures**, liberated and controlled'. (UdS 423)

The second occurrence is found in *Chūyào jīng* 出曜經, quoted in (16a), one of the Chinese translations of the *Udānavarga*. A Chinese parallel is found in the other translation of the same text (i.e. *Fǎjí yàosòng jīng* 法集要頌經), see (16b). A Sanskrit parallel of the passage occurs in the Sanskrit *Udānavarga,* see (16c).

16.

a. 如月清明, 懸處虛空, 不染於欲, 是謂梵志.

rú	*yuè*	*qīngmíng,*	*xuán*	*chù*		*xūkōng,*
like	moon	bright	hang	place		sky
bù	***rǎn***	***yú***	***yù,***	*shì*	*wèi*	*fànzhì*
NEG	be_attached	to	desire	that	be_called	brāhmaṇa

'Like the moon, clear and bright, hanging in the sky, [one who] is not **attached to desire** is called a *brāhmaṇa*'. (T 212, 771c20-1=771c25)

b. 如月清明朗, 懸處於虛空, 不染於愛欲, 是名為梵志.

rú	*yuè*	*qīng*	*mínglǎng,*	*xuán*	*chù*	*yú*		*xūkōng,*
like	moon	bright	clear	hang	place	on		sky
bù	***rǎn***	***yú***	***àiyù,***	*shì*	*míng*	*wéi*	*fànzhì*	
NEG	be_attached	to	desire	that	name	COP	brāhmaṇa	

'Like the moon, clear and bright, hanging in the sky, [one who] is not **attached to desire** is called a *brāhmaṇa*'. (T 213, 798c4-5)

c.	*candro*	*vā*	*vimalaḥ*		*śuddho*,	
	moon:NOM	or	bright:NOM		pure:NOM	
	viprasanno		*hy*	*anāvilaḥ*,		
	unperturbed:NOM		for	clear:NOM		
	na	***lipyate***		*yo*	*hi*	***kāmair***,
	NEG	smear:PRS.PASS.3SG		REL:NOM	for	desire:INS.PL
	bravīmi	*brāhmaṇaṃ*		*hi*	*tam*	
	call:PRS.1SG	brahman:ACC		for	3SG:ACC	

'[Like] the moon is bright, pure, unperturbed and clear, whoever is not **smeared by sensual pleasures**, him I call a brahman'. (Ud 33.31A)

A last group of examples needs to be quoted: in this case, the Chinese parallels of the passage in T 212 and T 213 do not include the syntagm *bù rǎn yú yù* 不染於欲 (the *wéi* 為 construction is employed in its place), see (17a) and (17b).[15] However, a precise parallel containing the variant *yú yù bù rǎn* 於欲不染 is included in a Chinese translation of a commentary on the *Vinaya*, namely *Shànjiànlǜ pípóshā* 善見律毘婆沙 (T 1462), see (17c).[16]

17.

a.	猶如眾華葉, 以鍼貫芥子, 不為欲所染, 是謂名梵志。							
	yóurú	*zhòng*	*huá*	*yè*,	*yǐ*	*zhēn*	*guàn*	*jièzǐ*,
	like	many	lotus	leaf	CONJ	awl	pass_through	mustard.seed
	bù	***wéi***	***yù***	***suǒ***	***rǎn***,	*shì*	*wèimíng*	*fànzhì*

15 Besides the two Chinese translations of the *Udānavarga* introduced above (i.e. T 212 and T 213), there are two extant Chinese compilations of the *Dharmapada*, namely the *Fǎjù jīng* 法句經 (T 210, translated by Zhú Jiāngyán 竺將炎 in 224 CE and subsequently revised by Zhī Qiān by supplementing it with material drawn from other sources, see Nattier 2023) and *Fǎjù pìyù jīng* 法句譬喻經 (T 211, translated by Fǎjù 法炬 and Fǎlì 法立 during 290-306 CE). These two translations do not contain precise parallels of the verses quoted in this paper (cf. Willemen 1974). Notwithstanding, one verse from chapter 35 of T 210 (572c19-20, corresponding to the *Brāhmaṇavagga*) appears to read a possible translation of *na lipyate kāmaiḥ* rendered by means of the *wéi* construction: *xīn qì èfǎ, rú shé tuōpí,* ***bù wéi yù wū****, shì wèi fànzhì* 心棄惡法, 如蛇脫皮, 不為欲污, 是謂梵志 (One whose mind has abandoned evil dharmas, like a snake liberating himself from its skin, **not contaminated by desire**, that is called a *brāhmaṇa*). The simile of the snake liberating itself from its skin is generally found in numerous verses of the *Bhikṣuvarga* in various Indic parallels (e.g. Sanskrit *Udānavarga*, *Udānavarga* from Subaši, Khotan *Dharmapada*, London *Dharmapada*) or as a separate section (*Uraga*) in the Patna *Dhammapada* (PDhp 209 ff.), or even as a separate sutta in the Pāli *Suttanipāta* (*Uragasutta*, Snp 1-3). None of the verses found among those parallels, however, appear to correspond to the Chinese verse discussed here.

16 Translated by Saṃghabhadra (Ch. Sēngqiébátuóluó 僧伽跋陀羅) in 488-9. *Shànjiànlǜ pípóshā* 善見律毘婆沙 (**Sudarśanavinayavibhāṣā*), partially corresponds to the Pāli *Samantapāsādikā*, a commentary on the *Vinaya* attributed to Buddhaghosa (fifth c. CE), cf. von Hinüber 1996, § 209; Heirman 2004.

NEG	COP	desire	NMLZ	smear	that	call		brāhmaṇa

'Like [water is not attached to] the leaves of the lotus, [like] with an awl going through mustard seeds [the latter do not adhere to the point of the former], [one who] is not **contaminated by desire** is called a *brāhmaṇa*'. (T 212, 771c3-4=771c9-10)

b. 猶如眾華葉, 以針貫芥子, 不為欲所染, 是名為梵志。

yóurú	*zhòng*	*huá*	*yè,*	*yǐ*	*zhēn*	*guàn*		*jièzǐ,*
like	many	lotus	leaf	with	awl	pass_through		mustard.seed
bù	***wéi***	***yù***	***suǒ***	***rǎn,***	*shì*	*míng*	*wéi*	*fànzhì*
NEG	COP	desire	NMLZ	taint	that	call	COP	brāhmaṇa

'Like [water is not attached to] the leaves of the lotus, [like] with an awl going through mustard seeds [the latter do not adhere to the point of the former], [one who] is not **tainted by desire** is called a *brāhmaṇa*'. (T 213, 798b29-c1)

c. 如蓮華在水, 芥子投針鋒, 若於欲不染, 我名婆羅門.

rú	*liánhuá*		*zài*	*shuǐ,*	*jièzǐ*		*tóu*	*zhēnfēng*
like	lotus.flower		on	water	mustard.seed		lodge	awl.point
ruò	***yú***	***yù***	***bù***	***rǎn***,	*wǒ*	*míng*	*póluómén*	
REL	to	desire	NEG	be_attached	I	name	brāhmaṇa	

'Like a lotus flower on water, or mustard seeds sticking to the point of an awl, one who is **not attached to desire**, him I call a *brāhmaṇa*'. (T 1462, 725a17-18)

For the purpose at hand, this passage is particularly relevant with regard to the Indic side of the discussion, since the majority of the *Dharmapada-Udānavarga* texts that have survived in Indic languages include a parallel of the passage. Parallels are found in Pāli, see (18a) from the Pāli *Dhammapada*, in the Hybrid Prakrit variety represented by the Patna *Dhammapada*,[17] (18b), in Buddhist Sanskrit, as in the *Bhikṣuṇī-Vinaya* of the Mahāsāṃghikas and in the Sanskrit

17 The Patna *Dhammapada* represents an interesting case among early Indic Buddhist texts, as it is written in a particular Prakrit variety more Sanskritized than Pāli but not as Sanskritized as the texts belonging to 'Group 1' and 'Group 2' in Edgerton's (1953, xxv) classification of Buddhist Hybrid Sanskrit sources (Dimitrov 2020, 79 ff.; von Hinüber 1989, 362-6; Norman 1989; Roth 1980). The text can be attributed with a certain degree of certainty to the Saṃmitīya school (Skilling 1997). Considering that the *Dharmapada* is a canonical text, it is reasonable to postulate the existence of a Saṃmitīya canon written in the same canonical language of the Patna *Dhammapada* (Dimitrov 2020, 162) and some traces of other texts in such language have been indeed recently discovered (Dimitrov 2020, 162 ff.; Tournier 2023). Dimitrov (2020, 155 ff.) has proposed the name 'Saindhavī' for this Prakrit variety, also arguing that such label was used by the Saṃmitīya communities themselves, a claim that has not been however accepted by all scholars (cf. Tournier 2023, 440 fn. 116).

