Edited by

Lucrezia Carnesale
Martina Giuliani

# THE PAVIA VERBS DATABASE

## Current progress and perspectives

Materiali Linguistici
Università di Pavia

FrancoAngeli

*Materiali Linguistici. Collana a cura del Dipartimento di Studi Umanistici – Sezione di Linguistica Teorica e Applicata – dell'Università di Pavia*

*Materiali Linguistici* è una collana fondata nel 1990, che pubblica studi su vari aspetti delle lingue naturali e della loro evoluzione storica. La collana riflette la molteplicità di punti di vista teorici e metodologici che la linguistica oggi comprende, con riferimento in particolare ai seguenti temi: descrizione dei sistemi linguistici, tipologia, sociolinguistica, semantica, pragmatica e apprendimento linguistico.
*Ogni volume è sottoposto a processo di peer review.*

Materiali Linguistici. *Book Series edited by the Department of Humanities – Section of Theoretical and Applied Linguistics – University of Pavia*

Materiali Linguistici *is a peer-reviewed series founded in 1990. It publishes studies on various aspects of natural languages and their historical evolution. The series reflects the wide range of methodological and theoretical approaches of contemporary linguistics, with particular reference to the following themes: description of linguistic systems, typology, sociolinguistics, semantics, pragmatics and language learning.*



**Segreteria – Editorial Assistant**
Ilaria Fiorentini *(Università di Pavia)*
Dipartimento di Studi umanistici – Sezione di Linguistica
C.so Carlo Alberto 5, I-27100 Pavia; tel. 0382984484.
Per maggiori informazioni: *www.lettere.unipv.it/diplinguistica*

 ISBN 9788835183631




Copyright © 2025 by FrancoAngeli s.r.l., Milano, Italy. ISBN 9788835183631

Edited by
Lucrezia Carnesale, Martina Giuliani

# THE PAVIA VERBS DATABASE

## Current progress and perspectives

FrancoAngeli

 ISBN 9788835183631

Il volume è stato pubblicato con il contributo Fondo Ricerca & Giovani dell'Università di Pavia.

Isbn e-book Open Access: 9788835183631



Copyright © 2025 by FrancoAngeli s.r.l., Milano, Italy. ISBN 9788835183631

# Table of Contents

 ISBN 9788835183631

# Introduction

*Lucrezia Carnesale*[*]
*Martina Giuliani*[*]

This volume brings together selected contributions originally presented at the PaVeDa Workshop held at Pavia University, in May 2023. The workshop was organized in collaboration with a network of Italian scholars working at the development of the *Pavia-Verbs-Database* (PaVeDa, https://paveda.unipv.it), an open-source typological resource designed for cross-linguistic and diachronic investigations of verbs' valency patterns and alternations (Zanchi et al. 2022).

PaVeDa builds upon the ValPaL database (https://valpal.info/, Hartmann et al. 2013), introducing several major innovations. The ValPaL database is one of the outputs of the Leipzig Valency Classes Project, which developed a large-scale cross-linguistic comparison of verbs valency classes in modern languages (Malchukov & Comrie 2015). This research is grounded on Levin's (1993) outbreaking study of verbal classes in English and on her intuition of arriving at a semantic classification of verbs through the study of their syntactic behaviour. In line with this, valency classes are regarded as groups of verbs with similar morphosyntactic properties, that are, argument encoding and alternations. The ValPaL database stores information on valency patterns and alternations of a selection of verb meanings for 36 modern languages belonging to 23 families. ValPaL members singled out 80 core verb meanings that are considered representative of the verbal lexicon and instantiate distinctive grammatical behaviour. These verbs meanings express various types of events characterized by a different number of participants and comprise zero-place verbs such as weather verbs (RAIN), one-place verbs of activities (RUN, LAUGH), two-place verbs of change-of-state (BREAK, KILL), three-place verbs of transfer (GIVE, SEND). This list also includes experiential verbs of perception (SEE, SMELL), cognition (THINK), emotion (LIKE), and bodily sensation (BE HUNGRY). Some

[*] University of Pavia

verb meanings were occasionally added for specific languages up to the total of 162 verb meanings currently represented in the database.

For each verb meaning, the semantically most fitting basic verb is selected by experts of each language. Verb lemmas are stored in the database together with their basic valency pattern and possible valency alternations, represented as 'coding frames'. Coding frames specify how verbs' participants are encoded, giving information on relevant aspects of argument flagging, argument indexing, and word order. For each verb, a number of associated participants called 'microroles' are also identified. Valency alternations are labelled as "coded" if they are overtly marked on the verb by an affix, a clitic or an auxiliary; alternatively, they are regarded as "uncoded" if they are not marked on the verb in this way. For example, the meaning SCREAM is expressed by the Italian basic verb *urlare* which is assigned the basic coding frame: 1 > V.subj[1] (as in 1). The unique participant of the event is assigned to the microrole "screamer" and is identified by the number 1. In coding frames, the symbol > indicates word order and the square brackets signal indexing. Word orders and indexing are relevant in Italian to distinguish argument types. This verb allows for the uncoded applicative alternation (as in 2), and the coded impersonal reflexive alternation (as in 3, examples are adapted from Cennamo & Fabrizio 2013).

(1) *La* *folla* *urlava* *furiosamente*
det.F crowd(F) scream:IMPF.3SG furiously
'The crowd was shouting furiously'

(2) *Marco urlò* *il* *suo* *dolore*
Marco scream:PRF.3SG DET POSS.3SG pain
'Mark screamed with rage.' literally 'Mark screamed his rage.'

(3) *Si* *urlò* *di* *dolore*
REFL scream:PRF.3SG of pain
'One/we/you/they (INDF) screamed in pain.'

PaVeDa expands and enhances ValPaL in different ways. The major innovations concern the addition of new languages, the corpus-based approach, and the implementation of new features.

First, PaVeDa aims to increase the number of represented languages by including data from families that are currently underrepresented or entirely absent in ValPaL, particularly Afro-Asiatic, Uralic, and Turkic, as well as from Indo-European branches not attested in ValPaL, such as Indo-Iranian. For the Uralic and Turkic language families, contributors are working on Finnish, Hungarian, Turkish, and Chuvash. For the Indo-Iranian languages,

contributors are providing datasets for both ancient languages (Sanskrit) and modern languages (Persian and Hindi).

In addition, PaVeDa incorporates data from ancient languages, enabling diachronic research. Contributors have started adding ancient languages whose modern counterparts are already stored in ValPaL, particularly Old Latin (Giuliani, this volume), Old English (Tron & Giarda, this volume), Old High German, Classical Armenian. We have also included languages for which no modern descendent is stored in ValPaL, as in the case of Gothic (Zanchi & Tarsi 2021), and Ionic-Attic Ancient Greek. As for the latter, modern Greek has also been added thereby allowing for diachronic comparison. Some collaborators are currently working on additional ancient Indo-European languages—particularly Old Italian (Corbetta, this volume), Old Church Slavonic, Old Icelandic, Sanskrit (see Biagetti & Montesi, this volume), Old Irish (Roma & Zanchi 2025), Middle Welsh (Tron & Giarda, this volume), Hittite. For Sanskrit and Old Irish, modern descendants (Hindi, see Carnesale & Drocco, this volume; and Irish) are being added.

Second, a novelty of PaVeDa approach to valency is favouring corpora over native speaker intuition for data elicitation. While corpus-illustrated examples were an obvious choice for ancient languages, modern languages could also benefit from corpus-based approaches. For this reason, PaVeDa contributors aim to include corpus data for languages, both newly added and already stored in ValPaL, for the sake of minimizing the impact of constructed data based on native speaker intuition (Pinelli, this volume).

Third, PaVeDa introduces new features and options for comparative visualization. While retaining ValPaL's architecture, it implements new cross-linguistic annotation layers that support contrastive data visualization. An intermediate level of alternations, 'alternation class', has been introduced to facilitate cross-linguistic comparison. Following Malchukov (2015), language- specific alternations have been classified into four groups: argument-decreasing, argument-augmenting, argument rearranging; argument identifying. The first two classes indicate alternations affecting verbs' valency. Examples of valency decreasing alternations are argument omissions and passives. An argument-augmenting alternation is the cognate/kindred object insertion. Alternations regarded as "argument-rearranging" do not imply a change in the number of arguments but, rather, are characterized by different argument coding and different mapping of semantic and syntactic roles. Applicatives are typical examples of argument-rearranging alternations. Finally, "Argument-identifying" refers to reflexive and reciprocal alternations. Contributors are also working on grouping language-specific alternations onto a layer of 'comparative concepts' (Haspelmath 2010) to enhance the comparative possibilities offered by the database.

Finally, two options for contrastive data visualization have been implemented. The first allows the simultaneous visualization of all the alternations stored in the database for each verb meaning across different languages, while the second allows the comparison of the alternations found in two given languages for each verb meaning. PaVeDa contributors are also planning to implement a comparison option based on etymological information (annotated for ancient languages but not yet uploaded into the database) that will enable to track changes in valency patterns and alternations over time.

The collaboration, documented by this workshop, between researchers from Italian and foreign universities, including many young researchers, aims to build datasets for the inclusion of new languages in the database. The purpose of the workshop was to provide an opportunity for the project participants to share progress and discuss challenges encountered during the annotation of their datasets, in order to foster their collaboration, promoting networking initiatives, and involving young researchers.

The volume thus seeks to offer an overview of the current state of the project and inspire further typological and diachronic studies on valency. In line with the purpose of the workshop, this volume presents a selection of papers that reflects the diversity of the contributions: some of the papers describe theoretical issues and alternations in specific languages, others address methodological challenges faced during the annotation. The selection also highlights the wide array of languages discussed in the workshop, encompassing both ancient and modern languages. The first part of the volume focuses on ancient languages, including Latin, Old Italian, British Isles varieties, and Vedic Sanskrit, while the second part focuses on modern languages and includes Modern Standard Russian, Hungarian, Turkish, and Hindi.

Giuliani's paper explores the domain of ditransitive verbs in Latin. It investigates the valency patterns and alternations of seventeen ditransitive verbs expressing different types of transfer, based on Early and Classical Latin corpora. By analyzing the syntactic behavior of Latin ditransitives, Giuliani addresses methodological issues concerning the selection of basic verbs, coding frames, and valency alternations. The author emphasizes the importance of considering lexical semantics alongside frequency, morphological, and etymological criteria in verb lemma selection. In line with previous findings in the literature, the study shows that Latin ditransitives are characterized by lexical splits and construction alternations, with indirective, secundative, and neutral alignment types all attested. The author argues that the frequency of valency patterns should be taken into account together with the verb's semantics and maximal argument structure to identify basic coding frames. The paper also discusses examples of valency-decreasing (r-passive and Recipient omission) and valency-rearranging alternations (secunda-

tive vs. indirective patterns) in Latin ditransitives. From a theoretical perspective, the description of valency patterns and alternations allows the author to critically discuss and rethink the concepts of 'basic coding frame' and 'valency alternation'.

Corbetta investigates object omission in Old Italian focusing on *Inferno*, the first *Cantica* of Dante Alighieri's *Divine Comedy*. The analysis is conducted semi-automatically on the treebank of Dante's comedy, which enables the automatic extraction of instances of object omission. The main goal of this study is to assess whether the classification of object omission types proposed for Modern Italian is applicable to Old Italian. The author examines the referential properties of omitted objects, together with additional factors potentially influencing object omission, such as the syntactic context — namely, whether the omitted object appears in main or embedded clauses. As in Modern Italian, object omission in Old Italian involves both definite and indefinite (generic) arguments. Indefinite omitted objects occur in both main and embedded clauses, whereas definite objects are more frequently omitted in embedded clauses, with only a few exceptions.

Tron and Giarda investigate the similarities in the valency profile of Insular Celtic and English, which have never been verified for the earlier stages of these languages. The authors focus on lability in Old and Middle English and Middle Welsh and discuss the valency profile of these languages in light of the contact hypothesis. The paper focus on two main issues. First, it assesses whether the prominence of lability especially in Welsh and Irish, more so than in English, can be traced back to the earlier stages of the languages. Second, it compares the behavior of (Middle) Welsh and (Old and Middle) English vis-à-vis (Old) Irish, to assess any correspondence in their valency profile. The study aims to demonstrate the potential of PaVeDa for studying basic valency orientation and lability in the languages of the British Isles.

Biagetti and Montesi's paper examines valency patterns and alternations of experiential verbs in Vedic Sanskrit. The authors analyse 16 experiential verbs, belonging to the subdomains of Bodily sensations, Perceptions, Cognitions and Emotions. The paper also discusses some methodological issues emerging from the annotation of Vedic Sanskrit valency alternations within PaVeDa and the solution that the authors decided to apply. In particular, Biagetti and Montesi focus on cases where the basic meanings are expressed by alternations of the same verb, rather than by different lexemes, and cases where a meaning is expressed by the combination of a verb and preverbs. The authors argue that in Vedic experiential situations are mostly encoded by two-place predicates where the Experiencer is a subject, and the Stimulus is marked with various cases. According to the authors, construction variation is influenced by semantic-pragmatic factors and that while sharing

patterns with other ancient Indo-European languages, Vedic Sanskrit also displays unique patterns, such as the three causative alternations.

Pinelli's paper discusses the revision process of the existing ValPaL Russian dataset that has to be integrated into PaVeDa, focusing in particular on some inconsistences found at different levels of the dataset: verb meaning, verb form, basic coding frame, and alternations. The aim of the revision is to create a more comprehensive and consistent dataset for Modern Standard Russian. At each level, the author illustrates the modifications made to enhance the representation of Russian valency patterns from an intra-linguistic perspective, while also aligning with PaVeDa's objective of enabling cross-linguistic comparison. Pinelli discusses cases where she chooses to add information, and cases where she removed or modified existing data.

Emeksiz, Özsoy, and Seghedoni address issues encountered while annotating Turkish data in PaVeDa, illustrating them with examples. They focus on two major challenges: (1) cases where one verb is morphologically derived from another and should be treated as an alternation rather than a separate entry; and (2) cases where the PaVeDa meaning is expressed by a complex verb, distinguishing compound verbs from verbs with incorporated arguments. The second part of the paper discusses valency alternations, including passives and causatives, along with other coded alternations such as impersonal, reflexive, and reciprocal. Further issues include the double passive and double causative constructions as well as arguments omission.

Rákosi examines issues related to the creation of the Hungarian database for PaVeDa, focusing on language-specific challenges that may also be relevant for other languages. Rákosi argues that valency-changing alternations in Hungarian are typically coded on the verb and that the language exhibits a rich morphology. In the first part of the paper, the author discusses the fine-grained variation in coding non-subject arguments, and he addresses the challenges involved in studying argument structures and in building the database. In the second part, Rákosi addresses argument omission in Hungarian, emphasizing the frequent licensing of implicit arguments of various types. Rákosi argues that these omissions are largely discourse-driven. He proposes to avoid providing a comprehensive survey of all such cases, while offering instead representative examples in selected instances.

Carnesale and Drocco examine the use of valency-decreasing morphology in argument-structure rearrangement strategies in Hindi. The paper surveys two such alternations, passive and anticausative, and compares their realization in simple verbs and complex predicates. After showing that these two types display different valency orientations—transitivizing and indeterminate respectively—the authors focus on constructions that, although built on valency-decreasing morphology, rearrange the argument structure rather

than simply reduce it. They argue that these strategies are especially productive with experiential complex predicates. The authors also discuss some methodological issues derived from the annotation of these alternations within PaVeDa, in particular the mismatches between Hindi orientation and the PaVeDa structure, together with the under-representation of Hindi complex predicates in the PaVeDa meaning list.

## References


Cennamo, M. 2015. *Valency patterns in Italian*. In A. Malchukov, B. Comrie (ed.), Valency classes in the world's languages, 417-81. Berlin: de Gruyter.

Cennamo, M., C. Fabrizio. 2013. Italian valency patterns. In Iren Hartmann, M. Haspelmath, B. Taylor (eds.), *Valency Patterns Leipzig.* Leipzig: Max Planck Institute.

Hartmann, I., M. Haspelmath & B. Taylor. 2013. *Valency Patterns Leipzig*. Leipzig: Max Planck Institute.

Haspelmath, M. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. Languages 86(3): 663–687.

Levin, B. 1993. *English verb classes and alternations*. Chicago: CUP.

Malchukov, A. 2015. Valency classes and alternations: parameters of variation. In A. Malchukov, B. Comrie (eds.), *Valency classes in the world's languages*, 73-130. Berlin: de Gruyter.

Malchukov, A. & B. Comrie. 2015. *Valency classes in the world's languages*. Berlin: de Gruyter.

Roma, E., C. Zanchi. 2025. Old Irish in the PaVeDa database. In D. R. Cooper, R. Martin, G. S.D. O'Toole & S. E. Puopolo (eds), *Proceedings of the Harvard Celtic Colloquium. Vol. 42*, 212-38. Cambridge: HUP.

Zanchi, C., M. Tarsi. 2021. Valency patterns and alternations in Gothic. In S. Luraghi, E. Roma (eds*.), Valency over time: Diachronic Perspectives on Valency Patterns and Valency Orientation,* 31–88. Berlin: De Gruyter.

Zanchi, C., S. Luraghi, and C. R. Combei. 2022. PaVeDa – Pavia Verbs Database: Challenges and perspectives. In *Proceedings of the 4th workshop on research in computational linguistic typology and multilingual NLP*, 99-102. Seattle, Washington: Association for Computational Linguistics.

# Annotating Latin Ditransitives in PaVeDa: Methodological Aspects

*Martina Giuliani*[*]

## 1. Introduction

The Latin language attests a large number of verbs which take three arguments. As already noted in previous literature (see the seminal work by Bennett (1914) and, more recently, Pinkster's (2015) syntax of Latin), Latin three-place verbs are syntactically diverse, given that the verbs belonging to this group show different types of argument encoding. The three arguments of these verbs typically correspond to the following syntactic roles: (i) the subject, marked by the nominative case; (ii) the direct object, marked by the accusative case and typically denoting an inanimate entity; and (iii) the indirect object, usually encoded by the dative case, which corresponds to an animate entity and bears the semantic role of Recipient or Addressee (cf. Pinkster 1990: 13, 2015: 138). Nevertheless, there are three-place verbs featuring different constructions. For example, the *flagito* 'demand fiercely, entreat' features two accusative arguments denoting the person and the thing demanded (see (2) quoted in Giusti & Iovino 2016, Napoli 2018). This pattern is known as double accusative construction (DAC) and characterizes the so-called *verba rogandi* (e.g. *posco* 'ask', *reposco* 'demand back'). With *dico* 'say' instead, the 'said thing' receives the accusative case, whereas the Addressee of the speech receives the dative case.

Latin three-place verbs have been examined by Napoli (2018) from the perspective of linguistic typology, which conceives three-place verbs denoting transfer as 'ditransitives' (see Malchukov et al. 2010). Ditransitive constructions consist "of a (ditransitive) verb, an agent argument (A), a recipient-like argument (R), and a theme argument (T)" (Malchukov et al. 2010: 1). Prototypical ditransitive constructions contain physical transfer verbs

[*] University of Pavia

such as *give* and *hand* and denote events in which "the A causes an object to pass into the possession of an animate receiver (= recipient)" (Malchukov et al. 2010: 2, see (3)).

(1) English

| *Mary* | *gave* | *John* | *a pen.* |
|---|---|---|---|
| A | | R | T |

Less prototypical ditransitives include verbs of abstract and cognitive transfer (e.g. *offer, promise, show, teach*). The ditransitive schema has been extended to verbs of communication (e.g. *say*, *tell*) through the 'conduit metaphor' (Reddy 1979; 'information transfer', in Luraghi 2016), according to which messages are conceptualized as objects being transferred and addressees as the entities receiving those messages. These verbs do not take a Recipient in the narrow sense, but rather what can be considered a Recipient-like argument. Last, the label ditransitive has been assigned to dispossession verbs (e.g. *deny*, *steal*), as they behave like ditransitives cross-linguistically. In those cases, the R-like argument is a Maleficiary.

Ditransitive constructions feature three basic alignment types depending on whether it is either the R or the T of a ditransitive clause that is coded like the Patient (P) of a monotransitive clause (Malchukov et al. 2010). The resulting basic alignment types are three:

(i) *indirective alignment*: R is treated differently from P and T;
(ii) *neutral alignment*: T, R, and P are encoded in the same way;
(iii) *secundative alignment*: T is treated differently from P and R.

The verbs previously mentioned, *flagito* 'demand fiercely, entreat' and *dico* 'say' feature neutral alignment and indirective alignment, respectively. The verb *dono* 'donate' instead, shows secundative alignment: the Recipient is marked by the accusative case, while the entity transferred is expressed with the ablative case, as an Instrument.

R and T arguments of a ditransitive verb may behave like the monotransitive P not only in regard to encoding, but also in regard to various syntactic constructions or processes, such as incorporation, nominalization, passivization, relativization. These behavioral patterns tend to feature the same alignment as the coding properties, although mismatches are also attested (Malchukov et al. 2010: 25–45). In addition to the three basic alignment types, languages often show phenomena of splits and alternations (Malchukov et al. 2010: 18-20). *Lexical split* refers to "the situation where different verbs use different patterns" (Malchukov et al. 2010: 18). Lexical splits are cross-

linguistically frequent. In Latin ditransitive verbs select different constructions, as in the case of *dico* 'say' and *flagito* 'demand fiercely, entreat'. In other words, argument encoding types are differently distributed across the lexicon (Napoli 2018).

*Construction alternation* indicates "the situation where one and the same verb occur with different patterns with roughly the same meaning" (Malchukov et al. 2010: 18). Alternating patterns are possible in the same grammatical and lexical environment and that verbs used in different patterns may imply "subtle semantic/pragmatic differences" (Malchukov et al. 2010: 21). A well-studied case of pattern alternation is the English 'dative alternation' (see Rappaport Hovav & Levin 2008, among others). English ditransitives alternate between the double object construction (see (1)) and the prepositional dative construction with *to* shown in (2).

(2) *Mary* *gave* *a pen* *to John.*
A T R

In Latin *dono* 'donate' alternates between secundative and indirective alignments (on this see Bolkestein 1985, Pinkster 1990, 2015: 152, Napoli 2018: 64-68, Zanchi et al. 2023, among others).

This paper examines the domain of ditransitives in Latin addressing methodological aspects related to the study of their argument structure. Valency patterns and alternations are discussed based on the Latin verbs expressing the ditransitive meanings included in ValPaL/PaVeDa. This set comprises more or less prototypical transfer events, including:

(i) physical transfer: GIVE, BRING, CARRY, SEND;
(ii) cognitive transfer: ASK FOR, SAY, SHOW, TALK, TEACH;
(iii) dispossession: HIDE, STEAL;
(iv) treatment/supplying with (Pinkster 2015[1]): FILL, LOAD;
(v) grooming/bodily care: DRESS.