Udānavarga, (18c) and (18d), and in Gāndhārī, (18e) from the Khotan *Dharmapada*.[18]

18.

a.	*vāri*		*pukkharapatte*	*va,*	
	water:NOM		lotus.leaf:LOC	or	
	āragge-r-iva		*sāsapo,*		
	point_of_awl:LOC-like		mustard_seed:NOM		
	yo	*na*	*lippati*	***kāmesu,***	
	REL:NOM	NEG	be_attached:PRS.3SG	desire:LOC.PL	
	tam	*ahaṃ*	*brūmi*	*brāhmaṇaṃ*	
	3SG:ACC	1SG:NOM	call:PRS.1SG	brahman:ACC	

'Whoever **does not cling to sensual pleasures**, just as water does nor cling to a lotus leaf, or a mustard seed to the point of an awl, him I call a brahman'. (Dhp 401,* transl. by Norman 1997b, 57)

* Pāli parallels are also found in Snp 625 = Sp I, 273.5

b.	*vārī*		*pokkharapatte*	*vā,*	
	water:NOM		lotus.leaf:LOC	or	
	ārāgre-r-iva		*sāsavo,*		
	point_of_awl:LOC-like		mustard_seed:NOM		
	yo	*na*	***lippati***	***kāmesu,***	
	REL.NOM	NEG	be_attached:PRS.3SG	desire:LOC.PL	
	tam	*ahaṃ*	*brūmi*	*brāhmaṇaṃ.*	
	3SG.ACC	1SG:NOM	call:PRS.1SG	brahman:ACC	

'Like water [is not attached to] a lotus leaf, or a mustard seed to the point of an awl, one who is not **attached to sensual pleasures**, him I call a brahman'. (PDhp 38)

c.	*vāri*	*puṣkarapatre*	*vā,*	
	water:NOM	lotus.leaf:LOC	or	
	ārāgre	*iva*	*sarṣapaḥ,*	
	point_of_awl:LOC	like	mustard_seed:NOM	
	yo *na*	***lipyati***	***kāmeṣu,***	

18 The remaining attested Indic text-fragments of *Dharmapada-Udānavarga* texts, namely the aforementioned *Udānavarga* from Subaši, the London *Dharmapada* (Lenz 2003) and the Gāndhārī *Dharmapada* from the Split Collection (Falk 2015), do not contain parallels of the verse analysed in (18) – or at least the parallel verse has not survived. A final occurence of the expression *na lipyate kāmaiḥ* is contained in verse 37 of the *Brāhmaṇavarga* from the Sanskrit *Udānavarga* (Ud 33.37): *ākāśam iva paṅkena, rajasā candramā iva, na lipyate yo hi kāmair, bravīmi brāhmaṇaṃ hi tam* (Like sky by dirt, or moon by impurity, one who is not smeared by sensual pleasures, him I call a brahman). All the other Indic texts, as well as the translations in other languages, do not seem to have a parallel of this passage (cf. Willemen 1974, 49).

REL.NOM	NEG	be_attached:PRS3SG		desires:LOC.PL
tan	*me*	*śakra*	*varaṃ*	*dada*
3SG.ACC	1SG:GEN	Śakra:VOC	favour:ACC	give:IMP.2SG

'Like water [is not attached to] a lotus leaf, or a mustard seed to the point of an awl, one who is not **attached to sensual pleasures**, him, O Śakra, I ask you to favour'. (BhīVin 148.6)

d.

vāri	*puṣkarapatreṇevārāgreṇeva**			*sarṣapaḥ*,
water:NOM	lotus.leaf:INS=like=point_of_awl:INS=like			mustard_seed:NOM
na	***lipyate***	*yo*	*hi*	***kāmair***
NEG	smear:PRS.PASS.3SG	REL:NOM	indeed	desire:INS.PL
bravīmi	*brāhmaṇaṃ*	*hi*	*tam*	
call:PRS.1SG	brahman:ACC	indeed	3SG.ACC	

'Like water by a lotus leaf, or a mustard seed by the point of an awl, one who is not **smeared by sensual pleasures**, him I call a brahman'. (Ud 33.30)

* As aptly pointed out by one of the anonymous reviewers, the use of the instrumental singular here instead of the locative as in all the other parallels is awkward both syntactically and content-wise. It probably represents an emendation triggered by INS. PL. *kāmair* in the second hemistich.

e.

vari	*puṣkarapatre*	*va*	*arage-r-iva*		*sarṣava,*
water:NOM	lotus.leaf:LOC	or	point_of_awl:LOC-like		mustard_seed:NOM
yo	*na*	***lipadi***		***kamehi***,	
REL:NOM	NEG	smear:PRS.PASS.3SG		desires:INS.PL	
tam	*ahu*	*bromi*	*brammaṇa.*		
3SG:ACC	1SG:NOM	call:PRS.1SG	brahman:ACC		

'Like water [is not attached to] a lotus leaf, or a mustard seed to the point of an awl, one who is not **smeared by sensual pleasures**, him I call a brahman'. (Dhp G^K 21)

In the light of the set of examples quoted above, two main points are to be highlighted. Firstly, as is the case with the passage from the Chinese translations of *Larger Prajñāpāramitā*, examples (9) and (10) above, in the various Chinese translations the syntagm *bù rǎn yú yù* 不染於欲 alternates with the passive *wéi* 為 construction in translating the same Indic source expression. In the case of T 212 and T 213, the two constructions even alternate in the very same text. Secondly, the various Indic parallels agree with each other almost verbatim, except for the case endings of the word stem *kāma-*. The two forms with the instrumental and locative are distributed quite distinctly among Sanskrit and Gāndhārī, on one side, and Pāli (and some Hybrid texts), on the other side **[tab. 1]**. Against this background, the oscillation in the Chinese translations between the passive construction and construction with the locative complement introduced by *yú* is worthy of attention, in that a similar semantic and

grammatical opposition of the verb complements is also observable in the Indic sources.

Table 1 Instrumental vs. locative marking

Instrumental		**Locative**	
Ud:	*lipyate kāmebhir/ kāmair*	Culv:	*lippati kāmesu*
Saṅghabh:	*lipyate kāmair*	Dhp:	*lippati kāmesu*
UdS:	*lipyati kāmehi*	PDhp:	*lippati kāmesu*
Dhp G^{K}:	*lipadi kamehi*	BhīVin:	*lipyati kāmeṣu*

5 The Elusive Meaning of *lipyate*/*lippati*

The Indic parallels discussed above present two distinct patterns with the verb *lipyate*/*lippati* which, after Kulikov (2012, 208), we can summarize as follow:

[i] 'to stick' S_{NOM} sticks to R_{LOC};
[ii] 'to be smeared' R_{NOM} is smeared with/by S_{INS}.

In the first pattern, the subject of attachment is in the nominative and the locative encodes the recipient/object of attachment. In the second pattern, the nominative expresses the recipient/object of smearing and the instrumental expresses the instrument of smearing. The two patterns also bear a similar meaning, as being 'attached to sensual pleasures' can be seen as semantically contiguous to being 'tainted' by them. Nonetheless, the locative and instrumental formally encode semantic roles that are clearly different, a curious fact in light of the distribution of the two patterns in Buddhist sources. What is more, one finds it difficult to explain how a *-ya*-present could serve as a present passive with an instrumental agent and simultaneously be used intransitively with a locative complement without any apparent morphological modification. Given this peculiar opposition, we might want to look at the use of the verbal root *lip-* in Indo-Aryan in greater detail.

5.1 The Indo-Aryan Root *lip-*: Meaning and Case-Marking

Indo-Aryan *lip-* is derived from the PIE root **leip-*, whose basic meaning is 'to be sticky', 'to adhere' (Mayrhofer 1996, 460; Rix 2001, 408; Werba 1997, 228). Old Indo-Aryan continues the Indo-European root meaning by means of the two patterns introduced above (Kulikov 2012, 208-10). The earliest instance of pattern [i] is represented by (19a), quoted from the *Ṛgveda*, in which the *-ta* participle *riptám* (from the variant root form *rip-*) occurs with a locative complement expressing

the recipient of attachment/smearing. The same pattern is attested also with the present *lipyate* in the *Vājasaneyisaṃhitā* of the *Yajurveda* (repeated also in the *Īśopaniṣad*, cf. Thieme 1965, 90-1), see (19b). As noted by Kulikov (2012, 209), another instance of *lip-* occurs in the compound *vilipyáte* attested in the *Maitrāyāṇī Saṃhitā* (19c); here, the verb denotes the meaning of 'to come unstuck' and occurs with a subject of attachment/smearing, so it can be considered an instance of pattern [i].

(19)

a. *yád* *vā* ***svárau*** ***svádhitau*** ***riptám*** *ásti*
REL:NOM or post:LOC axe:LOC smeared:NOM be:PRS.3SG
'Or what is **smeared on the sacrificial post or on the axe**'. (ṚV 1.162.9, transl. by Jamison, Brereton 2014, 345)

b. *evấṃ* *tváyi* *nấnyáthetó*
thus 2SG:LOC NEG=otherwise=hence
'sti *ná* *kárma* ***lipyate*** ***náre***
exist:PRS.SG NEG action:NOM cling:PRS.3SG man:LOC
'Thus, in this way and not otherwise, (the action) is in you, (and yet) the action **does not stick to the man** (that you are)'. (VS 40.2 = ĪśUp 2. Transl. by Kulikov 2012, 209)

c. *skándati* *vấ* *etád* *dhavír* *yád* *viścótati*
split:PRS.3SG or that:NOM offering:NOM rel:NOM drop:PRS.3SG
yád ***vilipyáte***
REL:NOM come_unstuck
'That offering is spilt when it drops away or when it **comes unstuck**'. (MS III.9.7, 125.10-11=126.14-15=III.10.1, 130.4. Transl. by Kulikov 2012, 209)

The second pattern is attested from the *Brāhmaṇas* onwards, see (20a), quoted from the *Śatapathabrāhmaṇa*. The active counterpart with a nominative agent of smearing, an accusative recipient and instrumental of substance of smearing is also attested, see (20b).[19]

19 The *Ṛgveda* has also an instance of the perfect *riripúr* (5.85.8) used with the sense of 'to cheat'. Such a meaning is argued to be derived from an admittedly not very compelling semantic extension of the meaning 'to smear' (Grassmann 1873, 1165; Kümmel 2000, 428). Alternatively, as claimed by Thieme (1995, 538 fn. 14), this usage represents a denominal verbal root ("radix postnominalis" in Thieme's terms) homonym with the one continuing PIE **leip-*. The denominal root would have been abstracted from the adjective/noun *ripú-* (deceiftul, enemy), in turn a dissimilated form from **riripú-* < *rap-* (to chatter). The form *riripúr* aside, the only other instance of *lip-* attested in the *Ṛgveda* is the aorist middle *alipsata* occurring in 1.191.1, 3 and 4, where it follows the preverb *ní*. Thus, also in this case, *lip-* is used intransitively with a sense of 'to be attached, to cling on', which by means of the preverb *ní* attains the opposite meaning of 'to disappear' < 'to become unstuck', cf. Narten 1964, 26; Kulikov 2012, 210-11.

(20)

a. *na* *kármaṇā* *lipyate* *pā́pakena*
NEG action:INS smear:PRS.3SG evil:INS
'[He] is not smeared (i.e. tainted) by an evil action'. (ŚB 14.7.2.28)

b. *rudra* *oṣadhīr* *viṣeṇālimpat*
Rudra:NOM plant:ACC.PL poison:INS=smear:IPRF.3SG
'Rudra smeared the plants with poison'. (KS 6.5, 53.12)

In later Sanskrit sources, pattern [i] is significantly less common than pattern [ii] – Kulikov (2012, 210) even claims that it "seems to disappear" – being continued by other synonymous verbs such as *śliṣyate*, cf. (28) below. Pattern [i] is alive and well in Pāli and Buddhist Hybrid Sanskrit, as shown above and further illustrated below, but it is certainly true that the typical form found in standard Sanskrit is pattern [ii], see example (21) quoted from the *Bhagavadgītā*.[20]

(21) ***lipyate*** *na* *sa* ***pāpena*** *padmapatram* *ivāmbhasā*
smeared:PRS.PASS.3SG NEG 3SG:NOM sin:INS lotus.leaf:NOM like=water:INS
'He is not smeared (i.e. tainted) by sin like the leaf of the lotus [is untouched] by water'. (Bhag 5.10)

As rightly observed by Kulikov, pattern [i] certainly does not represent a passive, but rather denotes a non-passive intransitive ('anticausative', more precisely) expressing a spontaneous process, or better the state resulting from this spontaneous process ('becomes attached' > 'is attached'). The case is slightly more complicated with pattern [ii]: in § 2.1, we saw that, as a rule, present passives and class IV *-ya-*presents, are distinguished by the position of the accent, i.e. accented suffix in passives and accented root in class IV presents. The only accented instance of *lipyate* we possess is the one in (19c), which, despite the accented suffix, appears to represent a non-passive intransitive. Moreover, Kulikov notes that pattern [ii] should be more correctly described as the anticausative (rather than the passive) counterpart of active instances such as (20b), since the instrumental denotes the

20 A quick search for *lipyate* in GRETIL (`https://gretil.sub.uni-goettingen.de/gretil.html`) shows that the ratio of instrumental to locative is overwhelmingly in favour of the former. One of the few occurrences of a possible instance of pattern [i] in Classical Sanskrit I was able to locate is the following passage from the *Mahābhārata* (13.1.37): *asaty api kṛte kārye neha pannaga lipyate* (O serpent, when an evil act is done, the doer is not implicated in that [lit. does not cling on it]). However, the verb *lipyate* could be here also understood as taking a coreferentially deleted instrumental referring to *asat-*.

instrument and not the agent of smearing. Indeed, doublets such as (20a) and (20b) do not present any promotion of a hypothetical agent to the oblique case as one would expect from a prototypical passive; instead, the instrumental invariably marks the instrument of smearing in both cases, active and (pseudo-)passive. These pieces of evidence lead Kukikov (1998b, 347-8; 2012, 720) to conclude that OIA *lipyate* belongs to a group of OIA *-ya*-presents characterized by fluctuating accentuation between the root and the suffix, even though the form with root accentuation is by chance unattested.[21]

Pattern [i] and [ii] are both attested in Pāli, but in contrast with Sanskrit, it is pattern [i] which has the higher frequency. Some examples of pattern [i] have already been provided in § 4, see also (22a) below. Pattern [ii] is also relatively common, see for instance (22b), as well as its active counterpart in (22c).

(22)

a. *so* *ubh'anta-m-abhiññāya* ***majjhe*** *mantā* *na* ***lippati***
3SG:NOM both.end:ACC.PL-know:GER middle:LOC thinker:NOM NEG stick:PRS.3SG
'That thinker, knowing both ends, does not cling to the middle'. (Snp 1042, transl. by Norman 2001, 132)

b. *akāmakaraṇīyasmiṃ* *kuvidha* ***pāpena*** ***lippati***
involuntary.act:LOC where sin:INS smear:PRS.PASS.3SG
'Where in an involuntary act is one smeared by sin?' (Jā V.528, 237.139)

c. ***padumaṃ*** *yathā* ***agginikāsiphālimaṃ***
lotus:ACC like fire.resembling.blossoming:NOM
na *kadamo* *na* *rajo* *na* *vāri* ***limpati***
NEG mud:NOM NEG dust:NOM NEG water:NOM smear:PRS.3SG
'Like mud, dust and water do not smear a lotus fully blossoming like fire'. (Jā III.397, 320.6)

Despite only a handful of examples occurring in the texts, Gāndhārī also presents both patterns. We have already seen pattern [ii] in (18e). Another instance of such sort can be found in section no. 19 of a Gāndhārī Commentary edited by Baums (2009); here the *-ta* participle *aṇoalito* (unsmeared; Sk. *anupaliptaḥ*) from the root verse quotation is explained by the commentator using the present *lipadi*.

21 As discussed by Kulikov (1997; 1998a; 1998b; 2012), the verbs belonging to this group also show semantic affinity, expressing what Kulikov labels as 'entropy increase', such as destruction and destructuring. The root form *lipyate* does not appear to be semantically related to this group, but the association could have happened via the compound forms *vilip-* and *nilip-* which denote processes akin to destructuring.

(23) ***jaleṇa*** ***pakeṇa*** ***aṇoalito*:** *jalo* *udago,*
water:INS mud:INS unsmeared:NOM water:NOM water:NOM
pako *kadamo;* *yasa* *so* *tatra* *jado*
mud:NOM mud:NOM like 3SG:NOM there born:NOM
vudhva *teṇa* *ca* *ṇa* *lipadi*
grow:GER 3SG:INS and NEG smear:PRS.PASS.3SG
'**Unsmeared by water** (*jala* -) **and mud** (*paka* -): *jala* - is water, *paka* - is mud. As it, born and having grown there, still is not smeared by it'. (Nird, 445.183-4; transl. p. 306)*

* As in Baums' edition and translation, the text portions in bold represent the root verse quotations. The punctuation is mine and it is given to elucidate the syntax of the commentary; it does not reflect the original punctuation of the Gāndhārī manuscript provided in Baums' edition.

Moreover, the Khotan *Dharmapada* also presents an instance of pattern [i], see (24):

(24) *yo* *du* ***puñe*** *ca* ***pave*** *ca* ***duhayasa*** *na* ***lipadi***
REL:NOM but virtue:LOC and sin:LOC and in_both NEG stick:PRS.3SG
'One who does not **stick** neither **to virtue** nor **to sin**'. (Dhp G^K 183)

In § 4, it was shown how the two patterns are both attested in Buddhist Sanskrit, with pattern [i] occurring especially in slightly Sanskritized texts such as the Patna *Dharmapada* and the *Bhikṣuṇī-Vinaya*. I do not argue that the use of the locative instead of the instrumental is only related to the degree of Sanskritization of the *sūtras*; as a matter of fact, the Sanskrit *Udānavarga*, which shows a systematic use of the instrumental, is generally regarded as a Hybrid text proper too (von Hinüber 1989, 346-7). In light of this, there are probably also other causes of non-linguistic nature underlying this distribution which one has to consider, such as different lines of textual transmission reflecting different sectarian affiliations. Nonetheless, if the generalized use of pattern [ii] in place of pattern [i] is a feature of standard Sanskrit, it truly seems that in this regard the Sanskrit *Udānavarga* is more Sanskritized than the other *Dharmapada-Udānavarga* texts. In this respect, examples (22) from Pāli and (24) from Gāndhārī both present a similar use of pattern [i] and can indeed be considered parallel passages. Indic parallels also exist in the *Suttanipāta* and in the *Udānavarga*,[22] see (25a) and (25b).

22 A close parallel occurs also in the *Mahāvastu*. The oldest palm-leaf manuscript (MS Sa) and paper manuscript (MS Na, cf. Marciniak 2016, 2017) both read *sarve puṇyo ca pāpā pi kā ubhayatra na lipyase*, which Marciniak (2019, 518) emends to *sarve puṇye ca pāpe pi ca ubhayatra na lipyase*. Marciniak (fn. 21) takes *sarve puṇye ca pāpe* as intr. pl. *-e* (< *ai* < *aiḥ*, cf. von Hinüber 2001, § 316) with the sense of 'You are not stained

In this case too, the Sanskrit *Udānavarga* presents a distinctive use of pattern [ii], whereas the *Suttanipāta* shows the use of pattern [i].

(25)

a. *evaṃ* ***puññe*** *ca* ***pāpe*** *ca* ***ubhaye*** *tvaṃ* *na* *lippasi*
thus virtue:LOC and sin:LOC and both 2SG:NOM NEG stick:PRS.2SG
'So you do not **cling to merit** and **evil, both**'. (Snp 547, transl. by Norman 2001, 69)

b. *yas* *tu* ***puṇyais*** *tathā* ***pāpair*** ***ubhayena*** *na* ***lipyate***
REL:NOM but virtue:INS.PL so sin:INS.PL both:INS NEG smear:PRS.PASS.3SG
'One who is not **smeared** neither **by virtues** nor **sins**'. (Ud 33.28)

Another interesting instance of pattern [i] in Hybrid Sanskrit comes from the verses of the *Lalitavistara*, one of the *sūtras* placed by Edgerton (1953, xxv) in 'Group 2' of Buddhist Hybrid texts, see (26). In this case, the verb *lipyate* in the second *pada* parallels the class IV present *rajyate* (Pāli *rajjati*) found in the first *pada*, which possesses a similar meaning to *lipyate*, i.e. 'to be dyed, to be stained' and 'to be attached', as well as the use of the same two patterns with the same instrumental and locative. As it happens, however, in this instance *rajyate* occurs with pattern [ii], whereas *lipyate* with pattern [i], even though they are fundamentally used as synonyms.