The annotation of Latin ditransitives in PaVeDa gives us the chance to tackle problematic aspects related to the selection of basic verbs, coding frames, and valency alternations. In the remainder of the paper, I address the methodological challenges encountered in analyzing how these meanings are expressed in Latin. From a broader perspective, the discussion of methodological issues related to

1. Pinkster (2015: 146) uses the label 'treatment' or 'supplying with' of a heterogenous group of Latin three-place verbs denoting events of supplying-with (e.g. 'furnish', 'equip'), filling, mixing, competing and changing.

the annotation of Latin ditransitive verbs seeks to stimulate reflections on the study of verbal valency with different verb classes, in both other corpus languages and modern ones.

The paper is structured as follows. After this introduction, in Section 2 I discuss the methodology for the collection and selection of Latin data. Section 3 illustrates the valency patterns of ditransitives focusing on issues with the selection of basic coding frames and argument encoding alternations. Finally, Section 4 provides some concluding remarks.

## 2. Data and methods

In this section, I describe the reference corpus and provide details on how Latin equivalents for ValPaL/PaVeDa verb meanings are selected.

### *2.1 Corpus and data extraction*

In line with PaVeDa aims, I examined the valency of ditransitive verbs with a corpus-based approach. This paper focuses on valency patterns of ditransitive verbs in Early and Classical Latin[2]. For Early Latin I consider Plautus and Cato the elder's corpora. The Plautus corpus has been investigated with the Perseus Digital Library (http://www.perseus.tufts.edu/hopper/); whereas Cato the elder's texts are collected in LASLA (https://www.lasla.uliege.be/). For Classical Latin I selected the texts by Caesar, Cicero, and Sallust included in LASLA. The entire corpora of Caesar and Sallust and part of Cicero's texts are currently available in LASLA. I supplemented Cicero's corpus, which mostly comprises orations, with a selection of his letters, in particular *Epistulae ad Atticum* included in the Perseus Digital Library. The resulting corpus is differentiated in genres, and registers, as it comprises letters, technical texts (*De Agri Cultura* by Cato the elder), and dramatic pieces, along with political and historiographic texts. To assess under- or over-representation of valency patterns due to corpus selection, I checked other authors (e.g. Seneca, Livy and occasionally poets like Vergil and Horace) and I consulted lexicographic resources such as the TLL (https://tll.degruyter.com/). Verb lemmas realizing PaVeDa/ValPaL ditransitive meanings in the LASLA corpus were investigated with Hyperbase

2. Part of the data presented here have been collected in previous works: cf. Giuliani (2021) on valency patterns in the Plautus corpus (see the Old Latin section of PaVeDa available at https://paveda.unipv.it/contributions/oldl1238 - tverbs) and Giuliani & Zanchi (2024)'s preliminary study on Classical Latin.

(http://hyperbase.unice.fr/). The occurrences of the selected verb lemmas were manually analyzed.

## 2.2 *Data selection: Latin counterpart verbs of ValPaL/PaVeDa meanings*

As I could not rely upon native speaker intuition, I had to resort to other morphological, frequency, and historical criteria to rate the basicness of competing verb lemmas for any given meaning. Following previous endeavours on corpus-languages (Zanchi & Tarsi 2021, Giuliani 2021, Giuliani & Zanchi 2024), I regard as more basic verbal lemmas that are morphologically underived or that exhibit the simplest morphological structure (e.g. BRING *fero* in respect of the preverbed *adfero* 'bring to'). If a verb is underived but is scarcely attested in the reference corpus I opted for a derived verb, provided that its number of occurrences was significantly higher. For instance, the verb *pleo* 'fill' is rare in Latin, as it is quoted only by grammarians (e.g. Festo and Priscianus). Two preverbed forms are attested with the meaning FILL: *compleo* and *impleo*. In Early Latin corpus, the former occurs 8 times while the latter occurs 17 times. In Classical Latin corpus instead, *compleo* is found 62 times while *impleo* is found 7 times. Interestingly, while in Caesar, Cicero, and Sallust *compleo* is almost exclusively attested, in Seneca *impleo* is by far more frequent than *compleo* (75 occurrences vs. 13 occurrences). As a general rule, I selected the verb which is more frequent among forms morphologically similar. In this case, the frequencies of attestation of the two verbs seem to be tied to authors' individual preferences. I included multiple lemmas for the same meaning when neither morphological complexity nor frequency applies. In this case, I included both *impleo* and *compleo* for the meaning FILL. As I will also discuss later, the frequency criterion is somewhat problematic for lemma selection in large corpus languages. First, the frequency of attestation of specific forms might be due to corpus sampling. Second, the *a priori* selection of verb lemmas in the case of preverbed forms does not allow us to assess whether and how preverbs modify the meaning in context of the base verb. For these reasons, morphological simplicity is normally favored over frequency of attestation in corpora.

In addition, for some verb meanings I monitored the behavior of verb lemmas which continue in Italian or in other Romance languages as well. This was the case of *induo* and *vestio* both expressing the meaning DRESS. These two forms continue in Italian with the verbs *indossare* and *vestire*.

Finally, I also considered verbs' lexical semantics. This criterion is particularly useful in the case of polysemous verbs. Specifically, verbs that are primarily associated with a given meaning are preferred over those that

express the same meaning only secondarily or metaphorically (see Inglese 2021: 142). In this regard, Giuliani & Zanchi (2024: 267) argue that *posco* is preferred to *peto* for the meaning ASK FOR, as the latter, despite its higher frequency, frequently conveys other meanings (e.g. 'assault', 'attack').

Consider also the meaning STEAL and the two possible candidates *rapio* and *aufero*. The former is a base verb, and the latter is a preverbed form of *fero* containing the source preverb *ab* 'from'. In both the corpora of Early Latin and Classical Latin *aufero* occurs more frequently than *rapio*. *Aufero* is found 127 times in Early Latin corpus and 237 times in Classical Latin corpus; *rapio* instead, is attested just 32 times in Early Latin corpus and 65 times in Classical Latin corpus. Both the verbs express the meaning 'to steal' only secondarily (see TLL and Lewis-Short dictionary). The primary meaning of *aufero* is 'to take away, carry off, and remove'. As the event of 'taking away' is usually performed in a violent and secret way, it also extends to the meanings of stealing and robbing. Similarly, *rapio* means 'to seize, snatch and carry off something, usually by force'. Especially, when object arguments are represented by human beings *rapio* means 'to kidnap'. As *rapio* is associated to a specific subsense of 'stealing' I opted for *aufero*. The latter, in fact, despite being more morphologically complex, better fits the general meaning of stealing.

As pointed out by the discussion of lemma selection criteria, the choice of translational equivalents is not a trivial task. Some issues also arise from differences in lexicalization patterns. ValPaL/PaVeDa meanings' list includes both BRING and CARRY. These two meanings are hardly distinguishable in Latin (see also Inglese & Zanchi 2022 on Ancient Greek). Different Latin verbs realize these meanings, including *fero*, *veho* and *porto*. A passage from *Iustiniani Augusti Digesta seu Pandectae* explains that *fero* is associated to the idea of sustaining the weight of something with one's own body; whereas *porto* conveys the meaning of carrying something with a transportation means[3]. However, this difference is not clear-cut. Examples from (3) to (5) suggest that *fero*, *veho*, and *porto* are used interchangeably with the meaning CARRY.

(3) *quorum* *erant* *...* *corpora* *insueta* *ad*
REL.GEN.PL be:PRS.3PL body:NOM.PL inured:NOM.PL to
*onera* ***portanda***
burden:ACC.PL carry:GER.ACC.PL

3. I quote here the relevant passage. Dig 50.16.235: *ferri proprie dicimus, quae quis suo corpore bajulat portari ea quae quis in jumento secum ducit*, "we properly say 'to be carried' for things which a person bears with their own body, and 'to be transported' for things which a person takes along with a beast of burden".

‘whose bodies were inured to carry heavy burdens’ (Caes *Civ*. 1.78)[4]

(4) ***vehit*** *hic* *clitellas …* *senex*
carry:PRS.3SG here pack-saddle:ACC.PL old_fellow:NOM.SG
‘Here this old fellow is carrying his pack-saddle’. (Plaut. *Most*. 778)

(5) *qui* *arma* ***ferre***
REL.NOM.PL arm:ACC.PL bring:INF.PRS
*possent*
be_able:SBJV.IMPF.3PL
‘those who were able to bear arms’ (Caes. *Gall*. 1.29.1)

As example (3) reveals, the event of carrying denoted by *porto* does not necessarily imply a transportation means. The verb *fero* is highly polysemous and occurs with different argument encodings. When it conveys the meaning CARRY it usually takes a two-argument pattern, as in (5). Elsewhere *fero* is used as a three-place verb with the meaning BRING (see examples in Section 3.1). However, *porto* expresses the meaning BRING, too (cf. Sall. *Cat*. 6). Considering both the corpora of Early and Classical Latin, *fero* occurs 1321 times, *veho* occurs 33 times and *porto* occurs 44 times. I assigned *fero* to the meaning BRING and *veho* to the meaning CARRY, but I also monitored the behavior of *porto*. Interestingly the Italian verb for BRING *portare* is a cognate of the Latin form *porto*. Latin counterpart verbs of PaVeDa/ValPaL ditransitive meanings are shown in Table 1.

| **PaVeDa/ValPaL meaning** | **Latin verb** |
|---|---|
| ASK FOR | *posco* |
| BRING | *fero* |
| CARRY | *veho* |
| DRESS | *induo*<br>*vestio* |
| GIVE | *do* |
| FILL | *compleo*<br>*impleo* |
| HIDE | *celo* |
| LOAD | *onero* |
| SAY | *dico* |

4. Examples are glossed according to the Leipzig Glossing Rules retrievable at the link: https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf. Abbreviations for authors and works follow the *Thesaurus Linguae Latinae* (*TLL*; https://thesaurus.badw.de/en/tll-digital/index/a.html).

| | |
|---|---|
| SEND | *mitto* |
| SHOW | *monstro* |
| STEAL | *aufero* |
| TALK | *loquor* |
| TEACH | *doceo* |
| TELL | *narro* |

Table 1: *Latin counterpart verbs of PaVeDa/ValPaL ditransitive meanings.*

## 3. Valency patterns and alternations: methodological aspects

In this section I show the basic valency patterns and alternations of Latin ditransitive verbs. The description of the different patterns instantiated by these verbs raise interesting methodological questions that give us the chance to discuss the concepts of 'basic coding frame' and 'valency alternation'.

### *3.1 Basic coding frame selection*

Latin ditransitives display the three basic alignments individuated by Malchukov et al. (2010). Verbs of physical transfer feature indirective alignment: the T is encoded in the same way as the P of a monotransitive construction, that is, with the accusative case, and the R is expressed differently. The verb GIVE *do* prototypically occurs in a three-argument pattern whereby the A takes the nominative, the T takes the accusative, and the R takes the dative, as shown in (6).

GIVE *do*
basic coding frame: 1-nom 2-acc 3-dat V.subj[1]
microroles: 1-giver, 2-gift, 3-giving recipient
(6) *at enim* *pecuniam* *Staieno*
but in_fact money:ACC.SG Staienus:DAT.SG
***dedit*** *Oppianicus*
give:PRF.3SG Oppianicus:NOM.SG
'But in fact, Oppianicus gave money to Staienus ...' (Cic. *Cluent*. 84.1)

In PaVeDa basic coding frames specify the number of participants ('microroles'), and their coding properties. The basic coding frame of GIVE *do* is: 1-nom 2-acc 3-dat V.subj[1]. Number 1 is assigned to the microrole "giver", number 2 to the microrole "gift", and number 3 to the microrole "giving recipient". The verbs of cognitive transfer SHOW *monstro*, SAY

*dico*, and TELL *narro* display indirective alignment, too. Addressees are encoded with the dative case as prototypical recipients. Finally, the dispossession verb *aufero* instantiates indirective alignment. In this case, the third argument embodies the semantic role of Maleficiary, as shown in (7).

STEAL *aufero*
basic coding frame: 1-nom 2-acc 3-dat V.subj[1]
microroles: 1-stealer, 2-stolen thing, 3-stealing source
(7) *pallam ...* *quam* *hodie* *uxori*
mantle:ACC.SG REL.ACC.SG today wife:DAT.SG
***abstuli***
take_away:PRF.1SG 'the mantle that today I took from (my) wife' (Plaut. *Men*. 601a)

The verbs ASK FOR *posco*, HIDE *celo*, and TEACH *doceo* display neutral alignment and feature the DAC (see (8)).

ASK FOR *posco*
basic coding frame: 1-nom 2-acc 3-acc V.subj[1]
microroles: 1-asker, 2-requested thing, 3-askee
(8) *quantum* *lubet* *me*
as_much:ACC.SG like:PRS.3SG 1SG.ACC
***poscitote*** *aurum*
ask_for:IMPV.FUT.2PL gold:ACC.SG
'Ask of me as much gold as you please' (Plaut. *Bacch*. 693)

Last, secundative alignment is shown by the verbs of treatment' or 'supplying with': FILL *impleo* and *compleo* and LOAD *onero*. These verbs encode the R-like argument in the accusative case and the T-like argument with the ablative case as an Instrument (see Napoli 2018), as shown in (9).

LOAD *onero*
basic coding frame: 1-nom 2-acc 3-abl V.subj[1]
microroles: 1-loader, 2-loading place, 3-loaded thing
(9) *stipendio* *armis …* *nauis* ***onerat***
pay:ABL.SG arm:ABL.PL ship:ACC.PL load:PRS.3SG
'He loads the ships with pay, arms…'(Sall. *Iug*. 86)

The selection of basic coding frames for Latin ditransitive verbs is not always straightforward. This is due partly to the general lack of clear criteria

for identifying basic coding frames, especially in ancient languages, and partly to the considerable constructional variability exhibited by Latin ditransitives, which leads to diverse encodings of the R and T arguments. One factor that PaVeDa contributors have considered in determining a verb's basic valency is the frequency of valency patterns observed in corpora. Usage-based approaches to valency (see Perek 2015) conceive frequency as a crucial factor for establishing basic verb valency. For instance, Perek (2015: 43) states that "the cognitive status of a valency pattern of a verb is related to the frequency of occurrence of that valency pattern with that verb in usage". For a corpus language like Latin, this means counting how many times a single verb occurs with different coding frames. A consequence of this point is that the selection of the basic pattern determines whether individual alternations are viewed as argument-adding or deleting.

However, the frequency criterion presents a dual nature. On the one hand, it offers a frequency-informed perspective on valency patterns, enabling researchers to (i) reduce the impact of idiosyncratic choices in the selection of basic coding frames, and (ii) distinguish between regular and marginal alternations on the basis of frequency. On the other hand, it is essential to acknowledge that the observed frequency of a given valency pattern in a reference corpus may reflect the composition of the text sample. For example, passive constructions are relatively rare in the Plautus corpus. This low frequency should not be taken as evidence that such alternation is marginal in Latin; rather, it reflects specific features of the comic genre. Plautus comedies, in fact, predominantly address concrete, everyday topics, and are marked by frequent dialogic exchanges among characters. Likewise, the selection of the basic valency pattern may be influenced more by the composition of the corpus than by actual usage patterns in the language.

Consider the verbs of physical transfer: GIVE *do*, BRING *fero*, and SEND *mitto.* In Latin GIVE *do* qualifies as the prototypical three-place verb, as it invariably occurs with three-argument patterns. Verbs of giving, in fact, denote transfer events conceptualized as a whole, requiring the simultaneous activation of the three participants. Furthermore, the Recipient is an integral part of the event, and the phases of giving and receiving cannot be distinguished. Thus, in the case of GIVE *do*, its lexical semantics prevents the omission of the R participant (see Kittilä 2007 for details). In contrast, verbs of caused motion allow for the omission of the R participant. However, when used as a two-place verb, SEND *mitto* does not occur with its basic transfer meaning, but rather with metaphorical meanings: e.g. 'let go, release, dismiss somebody', 'quit, stop (to do) something', and also 'sending out, putting forth, emitting something' (see TLL and Lewis-Short dictionary). The three-argument pattern of SEND *mitto* instead, profiles the Goal (Recipient/Direction) of the

event. As ValPaL guidelines[5] suggest avoiding metaphorical meanings, I regarded SEND *mitto* as a three-place verb. BRING *fero* instead, shows a more complicated situation. The latter is found with the basic transfer meaning in both two-argument and three-argument patterns. Consider as an example the frequencies of attestation of BRING *fero* with two-argument and three-argument patterns in Early Latin corpus. This verb occurs with comparable frequencies in both the patterns: out of a total of 157 occurrences, two-argument patterns are 72 (45.9%) and three-argument patterns 85 (54.1%). An example of BRING *fero* in a three-argument pattern is shown in (13).

BRING *fero*
coding frame: 1-nom 2-acc 3-dat V.subj[1]
microroles: 1-bringer, 2-brought thing, 3-bringing recipient

(10) *qui* *L.* *Antonio* ...
REL.NOM.SG L.DAT.SG Antonius.DAT.SG
***ferret*** *expensum*
bring:SBJV.IMPF.3SG payment:ACC.SG
'... Who would have brought to Lucius Antonius a payment (of thousand sesterces).' (Cic. *Phil*. 6.5)

In two-argument patterns, the Goal of the transfer is not profiled, whereas the focus is on the activity itself of 'bringing/transporting something (along)' (see (11)). A comitative adjunct expressed with *cum* 'with' and the ablative, co-referential with the A argument, may be added (cf. Plaut. *Trin*. 495).

BRING *fero*
coding frame: 1-nom 2-acc V.subj[1]

(11) *qui* *decimae* *legionis* *aquilam*
REL.NOM.SG tenth:GEN.SG legion:GEN.SG eagle:ACC.SG
***ferebat***
bring:IMPF.3SG
'The one who was bringing the eagle of the Tenth Legion.' (Caes. *Gall*. 4.25.3)

If we consider the monotransitive pattern with a nominative A argument and an accusative T argument the basic one for BRING *fero*, constructions expressing the R argument will be regarded as argument-adding types of alternations. In contrast, if we assume that the default pattern of BRING *fero*

5. ValPaL guidelines are available at https://paveda.unipv.it/static/ValencyDBQuestionnaireManual.pdf.

takes three arguments, we will analyze cases where the R argument is omitted as argument-deleting types of alternations.

Thus, the frequency with which a verb pattern is attested is not sufficient on its own to determine the basic valency pattern. Other factors, such as the verb's semantics and its maximal argument structure, must also be taken into account (see Perek 2015 for discussion). For example, the semantics of verbs of bringing imply the presence of a Goal; therefore, their maximal argument structures require three arguments. I argue that the meaning BRING is realized by *fero* in a three-argument pattern. When *fero* appears in a two-argument pattern, it more closely aligns with the meaning CARRY.

As previously introduced, basic coding frame selection is complicated by different markings of T and/or R arguments. For example, the R argument of HIDE *celo* is expressed by *clam* 'secretly' + accusative (see Plaut. *Poen* 1239), when it denotes animate entities. This strategy is employed to facilitate semantic role disambiguation (see Napoli 2018). Especially from Classical Latin onwards, the R argument of ASK FOR *posco* is encoded by the Source PP with *de* 'out of' and the ablative. Such construction can be understood in terms of analogy with the verb *peto* 'ask for, attack' usually expressing the R-like participant with *ab* 'from' and the ablative (see also Fedriani & Napoli 2023). The verb STEAL *aufero* expresses Maleficiary and/or Source arguments with the PPs *de/ex/ab* + ablative, along with the dative case. T-like arguments of FILL *compleo*/*impleo* also receive the genitive case. Finally, verbs of caused motion and cognitive transfer, SEND *mitto* can express Recipients as Directions, that is, with *ad* 'to' and the accusative (*ad* + Acc), as in (12).

SEND *mitto*
basic coding frame: 1-nom 2-acc *ad* 3-acc V.subj[1]
microroles: 1-sender, 2-sent thing, 3-sending recipient

(12) *legatos* *ad* *Crassum* *mittunt*
deputy:ACC.PL to Crassus:ACC.SG send:PRS.3PL
'They send deputies to Crassus.' (Caes. *Gall.* 3.19)

SEND *mitto* occurs more often with Recipients prepositionally constructed, unlike BRING *fero* that preferentially selects dative Recipients (see Table 2).

| | 1-nom 2-acc 3-dat V.subj[1] | 1-nom 2-acc *ad* 3-acc V.subj[1] | Total |
|---|---|---|---|
| SEND *mitto* | 67 | 254 | 321 |

| | | | |
|---|---|---|---|
| BRING *fero* | 83 | 26 | 106 |

Table 2: *Coding frames of SEND mitto and BRING fero.*

Typically, motion verbs allow for different encodings of Locative/Direction arguments. For instance, Latin intransitive motion verbs such as GO *eo* occur with Direction arguments expressed by different PPs (e.g. *ad* 'to' and *in* 'into') + accusative. In these cases, PaVeDa guidelines[6] suggest selecting the PP that is most frequent and/or semantically neuter for the basic coding frame. Alternative PPs can be included as notes to the coding frames. The same can be done with ditransitives expressing R and/or T arguments in various ways, provided that different encodings do not result in valency alternations (on this see Section 3.2). Thus, based on frequency, I assigned the coding frame 1-nom 2-acc *ad* 3-acc V.subj[1] to SEND *mitto* and the coding frame 1-nom 2-acc 3-dat V.subj[1] to BRING *fero* (on alternative encodings of Recipients see also Section 3.2).

## *3.2 Valency alternations*

Within the ValPaL project a valency alternation is defined as "a set of two different coding frames that are productively (or at least regularly) associated with both members of a set of verb pairs sharing the same verb stem" (Malchukov 2015: 91). Valency alternation concerns arguments, not adjuncts. In addition, the alternations relevant for the purposes of the project are those that are sensitive to lexical classes. For this reason, variation in coding frames (conventionally known as differential case marking) that depends on the properties of the argument itself rather than the verb (Malchukov 2015: 91) are usually discarded. In PaVeDa, valency alternations are grouped into four categories, called 'alternation classes' (see Malchukov 2015: 73-130). In line with the aims of the ValPal project, PaVeDa contributors prioritize argument encoding alternations that affect a verb's valency, that is, alternations that either increase or decrease the number of arguments. These are labelled within the database as 'Argument-augmenting' and 'Argument-decreasing' alternations.

Among argument decreasing alternations Latin ditransitives instantiate the passive alternation and the R-omission.

The passive alternation in Latin is coded, as it is marked on the verb by the characteristic affix in *-r*. Synthetic mediopassive *-r* forms characterize the *infectum* stem, while periphrastic constructions with the passive perfect

6. PaVeDa guidelines are available at https://paveda.unipv.it/static/paveda-guidelines.pdf.

participle and the verb *sum* 'be' are used for the *perfectum* stem (Pinkster 2015). In line with typological expectations, Latin ditransitives passivize either the T participant or the R participant according to the type of alignment they have (Malchukov et al. 2010). Verbs of physical and cognitive transfer featuring indirective alignment have indirective passivization. The T is invariably passiv-ized, whereas the R, if expressed, is retained in the dative (Napoli 2018). Verbs of filling and loading displaying secundative alignment passivize the R. Ditransitives with neutral alignment are believed to prefer R-passivization (Napoli 2018: 75). However, T-passivization is also admitted, as shown in example (13) containing HIDE *celo.*

*r*-passive (T)
Derived coding frame: 2-nom 3-acc (a/ab 1-abl) passV'.subj[2]

(13) *quor* *haec ...* ***celata***
why DEM.NOM.F.PL hide:PTCP.PRF.NOM.F.PL
*me* ***sunt***
1SG.ACC be:PRS.3PL
'Why ... were these things concealed from me?'(Plaut. *Pseud*. 490-491)

When the T is passivized, the R, if expressed, preserves the accusative case as in the active construction (see (13)).