(26) *na* ***rajyate*** *puruṣavarasya* *mānasaṃ*
NEG taint:PRS.PASS.3SG man.best:GEN mind:NOM
nabho *yathā* ***tamarajadhūmaketubhiḥ***,
sky:NOM like darkness.dust.vapour.meteor:INS.PL
na ***lipyate*** ***viṣayasukheṣu*** *nirmalo*
NEG be_attached:PRS.3SG sense.pleasure:LOC.PL pure:NOM
jale *yathā* *navanalinaṃ* *samudbhūtaṃ*
water:LOC like fresh.lotus:NOM rising_up:NOM
'The mind of the best among men **is not tainted**, like the sky [is not tainted] **by darkness, dust, vapour and meteors**; a pure one **is not attached to sensual pleasures**, like a fresh lotus rising up in the water/like a fresh lotus rising up [is untouched] by water'. (Lal 15.52, 92)

by merit or evil'. I am not entirely convinced by such emendation and its grammatical interpretation, but I cannot provide here a different proposal. Nonetheless, the difficult reading found in the *Mahāvastu* betrays the problems that scribes encountered in interpreting the syntax of *lipyate* already in ancient times.

To conclude this section, we also need to note that the active counterpart of pattern [ii] is attested in Buddhist Sanskrit as well, see for instance example (27) quoted from the *Ratnamālāvadāna*.

(27) *pādayor ubhayos tena caṃdanena lilepa sā*
foot:ACC.DU both:ACC.DU 3SG:INS sandal:INS anoint:PRF.3SG 3SG:F.NOM
'She anointed [his] feet with that sandal oil'. (Ratna 22.32)

5.2 Diachronic Development: From Locative to Instrumental

The alternation between pattern [i] and pattern [ii] in MIA sources as illustrated above raises a number of questions, especially with regard to the examples discussed in Section 4, since the *Dharmapada-Udānavarga* texts ultimately represent different sectarian rearrangements of a common group of inherited verses. Thus, we can postulate that the two readings with the locative and instrumental complements are ultimately derived from a common 'urkanonish' formula which was transposed into the instrumental or locative forms during the process of transposition from the unidentified midland MIA dialect of the earliest predication into the various Buddhist Prakrits and subsequently into Buddhist Sanskrit. So, what was the pattern used in the original urkanonish source expression and how can we explain the alternation between the two patterns attested in the extant sources?

In MIA, the ending *-ehi* is used as a generalized oblique ending (von Hinüber 2001, § 321; Oberlies 2019, § 36; Pischel 1900, § 371). As a consequence, the substitution of a historical instrumental for a locative, especially in the plural, is a common feature of Early MIA: instrumentals used as locatives are found already in the language of the *Upaniṣads* (Salomon 1991, 58) and are well-attested in Buddhist Hybrid Sanskrit (Edgerton 1953, 44), Pāli (Lüders 1954, §§ 220-5) and Gāndhārī (Lenz 2003, 56). On the other hand, in Prakrit locatives are also used as instrumentals (Oberlies 2019, 224); as put by Oberlies (2019, 225), we can talk of a certain "interchangeability on the part of the instrumental and locative plural" in MIA. The alternation between the instrumental and the locative observed in our sources, hence, is not surprising.

In this regard, Watanabe (2010) aptly observes that the simile of the lotus untouched by water and mud, as one is not touched by sensual pleasures, found in the *Dharmapada-Udānavarga* texts has parallels in Jain sources as well and can be traced back to a common archetype already present in the *Upaniṣads*. To substantiate this claim, Watanabe quotes a passage from the *Chāndogyopaniṣad*, quoted in (28), which is of particular interest for the present investigation. In this case, instead of *lipyate* found in Buddhist sources, one finds the

class IV present *śliṣyate* taking a locative complement as in pattern [i] of *lipyate*. The semantics of the two verbs is virtually the same and indeed the second part of this passage has the same meaning as found in the passage from the *Vājasaneyisaṃhitā* quoted in (19b).

(28) *yathā* ***puṣkarapalāśa*** *āpo* *na* ***śliṣyanta***
like lotus.leaf:LOC water:NOM.PL NEG stick:PRS.3PL
evam ***evaṃvidi*** *pāpaṃ* *karma* *na* ***śliṣyata*** *iti*
thus thus.knowing:LOC evil:NOM action:NOM NEG stick:PRS.3SG QUOT
'Like water does not **stick to the lotus leaf**, in the same way an evil action does not **stick to one who knows thus**'. (ChUp 4.14.3)

The example quoted above is revealing and can help us draw more solid conclusions as concerns the questions presented at the beginning of this section. First, if the verses occurring in the *Dharmapada-Udānavarga* texts echo the archetype attested in the *Upaniṣads*, it is legitimate to assume that the instrumentals occurring with *lipyate* were originally used as locatives. Second, as shown by the use of *śliṣyate* in the same exact context, the present *lipyate* with a locative complement as in pattern [i] represents a class IV present as *śliṣyate* (or *rajyate*). The status of the verb was presumably still clear in Late OIA, but with the generalization of the oblique suffix *-ehi*, instances of pattern [i] with locative plurals were progressively reanalysed as cases of pattern [ii]. It is possible that this triggered the generalization of pattern [ii] also with singular complements (see 18d), which eventually led to the virtual disappearance of pattern [i] in Sanskrit and possibly also to the reanalysis of *lipyate* into an actual present passive. Pāli and Hybrid sources appears to have preserved (or possibly even restored) pattern [i], while the majority of Sanskrit sources continue the instrumental plural reading. Considering that in Classical Sanskrit *lipyate* is typically found with the instrumental, we can imagine that the 'passive' reading (i.e. pattern [ii]) ultimately became the standard and that later Buddhist Sanskrit texts reflect this process of standardization.

It is possible that not even the use of the locative was sufficient to solve the semantic and grammatical ambiguity of the source expression. As a matter of fact, Pāli texts often show a certain hesitation between the readings *lippati* and *limpati* in pattern [i], see for instance the parallel of (15a) in SN I.10.8.15 proving that some of the monks who transmitted the scriptures were presumably also analysing *lippati* as a present passive and hence not compatible with an

intransitive reading.[23] In this regard, Norman (1997a, 85 ff.) conjectures that this hesitation betrays the difficulty of the scribes to decide between an active and passive interpretation in the context of written transmission of the texts. Possibly due to the use of a writing system which did not distinguish geminated consonants (cf. Norman 1993, 240-1) and without the help of an oral tradition, metrically ambiguous syllables were susceptible of a double interpretation, particularly in those cases where the context did not prove useful for the disambiguation.

6 How Was the Expression Understood by Translators?

One may wonder whether the Chinese and Tibetan translators were also aware of the semantic ambiguity of the expressions and that the instrumental *kāmehi*/*kāmaiḥ* occurring with *lipyate* could be understood as a locative. The Tibetan translation of the *Udānavarga* provides some insights into this issue. In (29) are given the respective translations of the verses quoted in (15), (16) and (18). In (29a) and (29b), the perfect *gos* (smeared) is preceded by the noun *'dod pa* (desire) marked with the ergative/instrumental suffix *-s*. On the other hand, (29c) presents a different verb, namely the present *gnas*, lit. 'to abide, to remain', preceded by the locative noun-phrase *'dod la* built with the locative postposition *la*. There is no evident reason to believe that the Sanskrit source text used for the Tibetan translation of the verse in (29c) read **na lipyate kāmeṣu* instead of the instrumental found elsewhere.[24] Therefore, one can conclude that the Tibetan translator was aware of the possible locative reading of the passage and that the context, especially the presence of the two

23 The confusion was also facilitated by the fact that the opposition of active and passive is generally based only on the stem due to the use of the active endings for the middle ones, see § 2.1.

24 As discussed by Schmithausen (1970, 59 ff.), the Sanskrit manuscripts from Central Asia used by Bernhard for his edition and the Tibetan translation represent two separate recensions of the *Udānavarga*. One can thus not completely exclude that the source text used for the Tibetan translation read a locative form of *kāma-*, although it seems quite unlikely. The only exception in following the locative reading among the Sanskrit sources used by Bernhard appears to be a Sanskrit fragment manuscript from the Ming Öy caves in Kizil (DUc in Bernhard's notation), which in the portions corresponding to 33.30, 33.31A and 33.28 respectively reads (*lipyate yo na*) *kāme*[*ṣu*], *lipyate y*(*o*) [*n*](*a*) *kāme*(*ṣu*) and *ca nobhāyatra* (471-3). As for the rest, the three Tibetan verses quoted here virtually agree almost verbatim with the Sanskrit recension of the *Udānavarga*. Besides the locative complement in (29c), the only other difference between the Tibetan and Sanskrit versions is the verb *'jug* (he behaves) instead of Sanskrit *śete* (he rests).

locatives *puṣkarapatre* and *ārāgre* in the first half of the verse, were of help to disambiguate the meaning of the verb.[25]

(29)

a. *bram ze* *mya ngan 'das pa* *dag,* *rnam pa kun tu* *bde bar* *'jug,*
brāhmaṇa calmed and always at_ease act
gang zhig ***'dod pa-s*** ***ma*** ***gos*** *shing, zag med* *rab tu rnam grol ba*
whoever desire-INS NEG smeared and immaculate completely_liberated
'A *brāhmaṇa* is calmed and in every circumstance behaves at ease, one who is **not smeared by desire**, immaculate and completely liberated'. (UdT 30.30)

b. *zla ba* *dag* *cing* *dri med* *la,*
moon pure and bright and
skyon bral *rab tu dang ba* *ltar,*
clear completely_clean like
gang zhig ***'dod pa-s*** *mi* *gos* *de,*
whoever desire-INS NEG smeared that
bram ze *yin par* *nga-s* *gsungs* *so*
brāhmaṇa be I-ERG said FIN
'Like the moon is pure, bright, clear and completely clean, one who is **not smeared by desire**, him I call a *brāhmaṇa*'. (UdT 33.38)

c. *padma* *'i* *'dab* *la* *chu* *ltar* *dang,*
lotus GEN leaf on water like and
smyung bu *'i* *rtse* *la* *yungs kar* *ltar,*
awl GEN point on mustard like
gang zhig ***'dod*** *la* *mi* *gnas* *de,*
whoever desire in NEG abide that
bram ze *yin par* *nga-s* *gsungs* *so*
brāhmaṇa be I-ERG said FIN
'Like water [does not cling on] a lotus leaf, or mustard to the point of an awl, one who **does not abide in desire**, him I call a *brāhmaṇa*'. (UdT 33.35)