R-omission occurs when R-like arguments lacking referentiality, definiteness, and specificity are suppressed, as shown in example (14) containing TEACH *doceo*.

R-omission
Derived coding frame: 1-nom 2-acc V.subj[1]
microroles: 1-teacher, 2-taught content, (3-teachee)

(14) ***docent*** *litteras*
teach:PRS.3PL literature:ACC.PL
'They teach literature.' (Plaut. *Most*.126*)*

In (14) the addressee of the teaching event is the 'general public' (more or less equivalent to *they* in English). R-omission constitutes an uncoded type of alternation, since it lacks overt morphological marking on the verb.

'Argument augmenting' alternations typically involve the insertion of a cognate or kindred argument that is etymologically or semantically related to the verb. For example, the communication verb SAY *dico* occurs with cognate T arguments. However, whereas cognate object insertion with one- and two-place verbs leads to argument augmentation, this is not the case for ditransitive

verbs like SAY *dico* where the number of arguments remains unchanged. Consider now the deponent verb of communication *loquor* realizing the meaning TALK. This verb typically occurs in a two-argument pattern whereby the A-like argument (the 'talker') is marked by the nominative and the R-like argument (the 'talked to person') can be expressed in different ways, including *cum* + ablative, as a Comitative, *ad* + Acc, and the prepositionless dative. However, TALK *loquor* can be used ditransitively, as example (15) shows.

Ditransitive pattern
Derived coding frame: 1-nom 2-dat 3-acc V.subj[1]
Microroles: 1-talker, 2-talked to person, 3-talked about content

| (15) | *tibi* | *ego* | *haec* | ***loquor*** |
|---|---|---|---|---|
| | 2SG.DAT | 1SG.NOM | DEM.ACC.PL | talk:PRS.PASS.1SG |

'I am saying these things to you.'(Plaut. *Pseud*. 227)

The ditransitive construction implies the addition of a T-like participant (the 'talked about content') in the accusative (see (15)).

The last class of alternations instantiated by Latin ditransitives are 'valency-rearranging' strategies. According to Malchukov (2015: 99), these are characterized by the "swapping of subject and object positions." Subject-object inversions and applicatives resulting in object rearrangement belong to this group. In general, variations in argument encoding that imply different mapping of semantic roles onto syntactic functions are regarded as valency-rearranging alternations. An example is the alternation between secundative and indirective alignment that is instantiated by DRESS *induo,* and, marginally, by LOAD *onero*. The latter is attested with indirective alignment only in Vergil. As for DRESS *induo*, Pinkster (2015: 264-65) claims that "with a verb of dressing the expected three-place expression is 'to dress somebody (accusative) in some sort of garment (ablative)". Secundative alignment is shown in the poetic example in (16).

Secundative alignment
coding frame: 1-nom 2-acc 3-abl V.subj[1]
microroles: 1-dresser, 2-dressee, 3-clothes

| (16) | *hoc* | *tunc* | *Argolicae* | *sanctum* | *velamine* |
|---|---|---|---|---|---|
| | DEM.ACC.SG | then | Argive:NOM.PL | sacred:ACC.SG | |
| | *velamina* | *matres* | ***induerant*** | *ebur* | |
| | veil:ABL.SG | mother:NOM.PL | dress:PPF.3PL | ivory:ACC.SG | |

'Then the Argive mothers had clothed this sacred (statue of) ivory in a veil.' (Stat. *Theb*. 10. 66)

In (16) the R-like argument (the 'dressee') is encoded with the accusative as a direct object. The expression of the third argument (the 'clothes') through the ablative "formally resembles an instrumental adjunct" (Pinkster 2015: 137). In the corpus selected DRESS *induo* is found only with indirective alignment, as shown in (17).

Indirective alignment
coding frame: 1-nom 3-acc 2-dat V.subj[1]

(17) *cui* *cum* *Deianira ...* *tunicam*
REL.DAT.SG when Deianira:NOM.SG tunic:ACC.SG
***induisset***
put_on:SBJV.PPF.3SG
'When Deianira had put upon him (sc. Hercules) the tunic...' (Cic. *Tusc*. 2.20)

In (17), the T is realized as the direct object, whereas the R is marked with the dative and functions as an indirect object. DRESS induo displays R-passivization. The T-like argument (the 'clothes') is expressed with the ablative (derived coding frame: 2-nom 3-abl passV'.subj[2]). However, as noted by Pinkster (2015: 264), "from Plautus onwards the accusative for the garment is also regularly found with the passive participle *indutus*" (see (18)).

*r*-passive (R)
Derived coding frame: 2-nom 3-acc passV'.subj[2]

(18) *Non* *ego* *te* ***indutum*** *...*
not 1SG.NOM 2SG.ACC dress:PTCP.PRF.ACC
*vidi* *pallam*
see:PRF.1SG mantle:ACC.SG
'Didn't I see you (come out) wearing a mantle?' (Plaut. *Men*. 511– 12)

In postclassical Latin DRESS *induo* occurs in two-argument pattern, too (see (19)).

Understood indirect reflexive
Derived coding frame: 1=2-nom 3-acc V.subj[1=2]

(19) *vestes ...* ***induunt***
robe:ACC.PL put_on:PRS.3PL
'They put on the robes.' (Curt. 3.13)

In (19) DRESS *induo* is used with a reflexive meaning, that is, 'to put something on oneself'. However, there is no reflexive pronoun in the dative

case referring to the R. Thus, *induo* can occur in a two-argument structure without an overt reflexive (i.e., without an indirect object in the dative), with the implication that the agent (A) dresses himself.

Determining which argument alternations should be stored as alternations is not a straightforward task. A case in point is argument encoding alternations. Typically, encoding alternations that do not involve a remapping of semantic roles and syntactic functions are not classified as alternations. For this reason, alternative encodings of Recipient arguments—such as those illustrated in Section 3.1 —are not recorded as alternations in PaVeDa. However, the alternation between Recipients expressed in the dative and those introduced by *ad* + Acc is meaningful in Latin. First, it is widespread in the domain of ditransitives, characterizing verbs of physical transfer, dispossession, and cognitive transfer (see Fedriani & Prandi 2014; Napoli 2018 for discussion). Second, it appears in specific contexts, particularly in the presence of animate Ts (see Luraghi 2010) and when Rs are only minimally affected by the transfer (see Giuliani 2024 for discussion). Third, it closely resembles the dative alternation attested across several languages represented in ValPaL (e.g. English).

## 4. Concluding remarks

This paper investigated the valency patterns and alternations of 17 ditransitive verbs expressing different types of transfer in Early and Classical Latin corpora. The resulting data are currently being integrated into PaVeDa and will contribute to broadening the diachronic coverage of Latin, which is presently limited to Old Latin (based on corpus data from Plautus).

The study has also highlighted methodological issues related to data gathering and annotation. I have examined problematic cases in the selection of verb lemmas, emphasizing the importance of considering lexical semantics alongside frequency, morphological, and etymological criteria. In line with previous findings in the literature (e.g. Napoli 2018), the analysis of the data reveals that the ditransitive domain in Latin is characterized by lexical splits and construction alternations, with indirective, secundative, and neutral alignment types all attested. In addition, I have addressed the challenges involved in identifying basic coding frames, arguing that attested frequency of valency patterns should be evaluated in conjunction with the verb's semantics and its maximal argument structure. This methodology, tailored for corpus languages, has the potential to inform and enhance the study of verb valency in modern languages as well.

Furthermore, I have illustrated examples of valency-decreasing (*r*-passive and R-omission), valency-augmenting (T argument insertion with TALK *loquor*), and valency-rearranging alternations (secundative vs. indirective patterns) with Latin ditransitives. The description of valency patterns and alternations of Latin ditransitives has inspired theoretical reflections on the concepts of 'basic coding frame' and 'valency alternation'. Providing precise and cross-linguistically applicable definitions of basic verbs, basic coding frames, and valency alternations facilitates the comparative analysis of valency-related phenomena across languages and through diachrony, in alignment with the aims of the PaVeDa project.

## References


Bolkestein, A. M. 1985. Discourse and case-marking; three-place predicates in Latin. In C. Touratier (ed.), S*yntaxe et latin, Actes du 2eme Congrès international de linguistique latine: Aix-en-Provence*, 28-31 mars 1983, 191–225. Marseilles: Université de Provence.

Fedriani, C. & M. Prandi. 2014. Exploring a diachronic (re)cycle of roles: The Dative complex from Latin to Romance. *Studies in Language* 38(3): 566–604.

Fedriani, C. & M. Napoli. 2023. The missing dative alternation in romance: explaining stability and change in the argument structure of Latin ditransitives. *Transactions of the Philological Society* Volume 121(1): 33–64.

Giuliani M. 2021. *Valency patterns in Latin.* MA thesis, Pavia University.

Giuliani, M. 2024. Alternating constructions with mitto 'send': the encoding of third arguments. In M. Batinić Angster & M. Angster (eds.), The verbal kaleidoscope Perspectives on the syntax and semantics of verbs, 172-99, Zadar: Morepress.

Giuliani, M. & C. Zanchi. 2024. Contributing to a new database on valency classes with Latin data. Challenges and perspectives. In L. Pultrová, M. Vaníková & M. Ctibor (eds.), *Exploring Latin: Structures, Functions, Meaning Volume I: Word,* 263-281, Berlin: De Gruyter.

Inglese, G. 2021. Anticausativization and basic valency orientation in Latin. In S. Luraghi, E. Roma (ed.), *Valency over Time,* 133-68. Berlin: De Gruyter.

Inglese, G & C. Zanchi. 2022. Ancient Greek valency patterns and alternations. Paper presented at the 3*rd International Colloquium on Ancient Greek Linguistics*, Universidad Autónoma de Madrid, 16-18 June 2022.

Giusti, G. & R. Iovino. 2016. Analisi comparativa del doppio accusativo latino. *Pallas*,102. *Études de linguistique latine,1* (Special issue O. Spevak): 89–97.

Kittilä, S. 2007. On the encoding of transitivity-related features on the indirect object. *Functions of Language (Ditransitivity, special issue dedicated to ditransitive constructions)* 14: 147–164.

Luraghi, S. 2010. Adverbials. In P. Baldi & P. Cuzzolin (eds.), *New Perspectives on Historical Latin Syntax*. Volume 2, 19–96. Berlin: de Gruyter.

Luraghi, S. 2016. The mapping of space onto the domain of benefaction and some unpredicted trends in semantic change. *Linguistics* 54(2): 339–379.
Malchukov, A. 2015. Valency classes and alternations: parameters of variation. In A. Malchukov & B. Comrie (eds.), *Valency classes in the world's languages*, 73-130. Berlin: de Gruyter.
Malchukov, A., M. Haspelmath & B. Comrie. 2010. *Studies in Ditransitive Constructions. A Comparative Handbook*. Berlin: de Gruyter.
Napoli, M. 2018. Ditransitive Verbs in Latin: A Typological Approach. *Journal of Latin Linguistics* 17(1): 51–91.
Perek, F. 2015. *Argument structure in usage-based construction grammar*. Amsterdam: John Benjamins.
Pinkster, H. 1990. The development of cases and adpositions in Latin. In H. Pinkster & I. Genee (eds.), *Unity in diversity*, 195-209. Dodrecht: Foris.
Pinkster, H. 2015. *The Oxford Latin Syntax* (Vol. I). Oxford: OUP.
Rappaport Hovav, M. & B. Levin. 2008. The English dative alternation: the case for verb sensitivity. *Journal of Linguistics* 44(1):129–167.
Reddy, M. 1979. The Conduit metaphor. In A. Ortony (ed.), *Metaphor and Thought*, 164-201. Cambridge: CUP.
Zanchi, C. & M. Tarsi. 2021. Valency patterns and alternations in Gothic. In S. Luraghi & E. Roma (eds.), 31–88. Berlin: De Gruyter.
Zanchi, C., S. Luraghi & M. Giuliani. 2023. NomAccAbl/NomAccDat: A corpus-based diachronic study of construction alternation with Latin ditransitives. Paper presented at *56th Annual Meeting of the Societas Linguistica Europaea* (SLE 2023), National and Kapodistrian University of Athens, 30 August 2023.

# Investigating Object Omission in Old Italian: a Case-Study from the Inferno in the Divine Comedy

*Claudia Corbetta*[*]

## 1. Object omission

When discussing object omission, I refer to cases in which the second argument of a two-place and three-place verb (or the second and/or third argument of three-place verb) is omitted. Object omission represents a type of valency alternation that is attested in several languages.

The PaVeDa database (Zanchi et al. 2022), an open-access relational database developed to investigate verb argument structure across languages[1], provides valuable insights into this alternation by listing languages in which object omission occurs[2]. Among these is Italian, specifically Standard Italian, henceforth referred to as Modern Italian to distinguish it from Old Italian (see Section 2). At the current stage, Old Italian is not included in the PaVeDa database. This study represents a preliminary step toward documenting and integrating valency alternations involving object omission in Old Italian into the database.

An example of object omission in Modern Italian, taken from the PaVeDa database[3], is provided in (1). Modern Italian has been described in this respect by Cennamo and Fabrizio (2013):

[*] University of Bergamo/Pavia

1. PaVeDa builds upon the ValPaL database (Hartmann et al., 2013; Haspelmath & Hartmann, 2015) by incorporating a broader range of languages and additional features.

2. The languages that, at the current stage, present object omissions in the database are: Classical Armenian, Eastern Armenian, Emai, German, Gothic, Italian, Modern Standard Arabic, Nen, Old High German, Old English, Old Latin, Xârâcùù, Zenzontepec Chatino. Even and Russian use the label "object deletion", whereas Nllng uses "deletion of direct object". See: https://paveda.unipv.it/alternations (accessed 27 April 2025).

3. Refer to: https://paveda.unipv.it/alternationvalues/106329 (accessed 27 April 2025).

(1) *Questo* *atteggiamento* *non* ***aiuta***.
DEM.M.SG behaviour:M.SG NEG help:PRS.3SG
'This behaviour does not help.'

In (1), the transitive verb *aiutare* 'to help' appears without an object, thus exhibiting a decreasing alternation (Cennamo 2015: 440) compared to its prototypical use, as reported in (2), where the object, *Luca*, is overtly expressed:

(2) *Anna* ***aiutò*** *Luca* *a* *studiare* *la*
Anne help:PST.3SG Luke to study:INF DET.F.SG
*matematica*.
maths:F.SG
'Anne helped Luke study maths.'

Object omission in Modern Italian has been studied by Lo Duca (2010) and further expanded by Cennamo (2015; 2017), with reference to the work of Fillmore (1986) and Croft (2001). Cennamo's classification (2017), which expands on Lo Duca (2010), is based on the lexico-aspectual properties of verbs, as well as on semantico-pragmatic and discourse-related features such as definiteness, referentiality, and context-recoverability (Cennamo 2017: 258). Following Lo Duca (2000), Cennamo identifies two main categories of object omission in Modern Italian:

(i) verbs with an indefinite null object;[4]
(ii) verbs with a definite null object[5].

Examples (3) and (4), reproduced from Cennamo (2017: 263, 267), illustrate these two types:

(3) *Marco* ***mangiò*** *e* *poi* *uscì*
Mark eat:PST.3SG and then exit:PST.3SG
'Mark ate and then went out.'

(4) *Ho* *ascoltato* *la*
AUX.PST.1SG listen:PTCP.M.SG DET.F.SG
*proposta* *e* *ho* ***rifiutato***
proposal:F.SG and AUX.PST.1SG refuse:PTCP.M.SG
'I listened to the proposal and I turned (it) down.'

4. This category also encompasses what Lo Duca (2000) refers to as 'generalized null objects', instances where the omitted object can be readily inferred from the broader syntactic context.
5. This type is referred to also as zero anaphora (Fillmore 1986).

In (3), the omitted object is indefinite, referring to an unspecified type of food. In contrast, in (4), it is definite and recoverable from the linguistic context—in this instance, *la proposta* 'the proposal'). According to the classification adopted in PaVeDa, only the first type of omission—when the object is generic and indefinite—qualifies as an instance of argument alternation.

Indefinite null objects tend to occur primarily with [–animate] referents and highly agentive verbs, though typically not with achievement or accomplishment verbs (Cennamo 2017: 271). By contrast, definite null objects display greater distributional flexibility, occurring with activity, achievement, and accomplishment verbs (Cennamo 2017: 268).

To date, no in-depth study has examined the antecedent of omitted objects in Old Italian, nor the lexico-semantic, pragmatic, and discourse-related features of the verb and the omitted argument. However, the phenomenon has been specifically investigated in relation to object clitic omission in coordination contexts, where Old Italian differs from Modern Italian (Luraghi 1998; Egerland 2003; Marello 2010: 1377-1379; Egerland & Cardinaletti 2010: 463-467) due to its allowance for omission in a broader range of contexts. More specifically, Old Italian exhibits instances of object omission that would be ungrammatical in Modern Italian, as illustrated in (5)[7]:

(5) Bono Giamboni, *Il Libro de' Vizî e delle Virtudi*, cap. 58, par. 5:

| | | | | |
|---|---|---|---|---|
| *come* | *in* | *quella* | *gente* | *vi* |
| as | in | DEM.F.SG | people | 2.PL.DAT |
| *vincemmo* | | *e* | ***cacciammo***... | |
| prevail:PST.1PL | | and | expel:PST.1PL | |

'how among those people we prevailed and drove (them) out...'

From a diachronic perspective, object omission has been analyzed by Luraghi (1998), who observed that the phenomenon becomes progressively more restricted from Latin to Old Italian and, ultimately, into Modern Italian. While Latin allowed a broader range of object omission[8], Old Italian restricted it primarily to coordination contexts (Luraghi, 1998). In Modern Italian, the pattern is even further constrained, with object omission occurring almost exclusively in coordination with de-emphatic pronouns. Luraghi's analysis underscores the relevance of syntactic position and highlights a historical shift toward the increasing overt realization of objects. This trend—from widespread omission in Latin to near obligatoriness in Modern Italian—is interpreted as evidence of a growing requirement for clitic

7. Example taken from Egerland & Cardinaletti (2010: 465, ex. 262, b).
8. For an insight into object omission in Latin, see Luraghi (1997, 1998, 2004).

expression. Notably, this development does not concern only object clitics, but clitics more broadly (Luraghi 1997: 12), and is linked to a broader process of grammaticalization (Giacalone Ramat 1990).

## 2. Old Italian, *Divine Comedy* and Italian-Old treebank

Old Italian is far from being an uncontroversial label among scholars, as the degree of continuity between Old Italian and Modern Italian has been subject of considerable debate (Ascoli 1882–1885: 124, cf. Tomasin 2019). However, more recent studies (Dardano and Frenguelli 2004), advocate for a clear distinction between the two stages, particularly from a syntactic perspective (Tesi 2004). In the present work, the label *Old Italian* is used to refer specifically to Old Florentine texts, following the definition provided in *Grammatica dell'italiano antico* (Salvi & Renzi 2010).

One of the key milestones in the linguistic, as well as literary, history of Italian is Dante Alighieri's *Divine Comedy*, an Old Italian – more precisely, Old Florentine – poem in verse, composed between 1308 and 1321. Dante's decision to adopt the Florentine vernacular rather than Latin marks a significant moment in the development of both Italian literature and language, contributing to the elevation of the vernacular to the status of a literary standard (Manni 2013) and laying the groundwork for the later codification and standardization of Modern Italian (Migliorini 1988).

The significance of this text led to the creation of annotated corpora of the *Divine Comedy* (and Dante's works in general), which facilitate the study of both the text and its language. More specifically, DanteSearch (Tavoni 2011) is a corpus containing the complete works of Dante (both in Italian and Latin), providing grammatical annotations such as lemmatization, part-of-speech tagging, and annotation at both the syntactic (clausal level) and discourse levels.

Building on DanteSearch corpus, the Italian-Old (Corbetta et al. 2023) treebank has the distinction of bringing the *Divine Comedy* and Old Italian into the larger project of Universal Dependencies (henceforth UD) (De Marneffe et al. 2021), a dependency framework encompassing more than 200 languages, including historical ones[9]. As the first Old Italian treebank in UD, Italian-Old offers detailed syntactic annotations (within a dependency framework), marking dependency relations at the word level and providing a fine-grained syntactic analysis.

9. https://universaldependencies.org.

The treebank is openly accessible[10] and encodes linguistic annotation that can be queried and extracted according to the specific needs of the user. Linguistic information is encoded with the CoNLL-U format[11], a tab-separated format where each line contains the annotation of words across 10 fields:

i) ID: numerical identifier of the word;
ii) FORM: the word form;
iii) LEMMA: the lemma of the word;
iv) UPOS: the Universal part-of-speech tag[12];
v) XPOS: a field for language-specific part-of-speech tag or morphological features, in Italian-Old this field reports the tagset of DanteSearch.
vi) FEATS: morphological features[13];
vii) HEAD: the ID number of the syntactic head of the current word;
viii) DEPREL: the dependency relation to the head[14];
ix) DEPS: a field that reports enhanced annotation[15];
x) MISC: a field for any additional annotation.

Figure 1 illustrates an annotated dependency tree of the final verse of the *Inferno*, visualized using the CoNLL-U format through ConlluEditor (Heinecke 2019):

(6) *E quindi uscimmo a riveder le stelle.*
'Thence we came forth to rebehold the stars.' (*Inf.*, XXXIV, v.139)

The syntactic tree provides a graphical representation of the linguistic information, displaying syntactic relations (e.g., cc, advmod, advcl, mark, obj, det, punct) through arrows pointing from a head node to its dependent.

Each node also includes morphological information about the part of speech (e.g., CCONJ, ADV, etc.) as well as the tagset used in DanteSearch (cscc, b, etc.). Outside the node, morphological features, if present, are represented in the form of binary values (e.g., Gender=Fem, Number=Plur, etc.).