The same issue also applies to the Chinese case: is it possible that the Chinese translators were aware of the semantic ambiguity behind the expression and of the possible locative reading of *kāmaiḥ*/*kāmehi*? The alternation between *bù rǎn yú yù* 不染於欲 and the *wéi* construction suggests that the locative interpretation of the passage was known by the Chinese translators as well. The Classical literary

25 The Tibetan parallel of (25b, Ud 33.28), i.e. UdT 33.31, seems to follow the instrumental reading of the Sanskrit passage: *gang zhig dge dang sdig pa dang, gnyis ka yis kyang mi gos pa* (One who is not stained neither by virtues nor sin).

expression *rǎn yú* 染於 was probably a good solution to the eyes of the translators to render the ambiguous Indic expression, as both meaning of 'to smear' ('to stain' < 'to dye'), and the oblique marking the locative complement (preposition *yú* 於) were simultaneously conveyed. As also seen in example (12), the relationship between the expression *rǎn yú* 染於 and a locative complement in the Indic source text, as well as an association with the meaning 'to cling on, to stick to', seems to be well-attested in the Chinese translated literature. Limiting the scope to Kumārajīva's translation corpus, the passive *wéi* constructions is regularly employed by the Kuchean translator, see for instance the use of *wéi* in (6b); thus, it is extremely dubious that *bù rǎn yú yù* 不染於欲 could simply represent a 'stylistic' variation of the passive construction.

In this regard, it is interesting to see that in the commentary part of example (17a), Zhú Fóniàn mentions the act of 'clinging' (Ch. *zhuó* 著) on the six external sensory objects, despite the fact that in the verse passage he employs the *wéi* construction.

(30) 猶如蓮華之葉不受塵水, 彼修行人亦復如是, 以離於欲, 不復著色聲香味細滑法.

yóurú	*liánhuā*	*zhī*	*yè*	*bù*	*shòu*	*chén*	*shuǐ,*
like	lotus	GEN	leaf	NEG	receive	dust	water
bǐ	*xiūxíngrén*	*yǐfù*	*rúshì,*	*yǐ*	*lí*	*yú*	*yù,*
that	practitioner	also	thus	to	depart	from	desire

bù	*fù*	***zhuó***	***sè***	***shēng***	***xiāng***	***wèi***	***xìhuá***	***fǎ***
NEG	also	cling	form	sound	smell	taste	touch*	dharma

'Like the leaves of the lotus are not touched by dust and water, that practitioner, in order to distance himself from desire, **does not cling to form, sound, smell, taste, touch and dharmas**'. (T 212, 771c5-6)

* Lit. 'soft and smooth', generally corresponding to Sk. *sparśa* 'touch' (Karashima 1998, 483; Vetter 2012, 177).

Moreover, Zhú Fóniàn employs the expression *rǎn yú yù* 染於欲 also to translate another verse from the *Paśyavarga* chapter of the *Udānavarga*. In this case, the verb in the Sanskrit parallel is not *lipyate*; the verb *rǎn* 染 seems rather to translate the Sanskrit *-ta* participle *mūḍha-* 'confused'. What is relevant to the present discussion, however, is the presence of the locative plural complement *kāmeṣu*, as well as the fact that the 'confusion' mentioned in the verse derives from 'clinging' (Sk. *saktāḥ* 'clinging' = Ch. *zhuó* 著) on desire.

(31)

a. 著欲染於欲, 不究結使緣.

zhuó	***yù***	***rǎn***	***yú***	***yù***	*bù*	*jiū*	*jiéshǐ*	*yuán*
cling	desire	be_attached	to	desire	NEG	understand	fetter	cause

'**Clinging on desire, being attached to desire**, [they] do not understand the cause of the fetter'. (T 212, 739a7)

b.	***kāmeṣu***	***saktāḥ***	*satataṃ*	*hi*	***mūḍhāḥ***,
	desire:LOC.PL	attached:NOM.PL	constantly	because	confused:NOM.PL
	saṃyojane	*vadyam*	*apaśyamānāḥ*		
	fetter:LOC	sin:ACC	not.see:PTCPL.PRS.PASS.NOM.PL		

'**Confused** because constantly **attached to sensual pleasures**, not seeing the sin in the fetter'. (Ud 27.27)

7 Conclusion

In this paper, I have argued that the evidence from MIA corroborates Kulikov's (2012) hypothesis on the status of *lipyate*: it originally served as a class IV present intransitive with an anticausative meaning. Owing to the use of the historical instrumental plural as a generalized oblique plural ending in MIA, such intransitive usage as found in *lipyate kāmehi* < **lipyate kāmeṣu* eventually became ambiguous, as the original recipient/object of attachment could be taken as the instrument of smearing and the intransitive verb reanalysed as a present passive. The locative reading of *kāma-* was generally preserved in Pāli and in some Hybrid Sanskrit texts, whereas Gāndhārī possibly reflects the process of transition towards the generalized use of the instrumental. Later texts with a higher degree of Sanskritization, such as the *Larger Prajñāpāramitā* discussed at the beginning of this paper, diverge from locative usage of Pāli and present instead the instrumental reading as well, presumably because in standard Sanskrit *lipyate* was generally used with the instrumental complement as the passive/anticausative counterpart of active *limpati*.

The Chinese and Tibetan translations reflect the semantic and grammatical ambiguity underlying the Indic source expression and even appear to show that the locative interpretation of the instrumental reading was known to the translators, in spite of the fact that that the Indic source texts used by them probably presented pattern [ii] with an instrumental plural.[26] The expression *bù rǎn yú yù* 不染於欲, borrowed from literary Chinese, can be understood as an attempt to convey the locative meaning, as well as the semantic nuance

26 One needs also to mention that it was a common practice for Chinese translators to rely on earlier popular or authoritative translations when producing a new one (Nattier 2008, 26). Thus, the use of *bù rǎn yú yù* 不染於欲 in place of the passive construction (and viceversa) could have also been influenced by reasons of stylistic choice of such sort.

of 'being stained', which derives from the literal sense of the expression 'to dye in'.

In closing, the Chinese translations of this expression also offer an insight into some methodological problems underlying the grammatical analysis of the Chinese Buddhist translations: the case of *bù rǎn yú yù* 不染於欲 shows how a precise grammatical interpretation of the linguistic material found in the Buddhist literature deeply relies on a thorough comparison of the Indic parallels.

Abbreviations

In the glosses

ACC	accusative
CONJ	conjunction
COP	copula
DU	dual
ERG	ergative
F	feminine
FIN	final particle
GEN	genitive
GER	gerund
GRND	gerundive
IMP	imperative
INDF	indefinite
INS	instrumental
IPRF	imperfect
LOC	locative
NEG	negation
NMLZ	nominalizer
NOM	nominative
PRF	perfect
PRS	present
PASS	passive
PL	plural
PTCPL	participle
QUOT	quotative particle
SG	singular
VOC	vocative
1	first person
2	second person
3	third person

In running text

A	agent
AN	*Aṅguttaranikāya* = Morris 1885-1900
Bhag	*Bhagavadgītā* = Belvalkar 1968
BhīVin	*Bhikṣuṇīvinaya* = Roth 1970
ChUp	*Chāndogyopaniṣad* = Olivelle 1998, 166-287
Ch	Chinese
Culv	*Cullavagga* = Oldenberg 1880
Dhp	Pāli *Dhammapada* = von Hinüber, Norman 1994
Dhp G^K	Khotan *Dharmapada* = Brough 1962
ĪśUp	*Īśopaniṣad* = Olivelle 1998, 405-12
Jā	Pāli *Jātaka* = Fausbøll 1877-96
KS	*Kāṭhakasaṃhitā* = von Schroeder 1900
Lal	*Lalitavistara* = Hokazono 2019
MIA	Middle Indo-Aryan
Mil	*Milindapañha* = Trenckner 1880
MN	*Majjhimanikāya* = Chalmers 1888-99
MS	*Maitrāyaṇī Saṃhitā* = von Schroeder 1885
Nird	Gāndhārī Commentary i.e. *Nirdeśa* = Baums 2009
OIA	Old Indo-Aryan
Pañca	*Pañcaviṃśatisāhasrikā* = Dutt 1934
PDhp	Patna *Dharmapada* = Cone 1989
PIE	Proto-Indo-European
R	recipient
Ratna	*Ratnamālāvadāna* = Takahata 1954
ṚV	*Ṛgveda* = van Nooten and Holland 1994
ŚB	*Śatapathabrāhmaṇa* = Weber 1855
S	subject
Saṅghabh	*Saṅghabhedavastu* = Gnoli 1978
SN	*Samyuttanikāya* = Feer 1884-98
Sk.	Sanskrit
Snp	*Suttanipāta* = Andersen, Smith 1913
Sp	*Samantapāsādikā* = Takakusu, Litt 1924
T	Taishō Canon = Takakusu, Watanabe 1924-32.
Ud	Sanskrit *Udānavarga* = Bernhard 1965
UdS	*Udānavarga* from Subaši = Nakatani 1987
UdT	Tibetan *Udānavarga* = Dietz, Zongtse 1990
Vikn	*Vimalakīrtinirdeśa* = SGBSL 2006
V	verb
VS	*Vājasaneyisaṃhitā* = Weber 1852

Bibliography

Abraham, W. (2006). "Introduction: Passivization and Typology. Form Vs. Function – A Confined Survey into the Research Status Quo". Abraham, Leisiö 2006, 1-28.