10. Data are available at: https://github.com/UniversalDependencies/UD_Italian-Old/blob/master/README.md.
11. https://universaldependencies.org/format.html.
12. https://universaldependencies.org/u/pos/index.html
13. https://universaldependencies.org/u/feat/index.html
14. https://universaldependencies.org/u/dep/index.html
15. https://universaldependencies.org/u/overview/enhanced-syntax.html

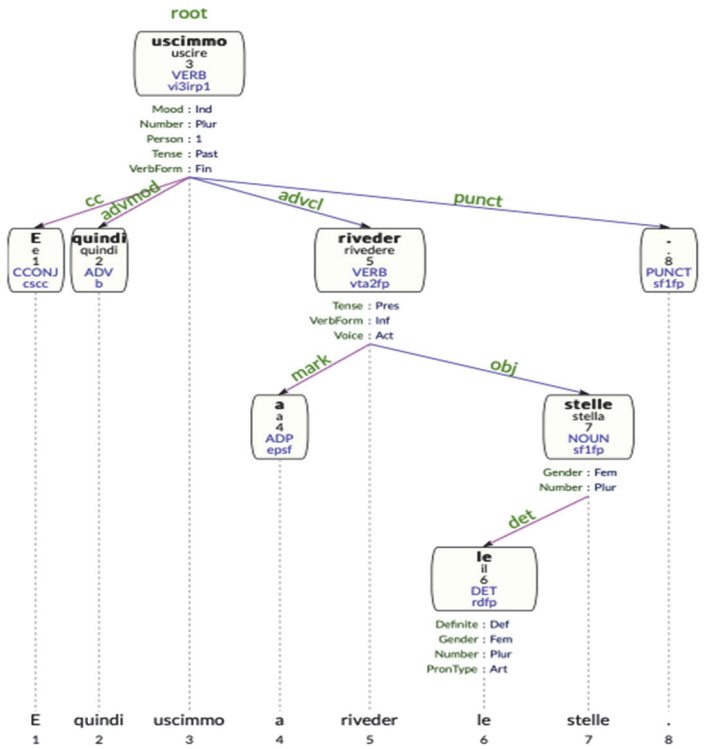


Figure 1: *Syntactic tree of Inferno, XXXIV, v. 139*

## 3. A case-study on object omission in the Divine Comedy's Inferno

The present work is intended as a preliminary corpus-based study of object omission in an Old Italian text, namely Dante's *Divine Comedy*. This text has been selected both because of its significance for the history of the Italian language (see Section 2) and because of the availability of a syntactically annotated corpus, the Italian-Old treebank, which allows for a large-scale extraction of instances of object omission. The analysis is conducted on the first Cantica, *Inferno*, with the aim of extending the study to the other two *Cantiche* in future research.

The study is carried out in semi-automatic manner, through the automatic extraction of object omissions (described in Section 4) followed by a manual verification and analysis of the retrieved instances (described in Section 5).

The analysis focuses on two main aspects: the examination of the antecedent (following the Modern Italian classification provided by Lo Duca (2010) and Cennamo (2015, 2017), see Section 1), and the investigation of the syntactic position of the verb with the omitted object, in line with the approaches of Luraghi (1998) and Egerland (2003), in order to determine whether the syntactic position of the verb plays a key role in object omission. Section 6 presents the conclusions and outlines directions for future research.

## 4. Data Extraction

Identifying something that is not overtly present in the text is inherently challenging, as the target of investigation is, by definition, absent and must be inferred from indirect or silent cues.

The UD framework, in its basic annotation scheme, does not explicitly mark ellipsis[16]. Therefore, in order to retrieve instances of object omission in transitive verbs without the necessity for exhaustive manual inspection of the entire corpus, it is necessary to develop workarounds that can support and streamline the manual identification process. To do that, the retrieval of object omissions was carried out automatically by exploiting the existing information encoded in the treebank. More specifically, the treebank includes the original tagset from DanteSearch (Tavoni 2012), which provides information on verb transitivity or intransitivity[17], encoded in the `xpos` column of the CoNLL-U format (see Section 2). Transitivity and intrasitivity are marked in the second position, immediately following the part-of-speech category (in this case v, indicating a verb), using t for transitive and `i` for intransitive verbs, respectively.

Figure 2 provides an image of the CoNLL-U format, featuring an example with the transitive verb vedere 'to see' (specifically *vedi* 'you saw'). In this example, the `xpos` field (fifth column) encodes transitivity in its second position, where `vt...` indicates that the verb is transitive.

16. For an insight into this specific issue, refer to: https://unidive.lisn.upsaclay.fr/lib/exe/fetch.php?media=meetings:general_meetings:3rd_unidive_general_meeting:23_how_to_ellipsis_a_proposal_poster_material.pdf.
17. Unfortunately, no documentation regarding the annotation process of transitivity and intransitivity has been released by the project. The only available information can be retrieved from Tavoni (2022).

```
# sent_id = Canto1-22
# text = Vedi la bestia per cu' io mi volsi; aiutami da lei, famoso saggio, ch'ella mi fa tremar le vene e i polsi».
1	Vedi	vedere	VERB	vta2mps2	Mood=Imp|Number=Sing|Person=2|Tense=Pres|VerbForm=Fin|Voice=Act	0	root	0:root	Canto=1|Verso=88
2	la	il	DET	rdfs	Definite=Def|Gender=Fem|Number=Sing|PronType=Art	3	det	3:det	Canto=1|Verso=88
3	bestia	bestia	NOUN	sf1fs	Gender=Fem|Number=Sing	1	obj	1:obj	Canto=1|Verso=88
```

Figure 2: *CoNLL-U sentence.*

The transitivity information was integrated with the syntactic annotation provided by the syntactic trees. Specifically, I selected all transitive verbs that do not have any 'object-like' nodes among their immediate syntactic dependents. By 'object-like' node, I refer to the following dependency relations (henceforth deprels): `obj`, indicating direct objects; `ccomp`, referring to complement clauses; and `xcomp`, denoting open clausal complements. When present, I also excluded subtypes of the mentioned deprels, such as `ccomp:reported`, which is used to indicate reported speech. An additional constraint involved excluding all expletive child nodes (`expl`, `expl:pv`, `expl:impers`, `expl:pass`), as these typically co-occur with verbs accompanied by the clitic *si*, a special marker used to express reflexive, middle, impersonal, and passive constructions (Pescarini 2015; Cennamo, Jezek 2011; Bentley 2004; Monachesi 1993, among others). These constructions fall outside the scope of the present study. Furthermore, I excluded child nodes labeled `nsubj:pass` (passive subjects), since the absence of an object in these cases does not carry meaningful implications for the purposes of this investigation. Lastly, in this preliminary analysis, I focused exclusively on finite verb forms, thereby excluding verbs in participial, infinitival, and gerundial forms.

To exemplify, the query was structured to retrieve instances like Example (7) and Figure (3), where the verb is marked as transitive, but it does not present as child node an "object-like" dependency relation, while discarding cases such as (8) and Figure (4), where an object child (`obj`) is explicitly present.

(7) **Vedrai** quando saranno/ più presso a
see:FUT.2SG when be:FUT.3PL more close to
noi…
us
'You'll see when they draw closer/ to us(…)' (Inf. V, vv. 76-77)

(8) **Vedi** la **bestia**
see:PRS.2SG DET.F.SG beast:F.SG
'You see the beast.' (Inf. I, v. 88)

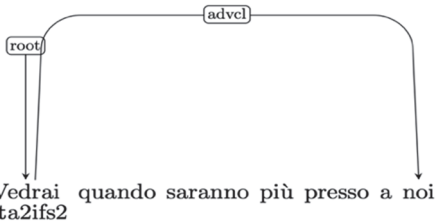

Figure 3: *Tree of a transitive verb without object.*

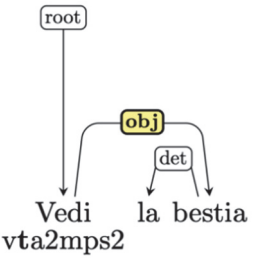

Figure 4: *Tree of a transitive verb with object.*

In example (7), the transitive verb *vedrai* 'you will see' (see.FUT.2SG) has as its only child node *noi* 'we', which is the head of an adverbial clause (`advcl`). Transitivity is indicated by the `t` in the `xpos` tagset (shown in brackets in Figure 1).

By contrast, in example (8), the verb *vedi* 'you see' (see.PRES.2SG) takes as its dependent the object node (`obj`) *bestia* 'beast'.

The query[18] and the automatic extraction were carried out using Udapi (Popel et al., 2017), an open-source framework that provides an API for processing UD data and is compatible with Python libraries. As previously mentioned, this is a preliminary test conducted on the *Inferno*, the first cantica of the *Divine Comedy*, using a portion of the Italian-Old UD treebank, version 2.15.

The retrieval, prior to manual verification, yielded 251 cases of potential object omission. After the manual check, some false-positive emerged. They are connected mainly with errors in the annotation or in the DanteSearch tagset. The total number of object omissions without false-positive in transitive verbs is 243 occurrences.

18. The query is openly available and can be shared upon request. For further information or access to the query, please contact the author of this article.

## 5. Analysis of object omission

As mentioned in Section 3, the analysis has been conducted into two directions:

- examining the referent of the omitted object. This step involves identifying the type of referent by determining whether it is indefinite and generic ("indefinite null objects"), or whether it is definite and an antecedent (or postcedent) can be traced in the text or is implicit in the semantic of the verb itself ("definite null objects").
- analyzing the syntactic context of verbs lacking an overt object. This includes determining the position of the clause (main or embedded) and its syntactic role (e.g., coordinated, comparative, or adverbial clause).

This study represents a preliminary step in a broader investigation of object omission in the *Divine Comedy*. Its goal is to contribute to the inclusion of this alternation, specifically, instances involving indefinite and generic objects, along with all others, into the PaVeDa database, which currently does not account for Old Italian.

### *5.1 Indefinite vs definite null objects*

When dealing with ellipsis, the referent, namely the linguistic element that is absent in the text, is a crucial element to analyze.

As mentioned in Section 2, object omission in Modern Italian (Lo Duca 2000; Cennamo 2015; 2017) has been classified in two different types: i) indefinite null objects; ii) definite null objects. In the data analyzed, both types of referent are present: in example (9) I report an example of indefinite and generic element, whereas in example (10) an instance of a definite one:

(9) [*si volge a l'acqua perigliosa e*] ***guata***
gaze:PRS.3SG
'Turns to the water perilous and gazes;' (*Inf.* I, v. 24)

In example (9), the omitted object is indefinite and generic. This choice appears to serve a narrative function within the author's broader stylistic strategy. The verse reported in (9) occurs at the beginning of the poem, where the author introduces the arrival of Dante's character in the *selva selvaggia* 'savage forest'. This incipit is a well-known depiction of Dante's psychological and physical journey (Inglese 2012). Upon reaching the end of the valley, Dante famously compares his emergence from the *selva* to that of a

shipwrecked man who reaches the shore and turns to gaze back at the perilous sea—*si volge a l'acqua perigliosa e guata*—without an explicit mention of the object of vision. The action described is one of turning and looking back, evoking a sense of retrospection and danger. While the omitted object in this verse is indefinite and generic, the scene gains expressive power precisely through this ellipsis: the lack of specification momentarily invites the reader to share in the character's disorientation. This stylistic device sharpens the dynamics of the scene and enhances the narrative rhythm.

It is worth noting that the verb *guardare* 'to look' is also used elsewhere in the poem with an explicit direct object, as in Example (10), suggesting that the omission in (9) is not a syntactic necessity but a deliberate narrative and stylistic choice.

| | | | | |
|---|---|---|---|---|
| (10) | ***Guarda,*** | *mi* | *disse,* | *le* |
| | gaze:PRS.3SG | 1SG.DAT | say:PST.3SG | DET.F.SG |
| | *feroci* | *Erine.* | | |
| | ferocious:M.S | Erinyes | | |
| | 'said: "Look at the ferocious Erinyes!" (*Inf*. IX, v. 45) | | | |

Among the cases of object omission, the text also includes instances involving definite objects. Interestingly, when analysing these cases, it is possible to distinguish between definite objects that have an identifiable antecedent (or postcedent) in the text, and those that are not explicitly mentioned, but are instead inferred from the context or from the lexico-semantic properties of the verb itself. Example (11) shows an instance of definite object with an antecedent, whereas example (12) is an example of definite object that is not textually expressed.

| | | | | | |
|---|---|---|---|---|---|
| (11) | [*graffia li spirti*] *ed* | | ***iscoia*** | *ed* | ***isquatra***. |
| | | and | flay:PRS.3SG | and | quarter:PRS.3SG |
| | 'He rends the spirits, flays, and quarters (them).' (*Inf*. VI, v.18) | | | | |
| (12) | [*Omè, vedete l'altro che*] | | ***digrigna****;* | | |
| | | | grind:PRS.3SG | | |
| | 'Ah me, see that one there who grinds (his teeth)!' (*Inf*. XXII, v.91) | | | | |

In example (11), the omitted objects of the verbs *iscoia* 'flays' and *isquatra* 'quarters' are both definite, referring to *li spirti* 'the spirits', which is the object of the preceding verb *graffia* 'rends'. The omission of the objects in this context would be ungrammatical in Modern Italian, where a clitic object pronoun is obligatorily expressed. This instance aligns with the observations made by Luraghi (1998) and Egerland (2003).

In example (12), the object is definite, even though it is not explicitly expressed in the text. It is recoverable from the lexico-semantic properties of the verb *digrignare* 'to grind', which is used exclusively with teeth as its object.

As previously observed with the verb *guardare* 'to look at' (Example 9 and Example 10), the verb *digrignare* is also attested in the poem also with its transitive use (Example 13):

(13) [*non vedi tu ch' e'*] ***digrignan*** ***li*** ***denti***
grind:PRS.3PL DET.M.PL tooth:M.PL
'can't you see how those demons grind their teeth?' (*Inf.* XXI, v.131)

## *5.2 Syntactic position of the verb with object omission*

Another parameter considered is the syntactic position of the verb with the omitted object, specifically whether the verb appears in the main clause or an embedded clause, and if in the latter, the type of embedded clause it occurs in. To retrieve this information, the dependency relations of the verb with the omitted object were extracted and classified. Table 1 reports the list of the dependency relations (deprels) along with their frequencies[19].

| **Frequency** | **Dependency relation** |
|---|---|
| 53 | `conj` |
| 40 | `advcl` |
| 40 | `acl:relcl` |
| 33 | `advcl:cmp` |
| 27 | `root` |
| 20 | `parataxis` |
| 17 | `ccomp:reported` |
| 4 | `ccomp` |
| 3 | `acl` |
| 2 | `csubj:relcl` |
| 2 | `csubj` |
| 2 | `flat:redup` |
| 1 | `orphan` |

Table 1. *Frequency of deprels for verbs with object omission.*

19. I specify that the frequencies reported in Table 1 serve to provide a numerical overview of the distribution of the phenomenon across syntactic positions. Relative frequencies will be calculated in future works.

Verbs in the main clause are marked with the `root` deprel, signifying that the verb is the root of the syntactic tree. All other relations listed in Table 1 correspond to embedded clauses. Specifically, `conj` refers to coordination, while `advcl` marks adverbial clauses, with subtypes such as `advcl:cmp` for comparative clauses. The `acl` relation denotes nominal modifiers, with `acl:relcl` being specific to relative clauses. The `ccomp` relation is used to express clausal complements, with `ccomp:reported` specific to reported speech, whereas `csubj` is used for subject clauses, with `csubj:relcl` referring to relative type. `Parataxis` is used in cases involving strong punctuation, such as the semicolon. Therefore, it can be considered as an embedded clause, though it is important to note that the punctuation in Old Italian texts is an editorial choice.

Finally, the `orphan` relation is used for predicate ellipsis, and `flat:redup` is employed in the case of reduplications.

### 5.2.1 Omission in root clauses

Omissions in root clauses are relatively infrequent compared to those in embedded clauses (see Table 1). In root clauses, omissions primarily involve generic objects, as illustrated in example (14):

(14) ***Cérca-****ti* *[al collo, e troverai la soga/ che 'l tien legato]*
search:IMP-2SG.DAT
'Search round thy neck, and thou wilt find the belt/ Which keeps it fastened' (*Inf.* XXXI, vv. 73-74)

In example (14), the omitted object is generic and indefinite. It is made explicit only through the coordinated verb *troverai* 'find', which present an object *la soga* 'the belt'.

Moreover, it is noteworthy that the use of generic objects in main clauses is often accompanied by the adverb *così* 'so'; an shown in example (15):

| | | | | |
|---|---|---|---|---|
| (15) | Così | **disse** | 'l | maestro; |
| | so | say:PST.3SG | DET.M.S | master:M.S |

'So said the Master;' (*Inf.* XXXI, v.130)

The adverb così ('so') functions as a correlative that resumes the preceding direct discourse.

Interestingly, cases of definite objects in root position are extremely rare. Only two occurrences have been detected, and both represent particular instances:

(16) *Or* ***movi***
now move-IMP.2SG
'Bestir (thee) now' (*Inf*. II, v. 67)

(17) *non* ***sonò*** [*sì terribilmente Orlando.*]
NEG sound:PST.3SG
'so terribly Orlando sounded not.' (*Inf*. XXXI, vv. 16-18)

In example (16), the missing element is the reflexive pronoun *ti* ('you'), which is co-referential with the implicit subject of the verb. This instance represents a specific case of endoreflexivity (Cennamo 2015), in which the verb *muovere* 'to move' is used with an inchoative (ingressive) meaning. Such transitive use of the verb *muovere* without an overt object is attested in early vernacular Italian (Ageno 1964: 73–74). In example (17), the object omission is definite, and the referent is *un alto corno* 'a loud horn' (*Inf*. XXXI, v. 12), mentioned a few verses earlier in the text. The context in which the omission occurs is a reference to another great epic, *La Chanson de Roland*.

The author draws a comparison between the sound of a horn heard by Dante, which captures his attention through the power of its sound, and the horn sounded by Orlando, the paladin of Charlemagne, who was defeated in the Battle of Roncesvalles. Although the object (the horn) is left unexpressed, the reader's familiarity with the context and the author's intention in framing it as a comparison allows its easy retrieval.

### 5.2.2 Omission in embedded clauses

Turning to object omission in embedded clauses, this phenomenon appears to be more frequent and is attested across various types of embedded structures.

In the present work, I will provide an overview of only the most frequent cases, namely object omission in coordination (`conj`), adverbial clauses (`advcl`) and relative clauses (`acl:relcl`), leaving a more fine-grained analysis of all dependency relations to future research. As shown in Table 1, object omission primarily occurs in coordinate clauses (`conj`). In these instances, the omitted object may be either generic (Example 18) or specific (Example 19):

(18) [*ma parla,*] *e* ***chiedi*** [*a lui, se più ti piace*]
and ask:PRS.2SG
'But speak, and question him, if more may please thee.' (*Inf*. XIII, v. 81)

(19) [*mostrav'alcun de' peccatori 'l dosso*] *e* ***nascondea***
*a*nd hide:PST.3SG
'one of the sinners would display his back,/ and in less time conceal (it) than it lightens.' (*Inf*. XXII, vv. 23-24)

In example (18), the omitted object of the verb *chiedi* 'question' is generic. The verse is drawn from an exchange between Virgil and Dante, following their encounter with Pier della Vigna, a chancellor of Frederick II, who is condemned to Hell and appears in the form of a trunk. After disclosing his identity and recounting his story, Pier della Vigna falls silent, thereby allowing Virgil to encourage Dante, *chiedi a lui*, to question him (i.e., the trunk) about whatever he wishes. The direct object is elegantly omitted, remaining open and indefinite, thus leaving space for Dante's own willingness.

In contrast to Example (18), the omitted object in Example (19) is specific and definite. Its referent can be traced back to the preceding conjunct, namely *'l dosso* 'the back'. The scene describes the movement of the damned souls, who are likened to dolphins swimming near the vessels, warning the sailors of an approaching storm. The sinners display their back and immediately conceal it in the pouch. Interestingly, cases such as (19) are not grammatical in Modern Italian, as an object clitic is required to preserve the grammaticality of the sentence (Calabrese 1988). It must be specified that in certain exceptions, the omission of the clitic object pronoun in coordination is still attested (Benincà & Cinque 1993), specifically when the omission of the clitic pronoun is permitted due to a strict morphological or semantic affinity between the conjuncts, as illustrated in Example (20), taken from Egerland (2003):

(20) Modern Italian:

| | | | |
|---|---|---|---|
| *Quando* | *la* | *vedo,* | *la* |
| when | F.3SG.ACC | see-PRS.1SG | F.3SG.ACC |
| *bacio* | *e* _ | ***abbraccio***. | |
| kiss-PRS.1SG | and _ | hug-PRS.1SG | |

'When I see her, I kiss her and _ hug.'

In example (19), however, no morphological or semantic affinity is attested. The absence of the object clitic pronoun in (19) aligns with the observations made by Luraghi (1998) and Egerland (2003), confirming that Old

Italian could behave differently from Modern Italian by allowing a a freer omission of the clitic object pronoun.

After occurrences in coordination, the syntactic context most frequently attested is that of adverbial clauses (`advcl`). It is worth noting that, within UD and in Italian-Old treebank, comparative clauses are marked with a specific subtype of adverbial clauses (`advcl:cmp`), which allowed me to isolate them during the extraction[20].

As in root clauses and in coordinate clauses, object omission in adverbial clauses (and comparative ones) can also involve either a specific or generic object. Examples of specific and generic object omission are provided in (21) and (22), respectively.

(21) [*mi disse, «riconoscimi,*] *se* ***sai****:*
if can:PRS.2SG
'he said to me, "recall me, if you can;' (*Inf*. VI, v. 41)

(22) [E mentre ch'io là giù con l'occhio] cerco,
search:PRS.1SG
[vidi un col capo sì di merda lordo]
'And whilst below there with mine eye I search,/ I saw one with his head so foul with ordure,' (Inf. XVIII, vv. 115-116)

In Example (21), the omitted object is specific and refers to the act of recognizing. As with Example (19), Example (21) would also be ungrammatical in Modern Italian, requiring the presence of an object clitic. In Example (22), by contrast, the omitted object is generic, referring to Dante searching for someone in moat crowded with people covered in filth.

Finally, in line with the other syntactic contexts, relative clauses also display both types of object omission, as illustrated in (23) and (24):

(23) [*fece li cieli e diè lor chi*] ***conduce***
guide:PRS.3SG
'The heavens created, and gave who should guide (them)' (*Inf*., VII, v. 74)

(24) [*E qual è quei che volontieri*] ***acquista****,/*
acquire:PRS.3SG
[*e giugne 'l tempo che perder lo face,*]
'And as he is who willingly acquires,/ And the time comes that causes him to lose,' (*Inf*. I, vv. 55-56)

20. Due to space constraints, I will not engage in detailed discussion of comparative clauses in this work. A discussion of these cases is deferred to future work.

Example (23) shows a specific omitted object, whose referent is *li cieli* ('the heavens'). Once again, such a structure would be ungrammatical in Modern Italian.

In contrast, in Example (24), the omitted object is generic. The verse occurs within a comparison between Dante's feelings and those of a miser or a gambler who gains material goods or money. As seen in Example (9), the allusiveness and genericity of the action can be interpreted as serving to engage the reader's imaginative capacity.

## 6. Conclusion

The present work is a preliminary corpus-based study of object omission in the *Inferno* of Dante Alighieri's *Divine Comedy*. The analysis was conducted using semi-automatic methods, exploiting the availability of Italian-Old, an annotated corpus (treebank) that provides transitivity and syntactic annotations.

A manual analysis was performed to examine the nature of the referent of the omitted object, specifically whether it is definite or indefinite, in order to assess whether the classification of object omission in Modern Italian (Lo Duca 2000; Cennamo 2015, 2017) is also applicable to Old Italian. The second aspect of the study involved the syntactic position of the verb with the omitted object, as it has been shown that, specifically in coordination, object omission differs from Modern Italian, highlighting a process of clitic grammaticalization (Luraghi 1998).

This preliminary analysis reveals that the classification of referents is indeed attested in Old Italian, which displays both indefinite and definite null objects. Interestingly, indefinite objects seem to serve a stylistic function, supporting the author's narrative.

Regarding the syntactic position of the verb with omitted objects, indefinite objects appear both in main and embedded clauses, while definite objects primarily appear in embedded clauses, although some instances have also been found in main clauses.

Future research will focus on exploring object omission in the other Cantiche, *Purgatorio* and *Paradiso*, with the aim of including this alternation—specifically, object omission involving generic and indefinite objects—alongside others, in the PaVeDa database, in order to provide valency information of Old Italian.

## References


Ascoli, G. I. 1882–1885. L'Italia dialettale. Archivio glottologico italiano, VIII:98–128.

Benincà, P, G. Cinque. 1993. Su alcune differenze fra enclisi e proclisi. In Omaggio a Gianfranco Folena, 2313–2326. Padova: Editoriale Programma.

Bentley, D. 2004. Unexpressed arguments: si-constructions in Italian. In RRG2004 Book of Proceedings: linguistic theory and practice: description, implementation and processing, 17-48. Institute of Technology Blanchardstown, Dublin, Ireland.

Brambilla, A. F. 1964. Il verbo nell'italiano antico: ricerche di sintassi. Milano-Napoli: R. Ricciardi.

Calabrese, A. 1988. I pronomi clitici. In Renzi, L. (ed.) Grande grammatica italiana di consultazione I, 549–592. Bologna: Il Mulino.

Cennamo, M. & E. Jezek. 2011. The anticausative alternation in Italian. In Atti del XLIII Congresso internazionale di studi della Società di linguistica Italiana (SLI), 809-823. Verona, 24-26.11.2009.

Cennamo, M. & C. Fabrizio. 2013. Italian valency patterns. In I. Hartmann, M. Haspelmath, B. Taylor (eds.), Valency Patterns Leipzig. Leipzig: Max Planck Institute for Evolutionary Anthropology. (accessed 29 Apr 25).

Cennamo, M. 2015. 13. Valency patterns in Italian. In A. Malchukov & B. Comrie (eds.) Volume 1 Introducing the Framework, and Case Studies from Africa and Eurasia, 417-482. Berlin, Boston: De Gruyter.

Corbetta, C., M. Passarotti, F. Cecchini & G. Moretti. 2023. Highway to Hell. Towards a Universal Dependencies Treebank for Dante Alighieri's Comedy. In Proceedings of the 9th Italian Conference on Computational Linguistics (CLiC-it 2023), 1-8. Venice, Italy.

Croft, W. 2001. Radical construction grammar: Syntactic theory in typological perspective. Oxford: OUP.

Dardano, M. & Frenguelli, G. (Eds.). (2004). SintAnt. Aracne, Rome, Italy.

De Marneffe, M.C., C. D. Manning, J. Nivre & D. Zeman. 2021. Universal Dependencies. Computational linguistics, 47(2): 255-308.

Egerland, V. 2003. Sull'omissione del pronome clitico oggetto in italiano antico. Verbum, 4(2): 349-366.

Egerland, V. & A. Cardinaletti. 2010. I pronomi personali e riflessivi. Grammatica dell'italiano antico 1: 401-467. Bologna: Il Mulino.

Fillmore, C. J. 1986. Pragmatically controlled zero anaphora. In Annual Meeting of the Berkeley Linguistics Society, 95-107.

Giacalone Ramat, A. 2011. Discourse functions and syntactic change. In Historical Linguistics 1987: Papers from the 8th International Conference on Historical Linguistics, Lille, August 30-September 4, 1987, 175-190. Amsterdam: John Benjamins Publishing Company.

Hartmann, I., M. Haspelmath & B. Taylor. 2013. Valency Patterns Leipzig. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Haspelmath, M. & I. Hartmann. 2015. Comparing verbal valency across languages. In A. Malchukov & B. Comrie (eds.), Valency classes in the world's languages, 41-71. Berlin/New York: Mouton de Gruyter.
Heinecke, J. 2019. ConlluEditor: a fully graphical editor for Universal dependencies treebank files. In Proceedings of the Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019), 87-93.
Inglese, G. 2012. Dante: guida alla Divina Commedia. Roma: Carocci.
Luraghi, S. 1997. Omission of the direct object in Latin. Indogermanische Forschungen 102: 239-257.
Luraghi, S. 1998. Omissione dell'oggetto diretto in frasi coordinate: dal latino all'italiano. In Sintassi storica. Atti del XXX Congresso SLI: 183-96.
Luraghi, S. 2004. Null objects in Latin and Greek and the relevance of linguistic typology for language reconstruction. In Proceedings of the 15th Annual UCLA Indo-European Conference, JIES Monograph 49: 234-256.
Manni, P. 2013. La lingua di Dante. Le vie della civiltà. Bologna: Il Mulino.
Marello, C. 2010. Ellissi. In G. Salvi & L. Renzi, Grammatica dell'italiano antico (Vol. 2), 1369-1386. Bologna: Il Mulino.
Migliorini, B. 1988. Storia della lingua italiana. Firenze: Sansoni.
Monachesi, P. 1993. On si constructions in Italian HPSG Grammar. In Proceedings of the TenthEastern States Conference on Linguistics.
Pescarini, D. 2015. Le costruzioni con si. Italiano, dialetti e lingue romanze. Roma: Carocci.
Popel, M., Z. Žabokrtský, M. Vojtek. 2017. Udapi: Universal API for universal dependencies. In Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies UDW 2017: 96-101.
Salvi, G. & L. Renzi. 2010. Grammatica dell'italiano antico (1,2). Bologna: Il Mulino.
Tavoni, M. 2011. DanteSearch: il corpus delle opere volgari e latine di Dante lemmatizzate con marcatura grammaticale e sintattica. In Omaggio a Vincenzo Placella per i suoi settanta anni (Vol. 2): 583-608. Università degli Studi di Napoli "L'Orientale": Il Torcoliere-Officine.
Tavoni, M. 2022. Allestimento, fruizione e prospettive di DanteSearch. In Corpora e Studi Linguistici. Atti del liv Congresso della Società di Linguistica Italiana 6: 255-273. Online, 8-10 settembre 2021.
Tesi, R. 2004. Parametri sintattici per la definizione di "Italiano antico". In M. Dardano & G. Frenguelli (eds.), SintAnt. La sintassi dell'italiano antico, 425–444. Roma: Aracne.
Tomasin, L. 2019. Il caos e l'ordine. Torino: Piccola Biblioteca Einaudi.
Zanchi, C., S. Luraghi & C. R. Combei. 2022. PaVeDa – Pavia Verbs Database: Challenges and perspectives. In Proceedings of the 4th workshop on research in computational linguistic typology and multilingual NLP, 99-102. Seattle, Washington: Association for Computational Linguistics.

# Valency in the British Isles: Lability Between English and Celtic

*Michele Tron*[*]
*Martina Giarda*[**]

## 1. Introduction[1]

Following Haspelmath et al. (2014: 590), causal verbs or verbal expressions can be characterized as including of a "cause" meaning component, while noncausal verbs share the same core meaning but omit the "cause" element. Similarly, Zúñiga & Kittilä (2019: 15, 41) describe causatives as verbs that add an external agent (causer) to the event, whereas anticausatives remove the agent component. Labile verbs, however, simultaneously represent both the causal and non-causal meanings within a single verb, without relying on morphological distinctions.

The orientation of present-day English, as well as that of Irish and Welsh, is heavily skewed in favour of labile pairs (cf. the list of 31 verbs in Poppe 2009: 260, 262-264), in contrast to what is true for Standard Average European (cf. Haspelmath 1993, 2001), although with the (partial) exception of German and Greek. Poppe (2009), in his investigation of a possible contact-based explanation for the development of identical intensive and reflexive pronouns in English and Insular Celtic, also discusses the issue of lability. Indeed, he perceptively observes that "[t]he concomitant typological similarities between Welsh, Irish, and SE with regard to the incidence of labile verbs is therefore a typological reflex of earlier developments, related to the emergence of new complex reflexives in SE" (cit. Poppe 2009: 261): however, the similarities in the valency profile of Insular Celtic and English has never

[*] University of Pavia
[**] University of Bergamo/Pavia

1. This paper is the result of close collaboration between the two authors. For the academic purposes of attribution, Michele Tron is responsible for Sections 1, 3, 3.3, 4, while Martina Giarda is responsible for Sections 2 and 3.2.



 ISBN 9788835183631

been verified for the earlier stages of the languages, in particular in light of the contact hypothesis.

Studies on Celtic-English contact are well-known in English Historical Linguistics (cf. Filpulla & Klemola 2010 and the literature cited therein). Allen (2019) suggests Celtic influence in the loss of dative external possessors, whereas Vezzosi (2005) proposed a Celtic origin for "himself" in Middle English, and Poussa (1990) linked periphrastic *do* to Celtic contact. Still, many of these substratal hypotheses have not always withstood careful examination, with some of them demonstrably showing "cogent internal arguments for their development so that contact can have provided support for their development but can hardly have been the sole origin" (cit. Hickey 2012: 505; also cf. Hickey 2020: 2-5).

However, two matters have yet to be settled: whether it is legitimate to project the prominence of lability backwards, more so for Welsh, as well as Irish, in its earlier stages than for English, for which the matter has received substantial attention in the literature; and whether there is a correspondence between the behaviour of (Middle) Welsh and (Old and Middle) English *vis-à-vis* (Old) Irish. Such an ambitious objective cannot be conclusively reached within this study: hence, our goals here will be mainly methodological in nature and will serve to demonstrate the potential of the Pavia Verbs Database (henceforth: PaVeDa; Zanchi et al. 2022) and its relevance to the topics just mentioned.[2]

## 2. Lability in English and Celtic

The development of lability in English is well-studied (cf. among others McMillion 2006). Van Gelderen (2011, 2018) and García García (2020) link the decline of Germanic causative formation to the rise of lability in English, discussing Old English basic valency. Visser (1963) covers the diachrony of English verbs, including lability. As for Middle English, Zimmermann (2024) statistically demonstrated that the verb *fear* transitioned from a causal to a non-causal meaning, illustrating its lability.

One development that needs to be mentioned is the evolution of morphological causatives: the formation of causative verbs using the *-ja* suffix is a

2. Our study is based on a selection of relevant verbs among those featured in the PaVeDa, in which data of Old English is already present, whereas Middle English, Old Irish, and Middle Welsh are currently under review and will soon be made available. The Welsh section of the PaVeDa ultimately aims to include three stages of the language, i.e., Middle Welsh, Early Modern Welsh, and Modern Welsh, thus covering the whole of the historical evolution of Welsh.

feature present in all Germanic languages, though to varying degrees, as this process likely existed in Proto-Germanic in an "embryonic state" (Ottósson 2013: 329). This might explain why Old English retains only traces of this suffixation, unlike languages such as Gothic, where "the causative member of the alternation is almost always expressed by *ja*-verbs" (Zanchi & Tarsi 2021: 63). García García (2020: 163) identifies 59 clear causative-noncausative verb pairs in Old English, distinguished by vowel alternation caused by the vowel raising effects of the original *-ja* suffix (Saibene & Buzzoni 2014: 116-117). However, the *-ja* causative formation was no longer productive in Old English, nor was it ever consistently applied during that phase of the language, as the association of strong verbs with intransitivity and weak verbs with transitivity/causativity was unstable (García García 2020: 170). Several of the pairs identified by García García (2020) align with the PaVeDa meanings, namely BOIL, BURN, KILL and DIE, SINK, TEAR, and BE COLD. It is worth noting that some of these verbs already showed signs of lability, albeit to a limited degree: although Old English still expressed this alternation through verbal pairs differentiated by vowel raising triggered by the original *-ja* suffix, some verbs are already found in an unmarked (i.e., labile) causal/non-causal alternation.

Interestingly, although labile verbs were only marginally present in Old English, their number significantly increased during Early Middle English (1150–1300, cf. Ingham 2020; McMillion 2006). Ingham (2020) suggests that this rise may have been driven by the substantial French influence of the time, which was both lexical and structural. While both Old English and Old French exhibited lability, their linguistic paths diverged: English saw an increase in lability, whereas French adopted new strategies to encode causality, such as the reflexive middle, hence losing many lability as an alternation for a large amount of verbs: still, Old French, namely Anglo-Norman, retained lability to a notable degree (Ingham 2020: 448). Ingham (2020) also highlights significant changes in Middle English, noting that fewer than 20% of native Old English verbs in the semantic domains of change-of-state and change-of-position were labile. By Middle English, this figure had risen to two-thirds, with an even greater proportion of labile verbs within these semantic fields introduced through French borrowings.[3] Conversely, García García and Ruiz Narbona (2021) propose that the origins of lability date back to Early Old English, or perhaps even the transition from Common Germanic, downplaying the extent of French influence on Middle English.

3. This could indeed be said to represent the core of labile verbs in Modern English, as well as in Modern Welsh and Modern Irish (cf. Poppe: 2009: 259-264).

The topic of lability, however, has not been explored to the same extent for the (Insular) Celtic languages nor for Middle Welsh specifically:[4] a thorough description of the overall valency profile of Middle Welsh is lacking at the moment, although several specific topics have received some attention in the literature, such as the expression of agents in the older stages of Welsh as well as Irish (Müller 1999), the marking of S-arguments on non-finite verbal forms (Manning 1995, Borsley et al. 2007: 327-330, Sackmann 2022, Rouveret 2023: 227-228, 234-251), the peculiar syntax of the latter in the Middle Welsh literature (Evans 1964: 160-161, Schumacher 2011: 202-203), or the existence of a prefix *ym-* involved in a number of valency-reducing operations (Irslinger 2017).

## 3. Data analysis

This section introduces the data taken from the PaVeDa that were exploited here after a brief historical sketch of each language and the periodisations here adopted. For the purposes of this study, 12 pairs of verbal concepts were chosen (see Table 1), 11 of which had at least one corresponding verbal concept, either transitive or intransitive or both, in the PaVeDa, to which WAKE UP was further added. The verbs were chosen among those in Roma & Zanchi's (2025) list in order to allow the simultaneous comparison with Old Irish, but for reasons of space we departed from their tripartite analysis by selecting only the relevant pairs in a causal/non-causal alternation without differentiating between non-causal continuous (atelic) meanings and non-causal bounded (telic) ones, as Nichols (2017) does in her revised list.

| | non-causal | causal counterpart |
|---|---|---|
| 1. | LAUGH | AMUSE |
| 2. | DIE | KILL |
| 3. | SIT | SEAT |
| 4. | EAT | FEED |
| 5. | LEARN | TEACH |
| 6. | SEE | SHOW |
| 7. | FEAR | FRIGHTEN |
| 8. | HIDE | HIDE (TR.) |
| 9. | BOIL | BOIL (TR.) |
| 10. | BURN | BURN (TR.) |

4. But see Borsley, Tallerman & Willis (2007: 278-279) on Modern Welsh.

| | | |
|---|---|---|
| 11. | BREAK | BREAK (TR.) |
| 12. | WAKE UP | WAKE UP (TR.) |

Table 1: *non-causal and causal verb pairs.*

## *3.1 Old and Middle English*

Old English is an ancient Germanic language, part of the West-Germanic subbranch, specifically within the Ingvaeonic group, alongside Old Saxon and Old Frisian. Conventionally, the Anglo-Saxon period in England's history spans from 449 CE (according to Bede's *Historia ecclesiastica*) to 1066, the year of the Norman Conquest: regardless of the exact beginning of the Anglo-Saxon period, written works in Old English began to flourish in the 8th century CE. The Norman Conquest in 1066 ended the Old English period and began Middle English. French-speaking nobility replaced the Anglo-Saxon elite, and Norman prelates took over church positions. For 200 years, French was the upper-class language, while English persisted among the masses, disappearing from written records until the 13th century. Thus, the Middle English period is conventionally considered to last from the Norman Conquest in 1066 to the early 15th century. Unlike Old English, which shows little diachronic evolution over the cen-turies, Middle English is characterized by significant changes (Baugh & Cable 2013: 152), affecting its phonology, morphosyntax, and vocabulary.

Data for this analysis (and for the OE/ME PaVeDa database in general) come from balanced subcorpora of the historical treebanks of English, namely the *York-Toronto-Helsinki Parsed Corpus of Old English Prose* (YCOE, Taylor et al. 2003), *The York-Helsinki Parsed Corpus of Old English Poetry* (YCOEP, Pintzuk & Plug 2002), the *Penn-Helsinki Parsed Corpus of Middle English, second edition* (PPCME2, Kroch & Taylor 2000), and *The Parsed Corpus of Middle English Poetry* (PCMEP, Zimmermann 2014).[5]

Old English shows only marginal occurrences of lability, namely for one of the verbs in the list: *bærnan* 'burn' is a two-place verb, which nonetheless can be found in some instances with a non-causal meaning, as in (1).[6]

5. The texts selected to be part of the subcorpora are available at the following link: https://drive.google.com/file/d/1kGDLULwAoNdWDUwQ_8YeN7qZq47_-GT9/view?usp=drive_link.

6. All examples are glossed according to the Leipzig Glossing Rules, retrievable at the following link: https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf, with a lower degree of precision for the morphological features that are not under discussion here. The list of abbreviations that are not found in the standard list is the following: PRET = preterite; PTCL = particle;

(1) *Fyr* *ic* *sende* *on*
fire(N).ACC.SG 1SG.NOM send.1SG on
*eorþan* *&* *hwæt* *wylle* *ic*
earth(F).ACC.SG and what.ACC.SG.N want.1SG 1SG.NOM
*buton* *þæt* *hit* *bærne*
except that 3SG.NOM.N burn.SBJV.PRS.3SG
'I have sent fire on the earth, and how I wish it was already burning!' (cowsgosp,Lk_[WSCp]:12.49.4720)

Suppletion (above all through opaque *-ja* causatives) is on the contrary well represented, being present for most of the other verbs on the list (9 out of 12). Concerning the alternation between KILL and DIE, several verbal lemmas are attested, thus suppletion can be said to be the most frequent strategy. Remarkably, The Dictionary of Old English (Cameron et al. 2024) lists *cwelan* 'die' and *cwellan* 'kill' as separate lemmas. However, *cwellan* 'kill' is also used in non-causal contexts, meaning 'die'. Given that causal vs. non-causal alternation often involves verb pairs with different thematic vowels (a relic of the old Germanic *-ja* suffix), and graphical variants are frequent in Old English, it can be argued that *cwelan* and *cwellan* might be different spellings of the same verb, indicating lability (cf. example (2)).

(2) *se* *bið* *swylce* *he* *sie*
DET.NOM.SG.M be.3SG like 3SG.NOM.M be.SBJV.PRS.3SG
*ealne* *weg* *cwellende*
all.ACC.SG.M way(M).ACC.SG kill/die.PTCP.PRS
'[He who is always afraid] is like one who is always dying' (codicts,Prov_1_[Cox]:1.76.140)

On the other hand, the number of labile verbs increases in Middle English (5 out of 12), though suppletion is still well attested. In particular, verbs such as *brennen* 'burn' and *boillen* 'boil' seem to merge the causal and non-causal meanings, as they were not labile in Old English (except for a few occurrences of *bærnan*).[7] The verb *leren* 'teach' also shows some marginal non-causal occurrences, meaning 'learn', in which the R-argument is omitted, as in (3)-(4).

VN = verbal noun; IND (= indicative) and PRS (= present) are implicit whenever no indication of the TAM features of the verb is found.
7. Note, however, that although BURN *brennen* is of Germanic origin, BOIL *boillen* is a loanword from French.

(3) *I* *sal* *yu* *lere* *þe* *dute* *of*
1SG.NOM shall 2PL.ACC teach.INF DET doubt of
*god*
God
'I shall teach you the fear of God' (CMBENRUL-M3,2.20)

(4) *So* *wolde* *he* *his* *mester*
so want.PRET.3SG 3SG.NOM.M POSS.3SG.M job
*lere*
teach.INF
'And was so eager to learn his trade' (Havelok,26.823.413)

The verb *breken* 'break' (< OE *brecan*) is clearly labile in Middle English, whereas the non-causal meaning was expressed through other strategies in Old English, namely with the adposition *ut* or through a "non-causal passive", as in (5)-(6).[8]

(5) *&* *hi* *næfre* *syððan* *ut* *brecan*
and 3PL.NOM never afterwards out break.INF
*ne* *magon*
NEG may.IND.PRS.PL
'and they never afterwards may burst out' (cocathom1,ÆCHom_I,_11:270.113.2082)

(6) *&* *se* *ðuma* *gebrocen*
and DET.NOM.SG.M thumb(M).NOM.SG break.PTCP.PST
*wæs*
be.PRET.3SG
'[when I was falling, I came driving with my head and hand on the stone.] And the thumb broke' (cobede,Bede_5:6.400.28.4025)

Concerning the alternation between FEAR and FRIGHTEN, Van Gelderen (2014: 106) argues that until the late 14th century, the verb OE *færan* (> ME *feren*) meant 'frighten', however Zimmermann (2024: 13) shows that the change in argument structure, and thus in meaning, begins far earlier than 1400.[9]

The verb *hiden* (from OE *hydan*) is the only verb of the sample showing a valency-reducing alternation ("direct reflexive"), as in (8). Reflexive and

8. https://valpal.apnetwork.it/alternations/13542966347.
9. "The emergence of the new meaning took place during a poorly attested period of the English language, perhaps as early as the beginning of the 13th century." (Zimmermann 2024: 38).

non-causal categories are linked in Old and Middle English (cf. Kemmer 1993, Poppe 2009), making them hard to distinguish from passive. *Hiden* also shows "understood reflexive objects", interpreted as non-causals. Due to a higher degree of intentionality, these intransitive usages are seen as having an "understood reflexive object".[10]

(7) *but* *Jhesus* *hidde* *hym*
but Jesus hide.PRET.3SG 3SG.ACC.M.REFL
'but Jesus hid himself' (CMNTEST-M3,8,40J.844)

(8) *Hides* *and* *helis* *als* *hende,* *For*
hide.IMP and cover.IMP as quickly because
*ge* *er* *cast* *in* *care*
2PL.NOM be.PRS.PL cast.PTCP.PST in sorrow
'hide and cover [yourselves] as quickly as possible, because you are cast in sorrow' (LaurMinot.[Poem_6],18.18.296)

## *3.2 Middle Welsh*

Middle Welsh is the second Celtic language to be added to the PaVeDa, the first being Old Irish (Roma 2021, 2024, Roma & Zanchi 2025). The textual data included in the database are extracted from a corpus of eleven narrative prose texts collectively known under the name of *Mabinogion*: despite the vast amount of technical literature in the language (i.e., medical, legislative, and historical), narrative prose was preferred on the basis of both its variety of syntactic structures and the overall absence of syntactic calques and interferences, which are instead common in translated texts, e.g., from Latin, as is the case for annalistic and religious literature. This corpus is ideal for providing a consistent description of the verbal syntax of Middle Welsh, which is one of the main aims of the inclusion of Middle Welsh in the PaVeDa project: Middle Welsh is in fact the oldest stage of the Welsh language for which the extant documentation allows us to produce a full description of its grammatical system.[11]

10. In contrast, in the same contexts, verbs like SHOW *sheuen* are labelled "non-causal" as they often mean 'to appear', lacking control and volition.