Abraham, W.; Leisiö, L. (eds) (2006). *Passivization and Typology: Form and Function*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Aldridge, E. (2013a). "Survey of Chinese Historical Syntax Part I: Pre-Archaic and Archaic Chinese". *Language and Linguistics Compass*, 7(1), 39-57. https://doi.org/10.1111/lnc3.12006.

Aldridge, E. (2013b). "Survey of Chinese Historical Syntax Part II: Middle Chinese". *Language and Linguistics Compass*, 7(1), 58-77. https://doi.org/10.1111/lnc3.12007.

Ān Jùnlì 安俊丽 (2009). "Hàn Wèi Liù Cháo Hànyì fójīng 'jiàn' zì bèidòngjù yánjiū" 漢魏六朝漢譯佛經"見"字式被動句研究 (Research on "Jian" Passive Sentences in the Sutras from Han to South-North Dynasty). *Nèiměnggǔ dàxué xuébào* 内蒙古大學學報, 41(3), 133-6.

Anderl, C. (2017). "Medieval Chinese Syntax". Sybesma, R. (ed.), *Encyclopedia of Chinese Language and Linguistics*. Leiden: Brill, 689-703.

Barchi, F. (forthcoming). "Chinese Buddhist Translations as a Locus of Grammatical Interference". *Journal of Language Contact*, 16(4).

Barchi, F.; Peschl, B. (2022). "Preterite Formations in Niya Prakrit and Khotanese: A Case of Grammatical Interference?". *Acta Orientalia Academiae Scientiarum Hungaricae*, 75(3), 405-37. https://doi.org/10.1556/062.2022.00223.

Baumgarten, N.; Özçetin, D. (2008). "Linguistic Variation Through Language Contact in Translation". Siemund, P.; Kintana, N. (eds), *Language Contact and Contact Languages*. Amsterdam: John Benjamins Publishing Company, 293-316. https://doi.org/10.1075/hsm.7.17bau.

Baums, S. (2009). *A Gāndhārī Commentary on Early Buddhist Verses: British Library Kharoṣṭhī Fragments 7, 9, 13 and 18*. Seattle: University of Washington.

Becher, V.; House, J.; Kranich, S. (2009). "Convergence and Divergence of Communicative Norms Through Language Contact in Translation". Braunmüller, K.; House, J. (eds), *Convergence and Divergence in Language Contact Situations*. Amsterdam: John Benjamins Publishing Company, 125-52. https://doi.org/10.1075/hsm.8.06bec.

Brough, J. (1962). *The Gāndhārī Dharmapada*. Oxford: Oxford University Press.

Bubeník, V. (1996). *The Structure and Development of Middle Indo-Aryan Dialects*. Delhi: Motilal Banarsidass Publishers.

Burrow, T. (1937). *The Language of the Kharoṣṭhi Documents from Chinese Turkestan*. Cambridge: Cambridge University Press.

Chao, Y.R. [1964] (2011). *A Grammar of Spoken Chinese*. Beijing: The Commercial Press.

Comrie, B. (1988). "Passive and Voice". Shibatani, M. (ed.), *Passive and Voice*. Amsterdam: John Benjamins Publishing Company, 9-23. https://doi.org/10.1075/tsl.16.04com.

Dahl, E. (2016). "The Origin and Development of the Old Indo-Aryan Predicated *-tá* Construction". Dahl, E.; Stroński, K. (eds), *Indo-Aryan Ergativity in Typological and Diachronic Perspective*. Amsterdam: John Benjamins Publishing Company, 61-108. https://doi.org/10.1075/tsl.112.03dah.

Delbrück, B. (1888). *Altindische Syntax*. Halle: Verlag der Buchhandlung des Waisenhauses.

Demoto Mitsuyo 出本充代 (1995). "Senshū hyaku innen kyō no yakushutsu nendai ni tsuite" 撰集百因縁経の訳出年代について (On the Date of Translation of the *Zhuànjí bǎi yuán jīng*). *Pārigaku Bukkyōgaku bunkagaku*, 8, 99-108.

Dimitrov, D. (2020). *The Buddhist Indus Script and Scriptures: On the So-Called Bhaikṣukī or Saindhavī Script of the Sāṃmitīyas and their Canon*. Wiesbaden: Harrassowitz.

Edgerton, F. (1953). *Buddhist Hybrid Sanskrit Grammar and Dictionary. Grammar*. New Haven: Yale University Press.

Falk, H. (2015). "A New Gāndhārī *Dharmapada* (Texts from the Split Collection 3)". *Annual Report of The International Research Institute for Advanced Buddhology at Soka University*, 18, 23-62.

Fortson, B.W. (2010). *Indo-European Language and Culture: An Introduction*. 2nd ed. Chichester: Wiley-Blackwell.

Geiger, W. (1916). *Pali: Literatur und Sprache*. Strassburg: Karl J. Trübner Verlag.

Gotō, T. (2013). *Old Indo-Aryan Morphology and Its Indo-Iranian Background*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Hashimoto, M.J. (1988). "The Structure and Typology of the Chinese Passive Construction". Shibatani, M. (ed.), *Passive and Voice*. Amsterdam: John Benjamins Publishing Company, 329-54. https://doi.org/10.1075/tsl.16.11has.

Haspelmath, M. (1990). "The Grammaticization of Passive Morphology". *Studies in Language*, 14(1), 25-72. https://doi.org/10.1075/sl.14.1.03has.

Heirman, A. (2004). "The Chinese 'Samantapāsādikā' and its School Affiliation". *Zeitschrift der Deutschen Morgenländischen Gesellschaft*, 154(2), 371-96.

von Hinüber, O. (1989). "Origin and Varieties of Buddhist Sanskrit". Caillat, C. (éd.), *Dialectes dans les littératures indo-aryennes*. Paris: Boccard, 341-67.

von Hinüber, O. (1996). *A Handbook of Pāli Literature*. Berlin; New York: De Gruyter.

von Hinüber, O. (2001). *Das ältere Mittelindisch im Überblick*. 2. erweiterte Auflage. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

von Hinüber, O. (2022). *Studien zur Kasussyntax des Pāli, besonders des Vinaya-Piṭaka*: *Durchgesehener und korrigierter Nachdruck im Neusatz, ergänzt um einen Sachindex von Petra Kieffer-Pülz*. Halle an der Saale: Universitätsverlag Halle-Wittenberg.

Hock, H.H. (1986). "'P-Oriented' Constructions in Sanskrit". Krishnamurti, B. (ed.), *South Asian Languages: Structure, Convergence and Diglossia*. Delhi: Motilal Banarsidass, 15-27.

Hock, H.H. (2022). "Passives and Anticausatives in Vedic Sanskrit: Synchronic and Diachronic Perspectives". Dahl, E. (ed.), *Alignment and Alignment Change in the Indo-European Family*. Oxford: Oxford University Press, 166-87. https://doi.org/10.1093/oso/9780198857907.003.0005.

Huáng Bǎoshēng 黃寶生 (2011). *Fàn Hàn duìkān Wéimójié suǒshuō jīng* 梵漢對勘維摩詰所説經 (Vimalakīrtinirdeśa: A Sanskrit-Chinese Comparative Analysis]). Beijng: Chinese Social Sciences Press.

Jamison, S.W. (1984). "The Vedic Passive Optative and Its Functional Equivalents: A Study in the Syntax of the Gerundive". *Journal of the American Oriental Society*, 104(4), 609-20. https://doi.org/10.2307/601895.

Jamison, S.W. (1990). "The Tense of the Predicated Past Participle in Vedic and Beyond". *Indo-Iranian Journal*, 33(1), 1-19. https://doi.org/10.1163/000000090790083248.

Jamison, S.W. (2000). "Lurching towards Ergativity: Expressions of Agency in the Niya Documents". *Bulletin of the School of Oriental and African Studies*, 63(1), 64-80. https://doi.org/10.1017/S0041977X00006455.

Karashima, S. (1998). *A Glossary of Dharmarakṣa's Translation of the Lotus Sutra* 正法華經詞典. Tokyo: The International Research Institute for Advanced Buddhology, Soka University.

Karashima, S. (2016). "Indian Folk Etymologies and Their Reflections in Chinese Translations – *brāhmaṇa*, *śramaṇa* and *Vaiśramaṇa*". *Annual Report of The International Research Institute for Advanced Buddhology at Soka University*, 19, 101-23

Kazenin, K.I. (2001). "The Passive Voice". Haspelmath, M. et al. (eds), *Language Typology and Language Universals*. Berlin, Boston: De Gruyter, 899-916. https://doi.org/10.1515/9783110194265-004.

Kranich, S. (2009). "Epistemic Modality in English Popular Scientific Texts and Their German Translations". *trans-kom*, 2(1), 26-41.

Kranich, S. (2014). "Translations as a Locus of Language Contact". House, J. (ed.), *Translation: A Multidisciplinary Approach*. London: Palgrave Macmillan, 96-115. https://doi.org/10.1057/9781137025487_6.

Kranich, S.; Becher, V.; Höder, S. (2011). "A Tentative Typology of Translation-Induced Language Change". Kranich, S. et al. (eds), *Multilingual Discourse Production: Diachronic and Synchronic Perspectives*. Amsterdam: John Benjamins Publishing Company, 9-44. https://doi.org/10.1075/hsm.12.02kra.

Kulikov, L. (1997). "Vedic *mriyáte* and Other Pseudo-Passives: Notes on an Accent Shift". Hegedűs, I. et al. (eds), *Indo-European, Nostratic, and Beyond: Festschrift for V.V. Shevoroshkin*. Washington, D.C.: Institute for the Study of Man, 198-205.

Kulikov, L. (1998a). "Passive, Anticausative and Classification of Verbs: The Case of Vedic". Kulikov, L.; Vater, H. (eds), *Typology of Verbal Categories: Papers Presented to Vladimir Nedjalkov on the Occasion of His 70th Birthday*. Tübingen: Max Niemeyer Verlag, 139-68. https://doi.org/10.1515/9783110913750.139.