11. Old Welsh designates the period from the first direct attestations of the languages around the mid-8th century, mainly consisting of Welsh language glosses in Latin manuscripts, up until the beginning of the Middle Welsh period in the 12th century. Despite not containing many verbal forms, one single form *cantarceint* 'close.PASS.3PL' of the verb, *cau* 'close', which is labile in Modern Welsh, is found in an Old Welsh gloss: the fact that the form is a passive, in Lambert's (1984: 187) proposal, could be taken to mean that the verb was not labile in Old Welsh, as the non-causal meaning could be encoded by the passive.

No single periodisation of the language is found in the literature: Evans (1964: XVI-XIX) defines Middle Welsh as the period extending from the 12th century to the end of the 14th, roughly coinciding with the start of Norman influence in Wales, and further divides it in an early period, which goes up to the middle of the 13th century, and a late period, which covers the remaining century and a half. A more straightforward chronology is proposed by Borsley et al. (2007: 286), who set the temporal boundaries of the Middle Welsh period between 1150 and 1500: this is the terminology adopted here. Although the texts included in the PaVeDa database are all found in manuscripts copied between the 14th and the 15th century, the date of their original composition has been argued never to be later than the 13th century, thus representing the older couch of the Middle Welsh period.[12]

The Middle Welsh equivalents of the first four intransitive verbs in the list, i.e., are not attested as labile verbs, deriving instead their causal counterparts either through suppletion or valency-augmenting alternations such as causative periphrases: these are *chwerthin* 'laugh', *(bod) marw* 'die', *eiste* 'sit', and *bwyta* 'eat'. Interestingly, the most frequent verb, *bwyta* 'eat', never appears in our corpus with an object argument, unlike its telic counterpart *ysu* 'eat up, devour', which is always found with one:[13] an example of the typical contexts in which they appears is found in (9)-(10), followed by an example of their causative equivalent, which is normally conveyed by the verb *llithio* 'feed'.[14]

| | | | | | | |
|---|---|---|---|---|---|---|
| (9) | *Mi* | *a* | *uynnaf* | *uwytta* | *o* | *honno* |
| | 1SG | PTCL | want.1SG | eat.VN | from | that.F |
| | 'I want to eat from that [basket]' (Bromwich & Evans: 1992: 23) | | | | | |
| (10) | *Gadwn* | *ef,* | *heb* | *y* | *Kei,* | *y* |
| | leave.IMP.1PL | 3SG.M | QUOT.3SG | PTCL | Cai | to |
| | *yssu* | *y* | *wala* | *o* | *'r* | *kic* |

Unfortunately, this interpretation relies on a conjectural reading of the actual form found in the manuscript, which is *cantarteint* 'they raise together', glossing lat. *perstruuntur* (cf. Falileyev 2000: 21-22), so the only piece of evidence which would have allowed us to argue in favour or against the existence of lability in Old Welsh remains highly speculative.

12. The datings of each of the texts and the issues surrounding them are far too complex to be discussed here: see recent accounts in Rodway (2005, 2007, 2013) on matters of dating.

13. Another interesting fact is that *bwyta* 'eat' is undoubtedly transitive in Modern Welsh (v. GPC, s.v. *bwyta*), while transitive constructions with this verb in Middle Welsh are surprisingly scant, even outside our corpus.

14. Here and in the rest of the paper Middle Welsh examples are given according to the Middle Welsh spelling, but when citing and discussing Welsh words in the text the Modern Welsh orthography is used instead, following the *Geiriadur Prifysgol Cymru*'s (GPC) conventions, as is standard in the literature.

eat_up.VN his fill of the meat
'We will leave him, said Cai, to eat his fill of the meat' (Bromwich & Evans 1992: 35)

The same lexically suppletive shift is found with the pair DIE/KILL: in this case, however, DIE is expressed by a complex predicate made up of the copula *bod* 'be' plus the adjective *marw* 'dead', while KILL is usually expressed by the verb *lladd* 'kill, strike'. Still, as can be seen from the example in (11) where *(bod) marw* and *lladd* are found side by side, the relationship between the non-causal verb and its causal counterpart is, again, one of suppletion.

(11) *ac* *ar* *nys* *llado* *marw* *vydant* *o* *newyn*
and those NEG kill.SBJV.3SG dead be.3PL of hunger
'[…] and those he does not kill die of starvation' (Goetinck 1976: 68)

The first verb encountered in our list which is labile in Middle Welsh is *dysgu* 'learn/teach': the same verb can mean either 'learn' or 'teach' depending on the valency frame in which it appears, as the addition of a third argument, i.e., the recipient, introduced by the preposition *i* 'to', is sufficient to mark the causative meaning and this is indeed true at any stage of the language (v. GPC, s.v. *dysgu*), so that virtually no diachronic variation can be identified.[15] The two examples in (12)-(13) present two instances of respectively the non-causal and causal meanings:

(12) *a* *chyt* *a* *mi* *y* *bydy* *y* *wers* *hon* *yn* *dysgu* *moes* *a* *mynut*
and together with 1SG PTCL be.2SG DET period that.F PRED learn/teach.VN manner and courtesy
'[…] and you will stay with me for a while, learning manners and etiquette' (Goetinck 1976: 18)

(13) *a* *mi* *a* *dyscaf* *it* *wniaw*
and 1SG PTCL learn/teach.1SG to.2SG sew.VN
'[…] and I will teach you how to stitch' (Williams 1930: 54)

15. Ostensibly, *dysgu* is a Latin loanword from *discō* 'learn'. Nonetheless, no stance can be taken on whether the causative or the non-causative meaning is diachronically primary, as the verb is also attested as meaning 'teach' in Late Latin (cf. TLL, s.v. *disco*), or even borrowed already as a labile verb.

The pair SEE/SHOW and FEAR/FRIGHTEN are again expressed, as for the first four verbs mentioned here, by suppletive pairs of lexemes. *Gweled* 'see' contrasts with *dangos* 'show' but, notably, *dangos* also has a *ym*-prefixed counterpart, *ymddangos* 'show oneself', in turn deriving the anticausative 'show (oneself)'.[16] Moreover, the same prefix as *ymddangos*, which is not relevant here, is found in the non-causal counterpart of *celu* 'hide', i.e., *ymgelu* 'hide (oneself)'. On the other hand, we have another complex predicate featuring the noun *ofn* 'fear' along with the copula and a prepositional phrase marked with *ar* 'on, upon' representing the fearer argument, as in (17); FRIGHTEN is in turn expressed by suppletion, either by a lexical verb or by another complex predicate with the same noun.[17]

Lastly, lability is found for all of the last four verbs of the list, which are *berwi* 'boil', *llosgi* 'break', *torri* 'break', and *deffroi* 'wake up'. One pair of examples given below in (14)-(15) will suffice to illustrate the behaviour of *deffroi*, (14) being the non-causal example and (15) the causal one:

(14) *kyn* *hanner* *noss* *kyscu* *a* *wnaeth*
before half night sleep.VN PTCL do.PRET.3SG
*pawb* *ohonunt* *a* *thu* *a* *'r* *plygeint*
each of.3PL and side with the cockcrow
*deffroi*
wake.VN
'[…] before midnight each one of them fell asleep and woke up towards cockcrow' (Williams 1930: 20)

(15) *A* *phan* *welas* *y* *meichat* *lliw* *o*
and when see.PRET.3SG the swinherd colour of
*dyd* *ef* *a* *deffroes* *Wydyon*
day 3SG.M PTCL wake.PRET.3SG Gwydion
'And when the swineheard saw daylight he woke Gwydion' (Wil liams 1930: 89)

16. There is, however, one attestation of *dangos* without any anticausative derivation possibly meaning 'show (oneself)', which is required by the context, in the first of the *Four Branches*, *Pwyll Pendefig Dyfed*. The example has garnered much of the editors' attention, but the passage is usually taken to be somewhat defective, as it requires either *ymddangos* or an object (Williams 1930: 139). For this reasons, it has been taken into account in the PaVeDa, although dubitatively, and it will not be considered here for the purposes of the paper, as it does not have any bearing on its relationship with *gweled*.

17. See Irslinger (2017) for the functions of this prefix. Incidentally, no adequate equivalents of the latter meaning are attested in the corpus, but some are found outside of it, e.g., *ofnocáu* or *ofnhau*, both meaning 'frighten'. A lexical verb for FEAR is also found in the corpus, i.e., *arswydo* 'fear, dread'.

### *3.3 Comparing across languages within the PaVeDa*

The results of our twofold analysis is presented in Table 1: in the first and second column Old and Middle English are presented side by side, while the Middle Welsh data is found in the third column alongside the relevant data taken from Roma & Zanchi (2025). Each label in the cells indicates the type of formal relationship and its orientation between the non-causal and causal meanings. Lability is attested in Middle Welsh for some of the verbs in the list, although the orientation of the language does not seem overall to be completely skewed towards lability as soon as the verbs outside of the change-of-state core meanings (cf. note 1) are taken into account. In particular, of the 12 verbs here under scrutiny, less than half exhibit a labile alternation, all of which, however, coincide with those that are labile in Poppe's (2009: 262-264) list of Modern Welsh verbs. The one verb which is labile both in Middle Welsh and Modern Welsh, i.e., LEARN/TEACH but is suppletive and not labile in modern English is attested as labile only in some ME occurrences, but not in OE. Suppletion is on the contrary well represented in our list, being present for most of the other verbs (6 out of 12), with one more occurrence if the pair *celu-ymgelu* is taken to be suppletive instead of a valency reduction alternation (represented by /R between brackets in the table), given the non-productivity of *ym*-prefixation in Welsh.

Old Irish features a higher amount of reduction, either by reflexivisation or voice alternation, here found for LEARN/TEACH, HIDE, BURN, and BREAK, and a very low incidence of lability, which is found only for WAKE UP. Note that the valency reducing alternation of HIDE is not derivational as in Middle Welsh, despite the Welsh and Irish verbs being direct cognates: however, the same strategy, valency reduction by voice alternation, i.e., the impersonal form, exists in Middle Welsh as well, although it is not represented in our sample for this verb and thus it has not been added here.

Two sets of discrepancies can be observed: on the one hand, it is clear that during the transition from Old English to Middle English, the language has clearly shifted towards lability, as we find 6 labile verbs in Middle English, i.e., half out of the total 12, against only 2 in Old English; on the other hand, Middle Welsh, as represented in our corpus, makes limited use of valency reduction and developed in turn a higher number of labile verbs (5 out of 12) in contrast to Old Irish, which is the most relevant point of comparison for philogenetic reasons, where lability is confined to a single verb.

| | OE | ME | MW | OI |
|---|---|---|---|---|
| LAUGH/AMUSE | S | S | S | S |
| DIE/KILL | L | S | S | S |

| | | | | |
|---|---|---|---|---|
| SIT | S | S | S | S |
| EAT/FEED | S | S | S | S |
| LEARN/TEACH | S | S/L | L | S/R |
| SEE/SHOW | S | S | S | S |
| FEAR/FRIGHTEN | S[18] | S/L | S | S |
| HIDE | R | R | S(/R) | S/R |
| BOIL | S | L | L | S |
| BURN | S/L | L | L | R |
| BREAK | A | L | L | R |
| AWAKE/WAKE | S[19] | S/L[21] | L | S/L |

Table 2: *Valency orientation of the 12 verb pairs in Old English (OE), Middle English (ME), Middle Welsh (ME), and Old Irish (OI). S = suppletion, R = reduction, A = augmentation, L = lability.*

## 4. Conclusions

This leads us to the last set of closing remarks: how do Middle Welsh and Middle English compare? Indeed, the last four verbs, i.e., BOIL, BURN, BREAK, and AWAKE/WAKE, only one of which was labile in Old English, are labile in both languages. The only other verb in the list which is found to be labile in Middle Welsh is *dysgu* 'learn/teach', which is not cognate with any of the verbs in Old Irish meaning either 'learn' or 'teach', as it is a direct loan from Latin; its Middle English equivalent also participates, although not exclusively, in a labile alternation, which is very marginal nonetheless. However, one other verb that was labile in Old English, as we argued above, has been substituted by a suppletive pair.[22]

18. The *Dictionary of Old English* (Cameron et al. 2024) also mentions the possibility that *fyrhtan* 'frighten' glosses lat. *tremere* 'to shake, tremble with fear', making it a non-causal occurrence of the verb. However, since it is not found in our sample and it is only found in that single gloss, we decided not to take it into account and not to add the label L to the cell.
19. García García (2020: 177) mentions *weccan* 'waken, arise, spring (intr.; caus.)' and *wæcnan* 'come into being, be born, spring' among the causative verb pairs in which one or both members become labile in OE. Otherwise both *wacian*, *a-wacian* and *a-wacan* are used intransitively (cf. *An Anglo-Saxon Dictionary Online,* s.vv. *wacian, a-wacian, a-wacan*).
21. Cf. Lewis et al. (1952-2001), s.vv. *awaken, waken*: *awaken* is used only intransitively, whereas *waken* is labile (cf. senses 1-3 vs. sense 4).
22. It has to be stressed that even if *cwelan* and *cwellan* were to be kept apart as different words instead of by-forms as argued here, the suppletive pair found in Middle English would not include the same two verbs as Old English.

It has to be borne in mind that labile pairs in common between Middle English and Middle Welsh need not have developed lability for the same reasons in the two languages, as far as can be argued: of the four, 'burn' has become labile because of a morphological merger between the base form and the derived *-ja* causative, which was already obsolescent in Old English, and 'boil', which also had such a causative in Old English, has become labile because its place has been taken by an Old French loanword (cf. note 6).[23] Obviously, the complexity underlying the parallel development of lability in English and Welsh could only be sketched here, using our selection of verbs as a thematic illustration of the issues at hand regarding valency and, more specifically, lability, and as a first step towards a deeper and larger analysis of its diachrony: therefore, untangling the varied and heterogeneous causes of processes such as this and the complex interplay of internal and external factors will constitute a major objective for the PaVeDa.

## References


Allen, C.L. 2019. *Dative external possessors in early English*. Vol. 39. Oxford: OUP.

Baugh, A.C. & T. Cable. 2013. *A History of the English Language*. London/New York: Routledge.

Borsley, R. D., M. Tallerman & D. Willis. 2007. *The Syntax of Welsh*. Cambridge/New York: CUP.

Bosworth, J. 2014. *An Anglo-Saxon Dictionary Online*. In T. N. Toller, C. Sean & O. Tichy (eds.). Prague: Faculty of Arts, Charles University.

Bromwich, R. & D. S. Evans. 1992. *Culhwch ac Olwen. An Edition and Study of the Oldest Arthurian Tale*. Cardiff: UWP.

Cameron, A., A.C. Amos, S. Butler, A. Healey. 2024. *Dictionary of Old English: A to Le online*.

Evans, D. Simon. 1964. *A Grammar of Middle Welsh*. Dublin: Dublin Institute for Advanced Studies.

Falileyev, A. 2000. *Etymological Glossary of Old Welsh*. Tübingen: Max Niemeyer Verlag.

Filppula, M. & J. Klemola. 2014. Celtic Influences in English: A Re-Evaluation. *Neuphilologische Mitteilungen* 115(1): 33–53.

García García, L. 2020. The basic valency orientation of Old English and the causative *ja*-formation: a synchronic and diachronic approach. *English Language and Linguistics* 24(1): 1–25.


23. Just as English gained some labile verbs by borrowing from Old French, Welsh has in turn been influenced in the same way by English, borrowing new labile verbs, e.g., *stopio* 'stop' (cf. Poppe 2009: 264).

García García, L. and E. R Narbona. 2021. Lability in Old English verbs: Chronological and textual distribution. *Anglia* 139(2): 283-326.
Goetinck, G. W. (ed.). 1976. *Historia Peredur vab Efrawc*. Cardiff: UWP.
GPC = GPC Online. *Geiriadur Prifysgol Cymru*. University of Wales Centre for Advanced Welsh & Celtic Studies: Aberystwyth.
Ingham, R. 2020. How contact with French drove patient-lability in English. *Transactions of the Philological Society* 118(3): 447–467.
Irslinger, B. 2017. Detransitive strategies in Middle Welsh. The preverbal marker *ym*-. In E. Poppe, K. Stuber & P. Widmer (eds.), *Referential properties and their impact on the syntax of Insular Celtic languages*, 101–143. Munster: Nodus Publikationen.
Haspelmath, M. 1993. More on the typology of inchoative/causative verb alternations. In B. Comrie & M. Polinsky (eds.), *Causatives and transitivity*, 87–120. Amsterdam: John Benjamins.
Haspelmath, M. 2001. The European linguistic area: Standard Average European. In M. Haspelmath, E. König, W. Oesterreicher & W. Raible (eds.), *Language typology and language universals* (2.2), 1492–1510. Berlin: de Gruyter.
Haspelmath, M., A. Calude, M. Spagnol, H. Narrog,, E. Bamyaci. 2014. Coding causal-noncausal verb alternations: A form-frequency correspondence explanation. *Journal of linguistics* 50(3): 587–625.
Hickey, R. 2012. Early English and the Celtic hypothesis. In T. Nevalainen & E. C. Traugott (eds.), *The Oxford Handbook of the History of English*, 497–507. Oxford: OUP.
Hickey, R. 2020. English and Celtic: contact-induced change in history. *Linguistics Vanguard* 6(2).
Kemmer, S. 1993. *The Middle Voice*. Amsterdam: John Benjamins.
Kroch, A. & A. Taylor. 2000. *Penn Helsinki Parsed Corpus of Middle English, second edition*.
Lambert, P. 1984. Les commentaires celtiques à Bède le Vénérable (suite). *Études Celtiques* 21: 185–206.
Lewis, R. E. et al. 1952–2001. *Middle English Dictionary*. Online edition in Middle English Compendium. F. McSparran, et al (ed). Ann Arbor: University of Michigan Library, 2000–2018. (Last accessed on 14/04/2025).
Luraghi, S., A. P. Aprosio, C. Zanchi & M. Giuliani. 2024. Introducing PaVeDa – Pavia Verbs Database: Valency Patterns and Pattern Comparison in Ancient Indo-European Languages. In *Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA)*, 79–88. Torino, Italia: ELRA and ICCL.
Manning, H. P. 1995. Fluid intransitivity in Middle Welsh: Gradience, typology and 'unaccusativity'. *Lingua*, 97/2: 171–194.
McMillion, A. 2006. *Labile verbs in English: Their meaning, behavior and structure*. Stockholm: Department of English, Stockholm University.
Müller, N. 1999. *Agents in early Welsh and early Irish*. Oxford: OUP.
Nichols, J. 2017. *Realization of the Causative Alternation: Revised Wordlist and Examples*.

Nichols, J., D. A. Peterson & J. Barnes. Transitivizing and detransitivizing languages. *Linguistic Typology*, 8/2: 149–211.
Ottósson, K. 2013. The anticausative and related categories in the Old Germanic Languages. In F. Josephson, I. Söhrman (ed.). *Diachronic and typological perspectives on verbs*, 329–382. Amsterdam: Benjamins.
Pintzuk, S. & L. Plug. 2002. *The York Helsinki Parsed Corpus of Old English Poetry* (1st edition). Department of Linguistics, University of York.
Poppe, E. 2009. Standard Average European and the Celticity of English intensifiers and reflexives: some considerations and implications. *English Language and Linguistics* 13: 251–266.
Poussa, P. 1990. A contact-universals origins for periphrastic DO with special consideration of old English-Celtic contact. In S. M. Adamson, V. A. Law, N. Vincent, S. Wright (ed.), *Papers of the 5th International Conference on English Historical Linguistics,* 407-34. Amsterdam: Benjamins.
Rodway, S. 2005. The date and authorship of *Culhwch ac Olwen*: a reassessment. *Cambrian Medieval Celtic Studies* 49: 21–44.
Rodway, S. 2007. The where, who, when and why of medieval Welsh prose texts: some methodological considerations. *Studia Celtica* 41: 47–89.
Rodway, S. 2013. *Dating Medieval Welsh Literature: Evidence from the Verbal System*. Aberystwyth: CMCS Publications.
Roma, E. 2021. Valency patterns of Old Irish verbs: finite and non-finite syntax. In S. Luraghi, E. Roma (ed.), *Valency over Time*, 89–132. Berlin: De Gruyter.
Roma, E. 2024. Differential case marking in Old Irish: nominal and pronominal arguments in valency alternations. *Studia Celtica Posnaniensia*, 9: 1–30.
Roma, E. & C. Zanchi. 2025. Old Irish in the PaVeDa database. In D. R. Cooper, R. A. Martin, G. S.D. O'Toole, S. E. Puopolo (eds.), *Proceedings of the Harvard Celtic Colloquium (42),* 212–238. Cambridge: HUP.
Rouveret, A. 2023. *Nonfinite Inquiries: Materials for a Comparative Study of Nonfinite Predicative Domains*. Berlin: Mouton De Gruyter.
Sackmann, R. 2022. Subjects of verbal nouns in Early Modern Welsh: evidence from Perl mewn Adfyd (1595). *Journal of Historical Syntax* 6/11.
Saibene, M. G. & M. Buzzoni. 2014. *Manuale di linguistica germanica*. Milano: Cisalpino-Monduzzi.
Schumacher, S. 2011. Mittel- und Fruhneukymrisch. In E. Ternes (Hrsg.), *Brythonic Celtic – Britannisches Keltisch. From Medieval British to Modern Breton*. Bremen: Hempen Verlag.
Taylor, A., A. Warner, S. Pintzuk & F. Beths. 2003. *The York Toronto Helsinki Parsed Corpus of Old English Prose* (*YCOE*). Department of Linguistics, University of York. Oxford Text Archive, 1st edition.
TTL = *Thesaurus linguae Latinae*. 1900-current. Berlin (formerly Leipzig): De Gruyter (formerly Teubner).
van Gelderen, E. 2011. Valency changes in the history of English. *Journal of Historical Linguistics* 1(1): 106–43.
van Gelderen, E. 2014. Changes in Psych-verbs: A reanalysis of little v. *Catalan Journal of Linguistics* 13: 99–122.

van Gelderen, E. 2018. *The Diachrony of Verb Meaning: Aspect and Argument Structure*. New York: Routledge.
Vezzosi, L. 2005. The development of himself in Middle English: a 'Celtic' hypothesis. *STUDIES IN ENGLISH MEDIEVAL LANGUAGE AND LITERATURE*, 228–243.
Visser, F. T. 1963. *A historical syntax of the English language*. Leiden: Brill.
Williams, I. 1930. *Pedeir Keinc y Mabinogi*. Caerdydd: Gwasg Prifysgol Cymru.
Zanchi, C., M. Tarsi. 2021. Valency patterns and alternations in Gothic. In S. Luraghi & E. Roma (eds.), *Valency over Time.* 31–88. Berlin/Boston: De Gruyter Mouton.
Zanchi, C., C. R. Combei & S. Luraghi. 2022. PaVeDa - Pavia Verbs Database: Challenges and Perspectives. In *Proceedings of the 4th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*, 99–102. Seattle/Washington: Association for Computational Linguistics.
Zimmermann, R. 2014. *The Parsed Corpus of Middle English Poetry* (*PCMEP*).
Zimmermann, R. 2024. How *fear* developed from an object to a subject experiencer verb: Remarks on argument structure change. *Journal of Historical Linguistics*.
Zúñiga, F. & S. Kittilä. 2019. *Grammatical Voice*. Cambridge: CUP

# Valency Patterns and Alternations of Experiential Predicates in Vedic Sanskrit

*Erica Biagetti*[*]
*Leonardo Montesi*[**]

## 1. Introduction

In this paper, we analyze valency patterns and alternations of experiential predicates in Vedic Sanskrit adopting the PaVeDa methodology (Zanchi et al. 2022).

Vedic Sanskrit, or Vedic, which can be roughly identified with Old Indo-Aryan, is the earliest attested language of the Indo-Aryan group of the Indo-European language family and one of the most ancient Indo-European languages attested. As a reference corpus for our study, we use the *Ṛgveda* (ṚV), whose language can approximately be dated to the second half of the second millennium BC and thus belongs to Early Vedic.

Occasionally, we also resort to the *Atharvaveda* (AV) (Early Middle Vedic; cf. Dahl 2015: 43) as well as later texts (Middle and Late Vedic). As sources for meanings and examples, we employ Grassmann's (1873) reference dictionary, the research platform VedaWeb (https://vedaweb.uni-koeln.de) for examples from the ṚV, and the *Digital Corpus of Sanskrit* (http://www.sanskrit-linguistics.org/dcs/index.php?contents=texte) for later texts.

For this paper, we focus on the analysis of 16 meanings expressing experiential situations and the verbs that lexicalize them. Of the 16 meanings, 15 belong to the 80 core meanings selected by developers of the ValPal project, whereas 1 meaning, APPEAR, is a later addition from the same project. Table 1 reports the analyzed meanings and verbs divided on the base of the experiential situation that they encode:

[*] University of Pavia
[**] University G. Marconi

| Experiential situation | Meaning | Vedic verb(s) |
|---|---|---|
| Bodily sensations | BE HUNGRY | *kṣudh-* |
| | FEEL COLD | *śīta- as-* |
| | FEEL PAIN | *tap-* |
| Perception | HEAR / LISTEN | *śru-* |
| | APPEAR | *dr̥ś-* |
| | SEE | *paś-* |
| | | *dr̥ś-* |
| | LOOK AT | PV *cakṣ-* |
| | | PV *paś-* |
| | | PV *khyā-* |
| | SHOW | *diś-* |
| | SMELL | *ghrā-* |
| | TOUCH | *(úpa) spr̥ś-* |
| Cognition | KNOW | *vid-* |
| | THINK | *man-* |
| Emotions | LIKE | *juṣ-* |
| | | *ran-* |
| | BE SAD | *tap-* |
| | FEAR | *bhī-* |
| | FRIGHTEN | *ví bhī-* |

Table 3. *Experiential situations, meanings and verbs analyzed in this study.*

## 2. Methodological considerations

Before moving on to the analysis of valency patterns and alternations of experiential predicates in Vedic, a few methodological considerations are in order. These are presented in Sections 2.1 to 2.4 and regard the choice of examples in our corpus, as well as matters related to the description of alternations.

### *2.1. Different meanings expressed by alternations*

Several of the meanings selected by the ValPal project and inherited by PaVeDa are expressed in Vedic by alternations of the same verb, rather than by different lexemes. For example, the meanings APPEAR, SEE, and SHOW can be expressed in Vedic by mediopassive, active, and causative forms of *dr̥ś-*, respectively (the latter only in post R̥gvedic Vedic, so for our

dataset we chose *diś-*). Although the meaning APPEAR belongs to ValPal's additional meanings, and not to the 80 meanings listed in the questionnaire, we decided to include this meaning in our database because mediopassive forms of *dṛś-* (APPEAR) are much more frequent than active ones (SEE, in a relation of present-aorist-perfect suppletivism with *paś-*), and it thus did not seem accurate to list the former as alternations of the latter.

## *2.2. Compounded verbs*

In Vedic, preverb-verb combinations still seem to operate at the syntactic rather than the lexical level.

Syntactically, preverbs (PV) exhibit a high degree of syntactic freedom as to their positioning, which however has no effect on the whole meaning of the composite (Danesi 2013: 62). Indeed, semantic shifts from their basic spatial meaning do not require univerbation, and correspondently univerbation does not imply either semantic shift from the basic spatial meaning or lexicalization. However, meanings of PV-verb pairs are sometimes non-compositional, which points to an increasing degree of lexicalization. Furthermore, changes in the lexical aspect of the verb, bringing about notions such as ingressivity or completion, points to an incipient grammaticalization. Finally, PVs can function as applicatives in certain contexts, which also points to ongoing grammaticalization (Kulikov 2012, for an overview, see Zanchi 2019: 95-117).

Due to their low degree of lexicalization at this stage, Grassmann's dictionary lists PV-verb pairs under the verbal root. In our database, we include PV-verb pairs if their meaning is close to the one of the bare root (1), but exclude them from the analysis in cases where their meaning is non-compositional (2):

(1) *pári* 'around' + *khyā-* 'see/look' > 'look around'
*mahás* *putrā́so ...* *urviyā́* ***pári***
great.GEN son.NOM.PL widely PV
***khyan***
look.INJ.AOR.3PL.ACT
'The sons of the great one … **look around** widely.' (ṚV 10.10.2cd)

(2) *pári* + *khyā-* > 'overlook'
*mā* *índro* *naḥ* *víṣṇur*
NEG Indra.NOM 1PL.ACC Viṣṇu.NOM
*marútaḥ* ***pári*** ***khyan***
Marut.NOM.PL PV look.INJ.AOR.3PL.ACT
'Let Indra, Viṣṇu, and the Maruts not **overlook** us.' (ṚV 7.93.8c)

For verbs that are compounded with more than one PV resulting in similar meanings, we initially included the most frequent PV in the lemma entry and mentioned other attested PVs in the notes. Since, however, some alternations that are attested with one PV may not be attested with others, we felt that including a single PV in the lemma entry was misleading in this respect. For instance, the factitive alternation (see Section 4.1) is attested with *prá cakṣ-* and not with *abhí cakṣ-*, which we had initially chosen as the lemma entry in our database:

(3) ***prá cakṣaya*** *ródasī*
PV see.FACT.IMPV.2SG.ACT world-half(F).ACC.DU
'Reveal the two world-halves.' (ṚV 1.134.3f)

Following these considerations, for cases where a meaning is expressed by the combination of a verb and one of a series of PVs, we decided to include a generic "PV" in the lemma entry and list the attested PVs in the notes (Table 2).

| **Lemma** | **Coding frame** | **Notes-lemma** |
|---|---|---|
| PV *cakṣ-* | 1-nom V.act.subj[1] 2-acc | PV = *abhí/ví/práti/prá/áva cakṣ-* |
| PV *paś-* | 1-nom V.act.subj[1] 2-acc | PV = *pári/abhí/ánu/áva paś-*; 'look at' also with a simplex verb in the imperative |
| PV *khyā-* | 1-nom V.act.subj[1] 2-acc | PV = *ví/práti/abhí/pári/áva khyā-* |

Table 4. *Annotation of PV-verb pairs for LOOK AT.*

## *2.3. Finite vs. non-finite forms*

Non-finite verb forms sometimes feature idiosyncratic behaviors. For instance, with PV *cakṣ-* (LOOK AT), p-lability is only attested in the infinitive:

(4) *ná* *ta* *indra* *sumatáyo*
NEG 2SG.GEN Indra.VOC favor(F).NOM.PL
*ná* *rā́yaḥ* ***saṁcákṣe***
NEG good.NOM.PL survey.INF
'Neither your favors nor your riches, o Indra, can be entirely surveyed.' (ṚV 7.18.20ab)

To make but another example, object omission is more frequent with participles, as they present background information; for instance, out of 115 participial forms of *vid-* (KNOW), 77 lack an object. For these reasons, we decided to exclude non-finite forms from the analysis. However, this rule does not apply when non-finite forms are the only attested ones for one verb, as in the case of *kṣudh-* (BE HUNGRY):

(5) *índra* ***kṣúdhyadbhyo***
Indra.VOC be.hungry.PTCP.DAT.PL.ACT
*váya* *āsutím̐* *dāḥ*
food(N).ACC drink(F).ACC give.AOR.INJ.2SG.ACT
'Give, o Indra, food and drink to (us) in hunger' (ṚV 1.104.7d)

## *2.4. Describing alternations*

In some cases, our verbs are subjected to morphological alternations which are however not paired with changes in their valency. For example, the verb *juṣ-* (LIKE) occurs 140 in the middle voice and 35 in the active, but no difference in meaning can be observed between the two forms, as shown by examples (6) and (7), where the verb is employed in its basic coding frame 1-nom V.subj[1] 2-acc:

(6) *gíro* ***juṣethām*** *adhvarám̐*
hymn(F).ACC.PL enjoy.IMPV.AOR.2DU.MID rite.ACC
***juṣethām̐***
enjoy.IMPV.AOR.2DU.MID
'Enjoy the hymns; enjoy the rite.' (ṚV 8.35.6a)

(7) ***jóṣi*** *bráhma* *jányam̐*
enjoy.IMPV.2SG.ACT formulation(N).ACC of.people.ACC.N
***jóṣi*** *suṣṭutím*
enjoy.IMPV.2SG.ACT lovely.praise(F).ACC
'Enjoy the sacred formulation stemming from your people; enjoy the lovely praise.' (ṚV 2.37.6ab)

A similar situation is found for the other verb of positive emotion, *ran-*, but with the opposite ratio between the two voices: the active is the most frequent one, occurring 28 out of 46 times. In such cases, we include both voices in the basic coding frame and in alternations that are not coded on the verb. Accordingly, the basic coding frame for *juṣ-* is 1-nom

V.mid/act.subj[1] 2-acc, whereas the one for *ran-* is 1-nom V.act/mid.subj[1] 2-loc(/acc).

## 3. Valency patterns and alternations of experiential predicates

The domain of experience encompasses diverse situations including bodily sensations, perception, cognition, emotions, and volition. Participants in experiential situations exhibit variable properties based on conceptualization of individual situations or on different conceptualizations of the same situation, making their encoding more complex than less polymorphic participants (Luraghi 2020: 2). For this reason, experiential verbs have received much attention in recent research and have been the subject of both language-specific and cross-linguistic studies. In Vedic studies, experiential predicates have been treated in studies on argument structure constructions (Dahl 2009, 2014, Dahl and Fedriani 2012, Luraghi, Caviglia and Pinelli 2014), on verbal morphology (Jamison 1983, Kümmel 1996, 2000, Kulikov 2006, 2013, Kulikov and Lavidas 2013), and valency-changing operations (Kulikov 2009, 2012), among others.

Following Verhoeven (2007), we define experiential situations as involving an experiencer and an expertum (the content being experienced), as in *John is thirsty*. Most experiential situations also include a stimulus that triggers the experience, as in *John loves Yōko*. The experience may focus on specific parts of the experiencer, whether physical as in *I feel pain in my back*, or immaterial, as in *He was meditating a plan in his mind* (cf. also Luraghi 2020: 1-2).

Being necessarily sentient and hence animate, experiencers share features with agents. However, since they undergo rather than initiate states of affairs, experiencers also share features of patients, especially in certain types of experiences (e.g. bodily sensations). Stimuli show greater referential variation than experiencers, being either animate or inanimate, and can often be states of affairs themselves (Verhoeven 2007: 53-69).

Most experiential verbs in Vedic function as two-place predicates with experiencer subjects, while stimuli appear as second arguments marked by various cases. Indeed, Vedic manifests a strong preference for nominative alignment in encoding experiential situations. However, since speakers can construe the same situation in different ways (Croft 2012: 13, Langacker 1999: 206-212), various alternations to the basic coding frame are attested for several of the verbs under analysis. In the following sections, we will present the basic coding frames and their alternations as attested in the ṚV.

### *3.1. Bodily sensations*

The subdomain of bodily sensations encompasses feelings of body temperature, physical needs such as feeling hunger or thirst, as well as more general states such as feeling well or feeling bad. This subdomain provides the source for an embodied conception of other subdomains of experience, as often bodily reactions are metonymically understood as standing for an experience (Luraghi 2020: 23-24). For instance, body temperature is often mapped onto the subdomain of emotions, a phenomenon that has been extensively analyzed in Cognitive Linguistics research (Lakoff and Johnson 1980, Kövecses 1995, Radden 2000).

In PaVeDa, meanings associated with bodily sensations are BE HUNGRY, FEEL COLD, and FEEL PAIN. In the ṚV, we only find one attestation of BE HUNGRY expressed by the root *kṣudh-* (5). *Śītā- (as-)* 'be cold' is also attested once; however, it refers to gambling dice and thus expresses touch temperature rather than personal feeling temperature (on the linguistics of temperature, see Koptjevskaja-Tamm 2015). The general lack of such predicates might be ascribed to the nature of Ṛgvedic hymns that, being dedicated to gods, are probably less interested in bodily human sensations such as hunger or pain.

### *3.2. Perception*

Perceptive situations are related to the intake of external information through the five sense modalities: sight, hearing, touch, taste, and smell. In PaVeDa, the meanings HEAR, LISTEN, SEE, LOOK AT, SMELL and TOUCH refer to perceptive situations. To these, APPEAR can be added, as the anticausative counterpart of SEE and LOOK AT, and SHOW as the causative ('make see'). However, despite the verb for SHOW, *diś-*, originally meaning 'point at' and therefore 'show' in the sense of 'make see', in the ṚV it is attested only with abstract stimuli and thus takes the meaning 'make known'. For this reason, SHOW is discussed together with verbs of cognition rather than perception.

In the ṚV, perception verbs are unevenly attested across perception modalities: while verbs that indicate sight and hearing are frequent, other modalities are scarcely if ever referred to. In particular, smell is attested only from the AV onward, and anyway rarely. Sight is the most represented modality for the quantity of verbs that encode it (*paś-*, *dṛś-*, *cakṣ-*, and *khyā-* are those analyzed), whereas hearing, always expressed by the verb *śru-*, is the most frequent (*śru-* occurs 349 times in the ṚV) and the most varied with respect to the alternations that it attests. The lexicalization of different event

types in the area of sight and hearing in the ṚV is in accordance with Viberg's (1984) universals of lexicalization, as a distinction between SEE and LOOK AT is, at least in part, indicated by different lexemes, whereas the distinction between HEAR and LISTEN TO appears to be conveyed only by mood variation (imperative vs. other moods). More specifically, SEE tends to be expressed by *paś-* and *dṛ́ś-* without preverbs, while LOOK AT, by *cakṣ-*, *paś-*, and *khyā-* accompanied by preverbs such as *abhí* 'to, onto, against', *ánu* 'after', *práti* 'in reverse direction, against', and *ví* 'apart, asunder, away' which confer directionality, and hence control to the situation. Finally, APPEAR is expressed by non-active forms of *dṛ́ś-* and is very well attested, being referred especially to the appearance of dawn.

Table 3 shows the basic coding frames of all perception verbs taken into analysis:

| **Meaning** | **Verb** | **Basic coding frame** |
|---|---|---|
| APPEAR | *dṛ́ś-* | 1-nom V.mid/pass.subj[1] |
| SEE | *paś-* | 1-nom V.act.subj[1] 2-acc |
| | *dṛ́ś-* | 1-nom V.act.subj[1] 2-acc |
| LOOK AT | PV *cakṣ-* | 1-nom V.mid.subj[1] 2-acc |
| | PV *paś-* | 1-nom V.act.subj[1] 2-acc |
| | PV *khyā-* | 1-nom V.act.subj[1] 2-acc |
| HEAR / LISTEN | *śru-* | 1-nom V.act.subj[1] 2-acc/gen |
| TOUCH | *spṛ́ś-* | 1-nom V.subj[1] 2-acc (3-instr) |
| SMELL | *ghrā-* (post ṚV) | 1-nom V.act.subj[1] 2-acc(-cognate) |

Table 5. *Basic coding frames of perception verbs in PaVeDa.*

As expected, the only monovalent verb is *dṛ́ś-* in the meaning APPEAR. In this meaning, *dṛ́ś-* occurs in middle-passive forms and its only argument is not an experiencer but a theme, as in (8). Active forms of *dṛ́ś-* (SEE) take a nominative experiencer and an accusative stimulus (9). In this, *dṛ́ś-* assimilates not only to *paś-*, with which it is in a relationship of complementarity with respect to lexical aspect,[1] but also to the other sight verbs, which consistently take an accusative stimulus. Differently from other sight verbs, *cakṣ-* feature middle morphology; rarely, it occurs in the active voice, with no changes in meaning or construction, but in two cases (ṚV 5.2.8c and 10.32.6c) the intensive stem may signal a higher degree of control.

1. On the fixed present-aorist-perfect suppletivism involving *paś-* and *dṛ́ś-* (*páśyati-dárśam-dadárśa*) see Casaretto (2002).

(8) *eṣā́* *divó* *duhitā́* *praty*
DEM.NOM.F heaven.GEN daughter(F).NOM PV
***adarśi***
appear.AOR.3SG.MID
'This Daughter of Heaven (Dawn) has appeared opposite.' (ṚV 1.124.3a)

(9) *utá* *tvaḥ* *páśyan* *ná* ***dadarśa***
CONJ INDEF.NOM see.PTCP.NOM NEG see.PF.3SG
*vā́cam*
Speech(F).ACC
'And many a one who sees has not seen Speech.' (ṚV 10.71.4a)

The verb *śru-* most frequently indicates the spontaneous event of hearing, but can also refer to the controlled activity of listening, without being combined with preverbs. In its basic coding frame, *śru-* also takes two arguments. Contrary to verbs of sight, however, *śru-* features construction variation based on the animacy of the stimulus: there is a clear tendency for the accusative to occur when the stimulus is inanimate, whereas the genitive is virtually restricted to animate stimuli (Dahl 2014). Touch is the third perceptual modality in Viberg's (1984) lexicalization hierarchy. In most of its occurrences, the verb *(úpa) spṛś-* 'touch' cannot be considered an experiential verb, as it typically takes the object *dyā́m* 'heaven', *diváḥ sā́nu* 'back of heaven', or *nā́kam* '(heaven's) vault', and thus means 'reach' in height, width, power, etc. An occurrence in which it may be taken to mean 'feel by touching' is (10), from a healing spell, where hands and fingers are explicitly mentioned:

(10) *hástābhyāṁ* *dáśa-śākhābhyāṁ* *tvā* ***úpa***
hand.INST.DU ten-branch.INST.DU 2SG.ACC PV
***spṛśāmasi***
touch.1PL.ACT
'With two hands with their ten branches (= fingers) … we touch you.' (ṚV 10.137.7)

The second argument of *spṛś-* alternates the accusative with the locative. With this verb, however, construction variation implies a different construal of the situation. In two passages (ṚV 4.41.1cd, reported in (11), and ṚV 10.91.13cd,), the subject is inanimate and abstract (the praise songs), and the locative second argument is the heart (*hṛdí*). Thus, the verb takes the metaphorical meaning of touching someone emotionally. In another passage, the stimulus is *tanvì śrutásya* 'the body of the famous one (Indra)' and the

instrument *dhiyā́* 'with an insightful thought' suggests that the meaning is also 'touch emotionally'.

(11) *yó* *vā́m* *hṛ́dí ...* *asmád*
REL.NOM.PL 2DU.GEN heart(N).ACC 1PL.ABL
*uktáḥ* ***paspárśad***
speak.PPP.NOM.PL touch.SUBJ.PF.3SG.ACT
*indrā-varuṇā*
Indra-Vāruṇa.VOC.DU
'(The praise song) spoken by us, which ... will touch your heart, o Indra and Varuṇa.' (ṚV 4.41.1cd)

Finally, the verb *ghrā-* 'smell' is only attested in post-Ṛgvedic texts, where it usually takes an accusative stimulus or a cognate object such as *gandha-* 'odour' (12):

(12) *yena* *vā* ***gandhā́ñ*** ***jighrati***
REL.INST or odour.ACC.PL smell.3SG.ACT
'(Which of these is the self?) Is it that by which one smells odors?' (AU 3.1.1)

### *3.3. Cognition*

In PaVeDa, the meanings KNOW and THINK pertain to cognition. As discussed in Section 3.2, we include SHOW in this Section, as *diś-* always means 'make known' in the ṚV, rather than 'make see'. Table 4 shows the basic coding frames of verbs of cognition taken into analysis:

| **Meaning** | **Verb** | **Basic coding frame** |
|---|---|---|
| KNOW | *vid-* | 1-nom V.act.subj[1] 2-acc/gen |
| SHOW | *diś-* | 1-nom V.mid.subj[1] 2-acc (3-acc/dat) |
| THINK | *man-* | 1-nom V.mid.subj[1] 2-acc/gen |

Table 6. *Basic coding frames of verbs of cognition in PaVeDa.*

The meaning KNOW is expressed by the root *vid-*, and more specifically by its perfect *veda*. This reflects PIE **uóid-h₂e*, originally the perfect of PIE **ueid-* 'see, look' and must have originally meant 'have seen', hence, by pragmatic inference, 'know from having seen'. In its basic coding frame, *veda* features a nominative experiencer and an accusative or genitive

stimulus. Contrary to what we have seen for *śru-*, construction variation is not based on the animacy of the stimulus, and no satisfactory explanation for it has been fund so far (see Dahl 2009 for a discussion). With Ancient Greek *oîda*, which is also a reflex of PIE **uóid-h₂e*, construction variation triggers different meanings, with NomGen limited to practical knowledge and skills (Luraghi 2020: 169-175). No clear distribution can be observed for *veda*, although practical knowledge tends to be expressed by the accusative (13), whereas the meanings 'take heed' and 'witness' are virtually restricted to genitive stimuli (14). Thus, intentionality and control seem to play a role in construction variation.

(13) ***vetthā*** *hí* ***nírr̥tīnāṁ***
know.PF.2SG.ACT PTC ruin.GEN.PL
*vájrahasta* ***parivŕ̥jam***
strength.in.hand.VOC disposal.ACC
*áhar-ahaḥ* *śundhyúḥ* *paripádām* *iva*
day-day preener.NOM snare.GEN.PL like
'For you know how to avoid calamities day after day, o you with mace in hand, as a preener does snares.' (Ṛ̊V 8.24.24)

(14) ***véda*** *nāv* ***asyá*** *pr̥thivī́*
know.PF.3SG.ACT 1PL.GEN DEM.GEN earth.NOM
*utá* *dyaúḥ*
and heaven.NOM
'[(Even) in the womb the Begetter made us two a married couple...] Heaven and Earth take heed of this about us.' (Ṛ̊V 10.10.5)

Somewhat unexpectedly for a caused situation such as SHOW, *diś-* also takes two arguments, namely a nominative causer and an accusative stimulus, whereas the experiencer is left unexpressed. The lack of an explicit experiencer and the use of the middle voice suggests that, in its basic coding frame, the causer is conceived as less agentive than in a ditransitive construction: this is confirmed by example (15), where the god Indra shows (i.e., makes known) his virility (*paúṁsyam*) not as a direct act of pointing or communicating, as in PaVeDa's example *The girls showed pictures to the teacher*, but indirectly by defeating the demon Vr̥tra.