Kulikov, L. (1998b). "Vedic *-ya*-Presents: Semantics and the Place of Stress". Meid, W. (ed.), *Sprache und Kultur der Indogermanen*. Innsbruck: Institut für Sprachwissenschaft, 341-50.

Kulikov, L. (2006). "Passive and Middle in Indo-European: Reconstructing the Early Vedic Passive Paradigm". Abraham, Leisiö 2006, 62-82. https://doi.org/10.1075/tsl.68.06kul.

Kulikov, L. (2011). "Voice Typology". Song, J.J. (ed.), *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press, 368-98.

Kulikov, L. (2012). *The Vedic -ya-Presents: Passives and Intransitivity in Old Indo-Aryan*. Amsterdam: Rodopi.

Kümmel, M.J. (1996). *Stativ und Passivaorist im Indoiranischen*. Göttingen: Vandenhoeck und Ruprecht.

Kümmel, M.J. (2000). *Das Perfekt im Indoiranischen: Eine Untersuchung der Form und Funktion einer ererbten Kategorie des Verbums und ihrer Weiterentwicklung in den indoiranischen Sprachen*. Wiesbaden: Reichert.

Lamotte, É. (1980). *Le traité de la grande vertu de sagesse de Nāgārjuna*. Tome 5, *Chapîtres XLIX-LII, et Chapître XX*. Louvain: Muséon.

Lenz, T. (2003). *A New Version of the Gāndhārī Dharmapada and a Collection of Previous-Birth Stories: British Library Kharoṣṭhī Fragments 16 + 25*. Seattle: University of Washington State.

Li, C.N.; Thompson, S.A. (1981). *Mandarin Chinese: A Functional Reference Grammar.* Berkeley: University of California Press.

Lin, Q.; Radich, M. (2021). "Computer-Assisted Analysis of Zhu Fonian's Original Mahāyāna Sutras". *Buddhist Studies Review*, 38(2), 145-68. https://doi.org/10.1558/bsrv.21194.

Liú Shìzhèn 柳士鎮 (1992). *Wèi Jìn Nánběi Cháo lìshǐ yǔfǎ* 魏晉南北朝歷史語法 (Historical Grammar of Wei, Jin, North and South Dynasties Chinese]). Nanjing: Nanjing University Press.

Lüders, H. (1954). *Beobachtungen über die Sprache des buddhistischen Urkanons*. Berlin: Akademie-Verlag.

Luraghi, S.; Inglese, G.; Kölligan, D. (2021). "The Passive Voice in Ancient Indo-European Languages: Inflection, Derivation, Periphrastic Verb Forms". *Folia Linguistica*, 42(2), 339-91. https://doi.org/10.1515/flin-2021-2033.

Macdonell, A.A. (1916). *A Vedic Grammar for Students*. Oxford: Clarendon Press.

Marciniak, K. (2016). "The Oldest Palm-Leaf Manuscript of the *Mahāvastu* (MS Sa). A Paleographic Description". *Annual Report of The International Research Institute for Advanced Buddhology at Soka University*, 19, 125-48.

Marciniak, K. (2017). "The Oldest Paper Manuscript of the *Mahāvastu*". *Annual Report of The International Research Institute for Advanced Buddhology at Soka University*, 20, 103-21.

Marciniak, K. (2019). *The Mahāvastu: A New Edition*, vol. 3. Tokyo: The International Research Institute for Advanced Buddhology, Soka University.

Masica, C.P. (1993). *Indo-Aryan Languages*. Cambridge: Cambridge University Press.

Méi Guǎng 梅廣 (2018). *Shànggǔ Hànyǔ yǔfǎ gāngyào* 上古漢語語法綱要 (An Outline of Classical Chinese Grammar). Shanghai: Shanghai Education Press.

Narten, J. (1964). *Die sigmatischen Aoriste im Veda*. Wiesbaden: Harrassowitz.

Nattier, J. (2008). *A Guide to the Earliest Chinese Buddhist Translations: Texts from the Eastern Han* 東漢 *and Three Kingdoms* 三國 *Periods*. Tokyo: The International Research Institute for Advanced Buddhology, Soka University.

Nattier, J. (2010). "Re-Evaluating Zhu Fonian's *Shizhu duanjie jing* (T309): Translation or Forgery?". *Annual Report of The International Research Institute for Advanced Buddhology at Soka University*, 13, 231-58.

Nattier, J. (2023). "On Two Previously Unidentified Verses in Zhi Qian's Hybrid *Dharmapada*". *Annual Report of The International Research Institute for Advanced Buddhology at Soka University*, 26, 215-52.

Norman, K.R. (1989). "Notes on the Patna *Dharmapada*". Samtani, N.H.; Prasad, H.S. (eds), *Amalā Prajñā: Aspects of Buddhist Studies: Professor P.V. Bapat Felicitation Volume.* Delhi: Sri Satguru Publications, 431-44.

Norman, K.R. (1993). "The Development of Writing in India and its Effect upon the Pāli Canon". *Wiener Zeitschrift für die Kunde Südasiens, Supplementband*, 36, 239-49.

Norman, K.R. (1997a). *A Philological Approach to Buddhism: The Bukkyo Dendo Kyokai Lectures 1994*. London: School of Oriental and African Studies, University of London.

Oberlies, T. (2019). *Pāli Grammar*. Bristol: The Pali Text Society.
Pān Yǔnzhōng 潘允中 (1982). *Hànyǔ yǔfǎ shǐ gàiyào* 漢語語法史概要 (Essentials of Chinese Historical Grammar). Zhengzhou: Zhōngzhōu shūhuà shè.
Peyraube, A. (1989). "History of the Passive Constructions in Chinese Until the 10th Century". *Journal of Chinese Linguistics*, 17(2), 335-72.
Peyraube, A.; Wiebusch, T. (1994). "Problems Relating to the History of Different Copulas in Ancient Chinese". Chen, M.Y.; Tzeng, O.J.L (eds), *In Honor of S-Y. Wang: Interdisciplinary Studies on Language and Language Change*. Taipei: Pyramid Press, 383-404.
Pischel, R. (1900). *Grammatik der Prakrit-Sprachen*. Strassburg: Trübner.
Pulleyblank, E.G. (1986). "The Locative Particles *Yü* 于, *Yü* 於, and *Hu* 乎". *Journal of the American Oriental Society*, 106(1), 1-12. https://doi.org/10.2307/602358.
Pulleyblank, E.G. (1995). *Outline of Classical Chinese Grammar*. Vancouver: UBC Press.
Roth, G. (1980). "Notes on the Patna *Dharmapada*. Supplement to Particular Features of the Language of the Ārya-Mahāsāṃghika Lokottaravādins". Bechert, H. (ed.), *Die Sprache der ältesten buddhistischen Überlieferung*. Göttingen: Vandenhoeck und Ruprecht, 93-7.
Salomon, R. (1991). "A Linguistic Analysis of the Praśna Upaniṣad". *Wiener Zeitschrift für die Kunde Südasiens*, 35, 47-74.
Schmithausen, L. (1970). "Zu den Rezensionen des Udānavargaḥ". *Wiener Zeitschrift für die Kunde Südasiens*, 14, 47-124.
Shibatani, M. (1985). "Passives and Related Constructions: A Prototype Analysis". *Language*, 61(4), 821-48. https://doi.org/10.2307/414491.
Shibatani, M. (2004). "Voice". Booij, G. et al. (eds), *Morphologie*. New York: De Gruyter, 1145-65. https://doi.org/10.1515/9783110172782.2.14.1145.
Silk, J. (2010). "The *Jifayue sheku tuoluoni jing*: Translation, Non-Translation, Both or Neither?". *Journal of the International Association of Buddhist Studies*, 31(1-2), 369-420.
Skilling, P. (1997). "On the School-Affiliation of the 'Patna *Dhammapada*'". *Journal of the Pali Text Society*, 23, 83-122.
Szemerényi, O. (1996). *Introduction to Indo-European Linguistics*. Oxford: Oxford University Press.
Teich, E. (2003). *Cross-Linguistic Variation in System and Text: A Methodology for the Investigation of Translations and Comparable Texts*. Hawthorne: De Gruyter.
Thieme, P. (1965). "Īśopaniṣad (= Vājasaneyi-Saṃhitā 40) 1-14". *Journal of the American Oriental Society*, 85(1), 89-99. https://doi.org/10.2307/597710.
Thieme, P. (1995). "Radices postnominales". Söhnen-Thieme, R. (Hrsg.), *Kleine Schriften*, Bd. II. Stuttgart: Steiner, 1046-53.
Tournier, V. (2023). "A 4th/5th-Century sūtra of the Saṁmitīya Canon? On the So-Called 'Continental Pāli' Inscription from Devnimori (Gujarat)". Cicuzza, C. (ed.), *Proceedings of the Third International Pali Studies Week — Paris 2018*. Bangkok; Lumbini: Fragile Palm Leaves Foundation; Lumbini International Research Institute, 403-70.
Vetter, T. (2012). *A Lexicographical Study of An Shigao's and His Circle's Chinese Translations of Buddhist Texts*. Tokyo: The International Institute for Buddhist Studies of the International College for Postgraduate Buddhist Studies.

Wáng Lì 王力 [1958] (2014). *Hànyǔ yǔfǎ shǐ gǎo* 漢語語法史稿 (Historical Outline of Chinese Grammar). Beijing: Zhōnghuá shūjú.

Watanabe Kenji 渡辺研二 (2010). "Hasuba to suiteki: Genson byakueha ko seiten to bukkyō seiten no gen denshō" 蓮葉と水滴—現存白衣派古聖典と仏教聖典の原伝承— (Lotus Leaf and Water Drop: The Two Different Versions in the Extant Jaina Agamas). *Indogaku Bukkyogaku Kenkyu*, 58(2), 905-898. https://doi.org/10.4259/ibk.58.2_905.

Wèi Péiquán 魏培泉 (1994). "Gǔ Hànyǔ bèidòngshì de fāzhǎn yǔ yǎnbiàn jīzhì" 古漢語被動式的發展與演變機制 (The Development of Passive Constructions in Ancient Chinese). Lǐ Rénguǐ; Huáng Jūrén; Tāng Zhìzhēn (eds), *Zhōngguó jìngnèi yǔyán jì yǔyánxué* 中國境內語言暨語言學. Taipei: Zhōngyāng yánjiūyuàn lìshǐ yǔyán yánjiūsuǒ, 293-319.

Willemen, C. (1974). *Dharmapada: A Concordance to Udānavarga, Dhammapada, and the Chinese Dharmapada Literature*. Bruxelles: Institut belge des hautes études bouddhiques.

Xu, D. (2006). *Typological Change in Chinese Syntax*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199297566.001.0001.

Yáng Bójùn 楊伯峻; Hé Lèshì 何樂士 (2001). *Gǔ Hànyǔ yǔfǎ jí qí fāzhǎn* 古漢語語法及其發展 (Ancient Chinese Grammar and Its Development). Beijing: Language Press.

Zacchetti, S. (2005). *In Praise of the Light: A Critical Synoptic Edition with an Annotated Translation of Chapters 1-3 of Dharmarakṣa's 'Guang zan jing'* 光讚經, *Being the Earliest Chinese Translation of the Larger Prajñāpāramitā*. Tokyo: The International Research Institute for Advanced Buddhology, Soka University.

Zacchetti, S. (2007). "Inventing a New Idiom: Some Aspects of the Language of the *Yin chi ru jing* 陰持入經 T 603 Translated by An Shigao". *Annual Report of The International Research Institute for Advanced Buddhology at Soka University*, 10, 395-416.

Zacchetti, S. (2015). "Prajñāpāramitā Sutras". Silk, J. (ed.), *Brill's Encyclopedia of Buddhism*. Vol. 1, *Literature and Languages*. Leiden; Boston: Brill, 171-209

Zacchetti, S. (2021). *The Da zhidu lun* 大智度論 *(*Mahāprajñāpāramitopadeśa) and the History of the Larger Prajñāpāramitā*. Bochum: Projektverlag.

Zeng, J. (2020). *A Reevaluation of So-Called Passive Constructions in Ancient Chinese: From Pre-Qin to the Han Dynasty*. Gent: Universiteit Gent.

Zeng, J.; Anderl, C. (2019). "The Formation of the Copula Function of *Wei* 为 and the Nature of the '*Wei* 为 V' Construction". *Journal of Chinese Linguistics*, 47(1), 82-124.

Zürcher, E. (1977). "Late Han Vernacular Elements in the Earliest Buddhist Translations". *Journal of Chinese Language Teachers Association*, 12, 177-203.

Text Editions

Andersen, D.; Smith, H. (1913). *Sutta-nipāta*. Oxford: The Pali Text Society.

Belvalkar, S.K. (1968). *The Bhagavadgītā: Being Reprint of Relevant Parts of Bhīṣmaparvan from B.O.R.Institute's Edition of the Mahābhārata*. Poona: Bhandarkar Oriental Research Institute.

Bernhard, F. (1965). *Udānavarga: Einleitung, Beschreibung der Handschriften, Textausgabe, Bibliographie*. Göttingen: Vandenhoeck und Ruprecht.

Chalmers, R. (1888-99). *The Majjhima-Nikāya*. 3 vols. London: The Pali Text Society.
Cone, M. (1989). "Patna Dharmapada, Part I: Text". *Journal of the Pali Text Society*, 13, 101-217.
Dietz, S.; Zongtse, C.T. (1990). *Udānavarga: Der tibetische Text*. Göttingen: Vandenhoeck; Ruprecht.
Dutt, N. (1934). *Pañcaviṃśatisāhasrikā Prajñāpāramitā*. London: Luzac.
Fausbøll, V. (1877-96). *The Jataka together with its Commentary Being Tales of the Anterior Births of Gotama Buddha*. 6 vols. London: Trübner.
Feer, L. (1884-98). *The Samyutta-Nikāya of the Sutta-Piṭaka*. 5 vols. London: The Pali Text Society.
Gnoli, R. (1978). *The Gilgit Manuscript of the Saṅghabhedavastu: Being the 17th and Last Section of the Vinaya of the Mūlasarvāstivādin*. Roma: Istituto Italiano per il Medio ed Estremo Oriente.
von Hinüber, O.; Norman, K.R. (1994). *Dhammapada*. Oxford: The Pali Text Society.
Hokazono Kōichi 外園幸一 (2019). *Raritavisutara no kenkyū (chūkan)* ラリタヴィスタラの研究(中巻) (Study on the Lalitavistara, vol. II). Tokyo: Daitō Shuppansha.
Morris, R. (1885-1900). *The Aṅguttara Nikāya*. 5 vols. London: The Pali Text Society.
Nakatani, H. (1987). *Udānavarga de Subaši: Éditions critique du manuscrit sanskrit sur bois provenant de Subaši; Bibliothèque Nationale de Paris, Fonds Pelliot / 1 : Texte et fac-similés*. Paris: Boccard.
van Nooten, B.A.; Holland, G.B. (1994). *Rig Veda: A Metrically Restored Text with an Introduction and Notes*. Cambridge (MA): Harvard University Press.
Oldenberg, H. (1880). *The Vinaya Piṭakaṃ: One of the Principal Buddhist Holy Scriptures in the Pāli Language*. Vol. 2, *The Cullavagga*. London: Williams; Norgate.
Olivelle, P. (1998). *The Early Upaniṣads: Annotated Text and Translation*. Oxford: Oxford University Press.
Roth, G. (1970). *Bhikṣuṇī-Vinaya: Including Bhikṣuṇī-Prakīrṇaka and a Summary of the Bhikṣu-Prakīrṇaka of the Ārya-Mahāsāṃghika-Lokottaravādin*. Patna: K.P. Jayaswal Research Institute.
von Schroeder, L. (1885). *Maitrāyaṇī Saṃhitā: Die Saṃhitā der Maitrāyaṇīya-Śākhā*, Bd 3. Wiesbaden: Steiner.
von Schroeder, L. (1900). *Kāṭhakam: Die Saṃhitā der Kaṭha-çākhā*, Bd. 1. Leipzig: Brockhaus.
Study Group on Buddhist Sanskrit Literature (SGBSL) (2004). *Vimalakīrtinirdeśa: Transliterated Sanskrit Text Collated with Tibetan and Chinese Translations* 梵蔵漢対照「維摩經」. Tokyo: Taisho University Press.
Study Group on Buddhist Sanskrit Literature (SGBSL) (2006). *Vimalakīrtinirdeśa: A Sanskrit Edition Based upon the Manuscript Newly Found at the Potala Palace* 梵文維摩経―ポタラ宮所蔵写本に基づく校訂. Tokyo: Taisho University Press.
Sturgeon, D. (2011). *Zhongguo zhezueshu dianzi jihua* 中國哲學書電子化計劃 (Chinese Text Project). `http://ctext.org`.
Sukthankar, V.S; Belvalkar, S.K.; Vaidya, P.L. (1963). *The Mahābhārata: The Anuśāsanaparvan*. Part 1, *Being the 13. Book*. Poona: Bhandarkar Oriental Research Institute.
Takahata, K. (1954). *Ratnamālāvadāna*. Tokyo: Toyo Bunko.

Takakusu, J.; Litt, D. (1924). *Samantapāsādikā: Buddhaghosa's Commentary on the Vinaya Pitaka*, vol. 1. London; Boston: The Pali Text Society.
Takakusu Junjirō 高楠順次郎; Watanabe Kaigyoku 渡辺海旭 (1924-32). *Taishō shinshū daizōkyō* 大正新脩大藏經 (Taishō Tripiṭaka). Tokyo: Taishō shinshū daizōkyō kankōkai; Daizō shuppan.
Trenckner, V. (1880). *The Milindapañho: Being Dialogues between King Milinda and the Buddhist sage Nāgasena*. London: Williams and Norgate.
Weber, A. (1852). *The Vâjasaneyi-Sanhitâ in the Mâdhyandina- and the Kâṇva-Çâkhâ with the Commentary of Mahîdhara*. Berlin; London: Williams and Norgate.
Weber, A. (1855). *The White Yajurveda*. Vol. 2, *The Çatapatha-Brâhmaṇa in the Mâhyandina-Çâkhâ*. Berlin: Dümmler.

Etymological Dictionaries

Grassmann, H.G. (1873). *Wörterbuch zum Rig-veda*. Wiesbaden: Harrassowitz.
Mayrhofer, M. (1996). *Etymologisches Wörterbuch des Altindoarischen*, Bd. 2. Heidelberg: C. Winter.
Rix, H. (2001). *Lexikon der Indogermanischen Verben*. Wiesbaden: Reichert.
Werba, C.H. (1997). *Verba Indoarica. 1: Radices primariae*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Translations into Modern Languages

Gómez, L.; Harrison, P. [with members of the Mangalam Translation Group] (2022). *The Teaching of Vimalakīrti: An English Translation of the Sanskrit Text Found in the Potala Palace, Lhasa*. Berkeley: Mangalam Press.
Jamison, S.W.; Brereton. J.P. (2014). *The Rigveda: The Earliest Religious Poetry of India*. Oxford: Oxford University Press.
Norman, K.R. (1997b). *The Word of the Doctrine* (*Dhammapada*). Oxford: The Pali Text Society.
Norman, K.R. (2001). *The Group of Discourses* (*Sutta-nipāta*). 2nd ed. Oxford: The Pali Text Society.
Thurman, R.A.F. (1976). *The Holy Teaching of Vimalakīrti: A Mahāyāna Scripture*. State College: Pennsylvania State University Press.

Semestral journal
Department of Asian
and North African Studies

Università
Ca'Foscari
Venezia