(15) *vr̥trahā́* ***adiṣṭa*** *paúṁsyam*
Vr̥tra.smiter.NOM show.AOR.3SG.MID virility(N).ACC
'The Vr̥tra-smiter displayed his masculine nature.' (Ṛ̊V 8.93.15b)

This analysis is validated by two occurrences where the verb is in the active voice and the experiencer is explicitly mentioned (once in the accusative and once in the dative). In these cases, the subject is agentively showing something to the experiencer, as further highlighted by the intensive stem *dediśam* in one of the two occurrences, presented in (16).

(16) *satyám* *ít* *tvā* *mahenadi*
true.ACC.N PTCL 2SG.ACC great.river.VOC
*páruṣṇy* ***áva*** ***dediśam***
Paruṣṇi.voc PV show.INTENS.SBJV.1SG.ACT
'This is really true – what I forcefully point out to you, o great river Paruṣṇi: [o waters, there exists no mortal who is a greater giver of horses than most powerful (Śrutarvan).]' (ṚV 8.74.15ab)

Patterns of polysemy attested cross-linguistically point toward a disparate variety of construal for THINK (see, e.g., Goddard 2003). The root *man-* constitutes no exception, as it can take several meanings and even border with the more complex concepts of 'remembrance' (Wierzbicka 2007). In its basic coding frame, *man-* takes two arguments and middle morphology. Within this construction, it means 'think' (17), 'recall' (18), or 'conceive (typically a praise song)' (19). In a few cases, the stimulus is in the genitive but, as in the case of *vid-*, an explanation for such alternation has not yet been worked out (cf. Dahl 2009). *Man-* never indicates a state, and accordingly never occurs in the perfect aspect;[2] rather, it indicates either an achievement or an activity, and can thus be found in the aorist and in the present.

(17) *tád* *rā́dho* *adyá* *savitúr*
DET.ACC.N largesse(N).ACC today Savitar.GEN
*váreṇyaṁ* *vayáṁ ...* ***manāmahe***
desirable.ACC.N 1PL.NOM think.SBJV.AOR.1PL.MID
'The desirable largesse of Savitar shall we think upon today.' (ṚV 1.159.5ab)

(18) ***mánai*** *nú* *babhrū́ṇām*
think.SBJV.AOR.1SG.MID now brown.GEN.PL.F
*aháṃ* *śatáṃ* *dhā́māni*
1SG.NOM one.hundred.ACC.PL.N form(N).ACC.PL
*saptá* *ca*
seven.ACC.PL CONJ

2. Two perfect forms of *man-* are attested: one with the preverb *ánu*, meaning 'yield to, concede' (ṚV 7.31.7), and one with the preverb *pári*, meaning 'disregard' (ṚV 7.93.6).

'I shall now call to mind the one hundred and seven forms of those brown (plants).' (ṚV 10.97.1cd)

(19) *agné* *stómam* ***manāmahe***
Agni.GEN praise.song.ACC think.SBJV.AOR.1PL.MID
'For Agni we will conceive a praise song.' (ṚV 5.13.2a)+

When it lacks a direct object, *man-* functions as an incise (see, e.g., ṚV 10.179.3a), whereas with a predicative object it means 'think something of someone, consider' (20).

(20) ***mánye*** *tvā* *yajñíyaṁ*
think.1SG.MID 2SG.ACC worthy.of.sacrifice.ACC
*yajñíyānām*
worthy.of.sacrifice.GEN.PL
'I consider you the most worthy of the sacrifice of those worthy of the sacrifice.' (ṚV 8.96.4a)

### *3.4. Emotions*

Emotions represent the widest and most diverse sub-domain of human experience. They are often categorized as positive or negative, primarily based on the experiencer's attitude toward the stimulus.

In PaVeDa, meanings belonging to the domain of emotions are BE SAD, FEAR, FRIGHTEN, and LIKE. Table 5 shows the basic coding frames of emotion verbs taken into analysis:

| **Meaning** | **Verb** | **Basic coding frame** |
|---|---|---|
| BE SAD | *tap-* | 1-nom V.mid.subj[1] |
| FEAR | *bhī-* | 1-nom V.act/mid.subj[1] 2-abl |
| FRIGHTEN | *ví bhī-* | 1-nom V.act.subj[1] 2-acc |
| LIKE | *juṣ-* | 1-nom V.mid/acc.subj[1] 2-acc |
| | *ran-* | 1-nom V.act/mid.subj[1] 2-loc(/acc) |

Table 7. *Basic coding frames of verbs expressing emotions in PaVeDa.*

The root *tap-* is well attested in the ṚV with its literal meaning 'burn'; BE SAD is a metaphorical extension of the literal meaning and is attested three times in the text, while it will become more common in later stages (on Epic Sanskrit, see De Rossi 2023). Note that, in later texts, *tap-* also occurs in the meaning FEEL PAIN, in this case a metonymical extension from the domain

of temperature (De Rossi 2023). In the ṚV, *tap-* (BE SAD) is found in the active and middle voices, and in both cases it takes a nominative experiencer (21). Furthermore, *tap-* is attested in the passive voice, as in (22), where the subject *hṛ́dayaṃ* 'heart' also suggests reference to an experiential situation. From the *Atharvaveda,* the root is also attested in the factitive alternation (see Section 4.1).

(21) *jāyā́* ***tapyate*** *kitavásya* *hīnā́*
wife.NOM be.sad.3SG.MID gambler.GEN leave.PPP.NOM
'The wife of the gambler, abandoned, is scorched.' (ṚV 10.34.10a)

(22) *hṛ́dayaṃ* ***tapyate*** *me*
heart(N).NOM be.sad.3SG.MID 1SG.GEN
'My heart is scorched.' (RV 10.95.17d)

The root *bhī-* (FEAR) can take one or two arguments. When *bhī-* takes only a nominative experiencer (23), the emotion is conceived as a non-dynamic, durative state, and can take the active or the middle voice:

(23) *ká* *īṣate* *tujyáte* *kó*
who.NOM flee.3SG.MID suffer.3SG.MID who.NOM
***bībhayat***
fear.PF.3SG.ACT
'Who flees? Who suffers? Who fears?' (ṚV 1.84.17a)

Alternatively, *bhī-* can take an overt stimulus, that is construed as a cause and thus takes the ablative:

(24) ***bháyante*** *víśvā* *bhúvanā*
fear.3PL.MID all.NOM.PL.N creature(N).NOM.PL
*marúdbhyo*
Marut.ABL.PL
'Every creature fears the Maruts' (ṚV 1.85.8c)

In one case (25), *bhī-* is compounded with the PV *ví* 'apart, asunder, away' and takes an accusative second argument that is not a stimulus, but rather an experiencer, whereas the stimulus is in the instrumental.[3] As a consequence, the meaning is not FEAR but FRIGHTEN:

3. Interestingly, with other roots *ví* serves as a marker of reciprocity, a function that Kulikov (2012) defines as 'intransitivizing'; instead, when *ví* is compounded with *bhī-*, the valency of the root increases.

(25) *ná* *vépasā* *ná* *tanyata*
NEG excitation(N).INST NEG thundering(F).INST
*índram* *vr̥tró* ***ví*** ***bībhayat***
Indra.ACC Vr̥tra.NOM PV fear.AOR.INJ.3SG.ACT
'Not by his trembling excitation, not by his thundering did Vr̥tra make Indra fear.' (Ṛ̊V 1.80.12b)

The only positive emotion included in PaVeDa is LIKE. In Early Vedic, this meaning is expressed by the roots *juṣ-*, *ran-*, *mad-* and *kam-*, of which only the first two are analyzed here (on *mad-* and *kam-* see Dahl 2014). In its basic coding frame, *juṣ-* takes an accusative stimulus and middle morphology; a few active forms are attested without observable changes in meaning (see Section 2.4). Typically, the gods are the experiencers of the emotion, while hymns, praise songs, oblations, and sacrifices offered to them are the stimuli. Once (26), the stimulus is the singer (*jaritṛ́-*) and, as is often the case for animate stimuli, it takes the genitive (cf. *śru-* in Section 3.2).

(26) *yác* *chasyáse* *dyúbhir* *aktó*
when praise.2SG.PASS day.INST.PL anointed.PPP.NOM
*vácobhis* *táj* ***juṣasva***
word(N).INST.PL then enjoy.IMPV.AOR.2SG.MID
*jaritúr*
singer.GEN
'When you are praised through the days, anointed with words, then take pleasure in the singer.' (Ṛ̊V 6.5.6cd)

Apart from one instrumental stimulus, four accusative ones, and a few cases of generic argument omission, *ran-* consistently takes a locative stimulus; in one case (Ṛ̊V 9.111.2e), we find a correlative adverb indicating place (*yátra* 'where') instead of a nominal stimulus. According to Dahl (2014: 38-41), the locative is associated with a definite interpretation, whereas the instrumental, which in the case of *mad-* and *kam-* alternates with the locative, seems to be underspecified with respect to definiteness.

## 4. Alternations

In this Section, we will discuss alternations featured by experiential predicates. Following a tradition established by the ValPal project and inherited by PaVeDa, we will distinguish between alternations that are coded on the

verb and those that are not. For each alternation, we will specify whether it decreases (D), augments (A), or rearranges (R) the valency of the verb, or whether it identifies two elements (I):

### *4.1 Coded alternations*

Coded alternations featured by experiential verbs are summarized in Table 6.

| **Comparative concept** | **Language-specific alternation** | **Derived coding frame** | **D/A/R /I** |
|---|---|---|---|
| Voice alternation | Voice alternation – selfbenefactive | 1=3-nom midV'.subj[1=3] 2-acc | I |
| Passive | Passive | 2-nom passV'.subj[2] (1-inst) | D |
| Causative | Factitive | 4-nom -*áya*-V'.subj[4] 2-acc | A/R |
| | Causative | 4-nom -*áya*-V'.subj[4] 1-acc | A/R |
| | Causative with instrument causee | 4-nom -*áya*-V'.subj[4] 3-acc | A/R |

Table 8. *Coded alternation displayed by experiential verbs.*

The selfbenefactive alternation is marked by continuants of the PIE mediopassive. In this alternation, middle morphology has the effect to add a beneficiary adjunct and simultaneously establish a coreference relationship with the subject. Thus, the event structure of the verb is augmented by one participant, but its valency remains unchanged.

The passive alternation is realized by three types of passive formations, one for each of the three main tense systems (Kulikov 2006): i) present passives with the accented suffix -*yá*-, ii) (medio-)passive aorists, ii) statives derived from either present or perfect stems. Formations ii) and iii) have a defective paradigm, being attested primarily in the 3rd person singular (aorists in -*i*, statives in -*e*) and plural (aorists in -*ran/-ram*, statives in -*re*). Apart from rare cases where an oblique experiencer is explicitly expressed, passives of experiential verbs are typically agentless passives (Kulikov 2011) or even anticausatives. Since, however, the one between agentless passive and anticausative is often a very subtle distinction, we treat the three categories as a single alternation. Examples (27)-(29) with the verb *śru*- illustrate the distinction between passive, agentless passive, and anticausative: while in (27) we have an instrumental experiencer, in (28) the experiencer is not expressed, and in (29) it is not even implied by the event structure, and the verb is rendered as 'be famed'. The difference between (28) and (29) lies in the

semantic traits of the stimulus, being a sound in (28), and a god in (29) (Kulikov 2011: 237).

(27) *sáṁ* *ghóṣaḥ* ***śr̥ṇve*** *avamaír*
PV battle.cry.NOM hear.STAT.3SG.MID closest.INST.PL
*amítrair*
foe.INST.PL
'All around your battle-cry is heard by the closest foes.' (Ṛ̊V 3.30.16a)

(28) ***śr̥ṇvé*** *yā́meṣu* *saṁtaníḥ*
hear.STAT.3SG.MID journey.LOC.PL thundering.NOM
'Its thundering is heard on its journeys.' (Ṛ̊V 5.73.7b)

(29) *kásminn* *adyá* *jáne* *mitró*
INTER.LOC today people.LOC ally.NOM
*ná* ***śrūyate***
like hear.3SG.PASS
'Among what people is he (Indra) famed today like an ally among the people?' (Ṛ̊V 10.22.1ab)

In Early Vedic, we find three alternations that have the effect of adding a causer as a new role. All are formed with the addition of the suffix *-áya-* and are called Factitive, Causative proper, and Causative with instrument causee. The Factitive alternation, for which we adopt a term proposed by Lubotsky (1989; cf. also Thieme 1929:22, Insler 1968, Kortlandt 1981:127), applies to intransitive verbs and to intransitive/transitive verbs, that is, verbs that can be constructed both transitively and intransitively, including also verbs of perception and cognition that features an alternation in the encoding of the object (Jamison 1983). The meaning of the Factitive alternation is 'to accomplish that somebody/something achieves some state'. The resulting coding frame consists of two arguments, but a causer role is added that is absent in the transitive construction, while the experiencer is often left unexpressed. The difference between transitives and factitives can be clearly seen in verbs of perception and cognition, where transitives mean 'see, hear, touch, know', etc., whereas factitives mean 'make seen (30), heard, touched, known', etc.

(30) *ágūhat* *támo* *vy*
hide.IMPF.3SG darkness(N).ACC PV
***àcakṣayat*** *svàḥ*
look.FACT.IMPF.3SG.ACT sun(N).ACC
'He hid the darkness and made the sun visible.' (Ṛ̊V 2.24.3d)

Factitives were later interpreted as causatives, and in Middle Vedic the causative alternation began to spread to transitive verbs (Jamison 1983; Lubotsky 1989; Kulikov 2011). Already in the ṚV, some real *-áya-* causatives can be found, in which the experiencer serves as the second argument, while the stimulus may be expressed, as in (32), or not.

(31) *tám* *ūtáyo* ***raṇayañ***
DET.ACC help(F).NOM.PL enjoy.CAUS.INJ.3PL.ACT
*chū́rasātau*
contest(F).LOC.PL
'Him do his help(er)s [=Maruts] cause to take pleasure in the contest of champions.' (ṚV 1.100.7a)

Finally, in the Causative with instrument causee alternation, the body part that is responsible for the intake of a perceptive situation (e.g., the ears in the case of HEAR) is promoted to the role of causee:

(32) ***śrāváya*** *íd* *asya* *kárṇā*
hear.CAUS.IMPV.3SG PTCL 3SG.GEN ear.ACC.DU
'Cause his ears to hear.' (ṚV 4.29.3a)

A similar alternation is found, in the *Atharvaveda*, for *tap-*, where the causees are the heart (*hṛ́dáḥ*) and spirit (*mánaḥ*) of the enemy, that is, the seat of sadness.

(33) *dviṣatás* ***tāpáyan*** *hṛ́dáḥ*
foe.GEN burn.CAUS.PTCP.NOM.ACT heart(N).ACC
'Burning the heart of the foe.' (AVŚ 19.28.2)

## *4.2 Uncoded alternations*

Uncoded alternations featured by experiential verbs are summarized in Table 7.

| **Comparative concept** | **Language-specific alternation** | **Derived coding frame** | **D/A/R/I** |
|---|---|---|---|
| Passive | Passive lability | 2-nom V.mid.subj[2] | D |
| Reflexive | Reflexive lability | 1=2-nom V.mid.subj[1=2] | I |
| Generic argument omission | Generic argument omission | 1-nom V.subj[1] | D |

| | | | |
|---|---|---|---|
| | Predicative subject | 2-nom V.subj[2] 3-nom | R |
| | Predicative double accusative | 1-nom V.subj[1] 2-acc 3-acc | A |
| | Locative experiencer | 2-nom V.act.subj[2] 1-loc | R |

Table 9. *Uncoded alternations featured by experiential verbs.*

With Passive and Reflexive lability, verbs that are transitive media tantum are used in a monovalent coding frame, either with the subject performing the role of the Patient (Passive), or with the subject simultaneously performing both Agent and Patient (Reflexive). Among the verbs overviewed here, these alternations are only attested with *man-* 'think'; as the meaning shifts to 'consider', these alternations consistently occur together with the Predicative subject alternation, where a predicative nominative subject complement is added:

(34) *suvī́ras* *te* *janitā́*
with.heroes.NOM 2SG.GEN begetter.NOM
***manyata*** *dyaúr*
think.INJ.3SG.PASS heaven.NOM
'Rich in heroes, Heaven is considered to be your begetter.' (ṚV 4.17.4a)

In the Generic argument omission alternation, the second argument of two-place verbs is omitted because generic/indefinite; the omission may yield an atelic reading of the event. While this alternation is attested with most verbs under analysis, it is most typical with verbs of looking that are compounded with a PV or that take an oblique adding directionality to the event:

(35) *divyáḥ* *suparṇó* ***'va*** ***cakṣi***
heavenly.NOM eagle.NOM down look.IMPV.2SG.ACT
*soma*
Soma.VOC
'As heavenly eagle, o Soma, gaze down.' (ṚV 9.97.33a)

In the Predicative double accusative alternation, a predicative accusative object complement coreferential with 2 is added:

(36) ***vidmā́*** *hí* *tvā* *vṛ́ṣantamaṁ*
know.PF.1PL.ACT PTCL 2SG.ACC best.of.bulls.ACC

'Since we know you as the best of bulls.' (ṚV 1.10.10a)

Finally, the Locative experiencer alternation is only attested for *spṛś-* (TOUCH) and was discussed in Section 3.2.

## 5. Conclusion

In this study, we have examined the valency patterns and alternations of experiential predicates in Vedic Sanskrit using the PaVeDa methodology. Vedic strongly prefers nominative alignment in encoding experiential situations, with most verbs functioning as two-place predicates where experiencers appear as subjects and stimuli as second arguments marked by various cases. Significant construction variation exists across domains, with case alternations often tied to semantic-pragmatic factors such as animacy of the stimulus, intentionality, and control. Finally, while maintaining patterns common to other ancient IE languages, Vedic Sanskrit exhibits language-specific characteristics in alternations, as shown for instance by the three causative alternations.

## References


Casaretto, A. 2002. Defektivität und Suppletion im vedischen Verbalsystem: ved. paś: darś, drav: drā und dhāv: sar. *Historische Sprachforschung/Historical Linguistics* 115(1): 37–56.

Croft, W. 2012. *Verbs: Aspect and Causal Structure*. Oxford: OUP.

Dahl, E. 2009. Some semantic and pragmatic aspects of object alternation in Early Vedic. In S.L. Chelliah and J. Barðdal (eds.), *The role of semantic, pragmatic and discourse factors in the development of case*, 23-55. Amsterdam: Benjamins.

Dahl, E. 2014. The morphosyntax of the Experiencer in Early Vedic. In S. Luraghi and H. Narrog (ed.), *Perspectives on semantic roles*, 181-204. Amsterdam: Benjamins.

Dahl, E. 2015. Toward a formal model of semantic change: a Neo-Reichenbachian approach to the development of the Vedic past tense system. *Lingua Posnaniensis* 57(1): 41-76.

Dahl, E. and C. Fedriani. 2012. The Argument Structure of Experience: Experiential Constructions in Early Vedic, Homeric Greek and Early Latin 1. *Transactions of the Philological Society* 110(3): 342-362.

Danesi, S. 2013. Particle-verb Constructions in Vedic: The Case of ápa. *Studi e Saggi Linguistici* 51: 57-100.

De Rossi, N. 2023. *"Bruciare" nel Mahābhārata: un'analisi delle radici dah-, tap-, uṣ-, jval- (e jvar-), ghṛ- e śuc- nel quadro della semantica cognitiva*. MA Thesis, University of Pavia.
Goddard, C. 2003. Thinking Across Languages and Cultures: Six Dimensions of Variation. *Cognitive Linguistics* 14(2–3): 109-140.
Grassmann, H. 1936[1873]. *Worterbuch zum Rig-Veda*. Leipzig: Brockhaus.
Insler, S. 1968. The origin of the Sanskrit passive aorist. *Indogermanische Forschungen* 73: 312-346.
Jamison, S. W. 1983. *Function and Form in the-áya-Formations of the Rig Veda and Atharva Veda*. Göttingen: Vandenhoeck and Ruprecht.
Koptjevskaja-Tamm, M. (ed.). 2015. *The linguistics of temperature*. Amsterdam: John Benjamins.
Kortlandt, F. 1981.1st sg. middle *-$H_2$. *Indogermanische Forschungen* 86: 123-136.
Kövecses, Z. 1995. Anger: Its language, conceptualization, and physiology in the light of cross-cultural evidence. *Language and the Cognitive Construal of the World* 82: 181-196.
Kulikov, L. 2006. Passive and middle in Indo-European: Reconstructing the early Vedic passive paradigm. In W. Abraham and L. Leisiö (eds.), *Passivization and typology: form and function*, 62-81. Amsterdam: John Benjamins.
Kulikov, L. 2009. Valency-changing categories in Indo-Aryan and Indo-European: A diachronic typological portrait of Vedic Sanskrit. In A. Saxena, Å. Viberg (ed.), *Multilingualism: Proceedings of the 23rd Scandinavian Conference of Linguistics* (*Studia Linguistica Upsaliensis* 8), 75–92. Uppsala: Uppsala Universitet.
Kulikov, L. 2011. Passive to anticausative through impersonalization: The case of Vedic and Indo-European. In A.L. Malchukov, A. Siewierska (ed.), *Impersonal constructions: A cross-linguistic perspective*, 229-254. Amsterdam: Benjamins.
Kulikov, L. 2012. Vedic Preverbs as Markers of Valency-Changing Derivations. *Studies in Language* 34(4): 721-746.
Kulikov, L. 2013. 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. N. Lavidas. 2013. Reconstructing passive and voice in Proto-Indo-European. *Journal of Historical Linguistics* 3(1): 98-121.
Kümmel, M. 1996. *Stativ und Passivaorist im Indoiranischen*. Göttingen: Vandenhoeck and Ruprecht.
Kümmel, M. 2000. *Das Perfekt im Indoiranischen*. Wiesbaden: Reichert.
Lakoff, G. and M. Johnson. 1980. *Metaphors we live by*. Chicago: CUP
Langacker, R. W. 1999. *Grammar and Conceptualization*. Berlin: Mouton de Gruyter.
Lubotsky, A. 1989. The Vedic-áya-formations. *Indo-Iranian Journal* 32(2): 89-113.
Luraghi, S. 2020. Experiential verbs in Homeric Greek, A constructional approach, Brill's Studies in Language, Cognition and Culture. vol.27, Brill, Boston.
Luraghi, S. mAlessandra C. and E. Pinelli. 2014. Accusative and genitive objects with verbs of perception and mental activity in Homeric Greek, Early Vedic, and Old Church Slavic S. Luraghi, A. Caviglia and E. Pinelli. *Paper presented at the*

*Workshop "Case, Arguments And Clause Structure In Indo-European*" Ghent, 5-6 May 2014.
Radden, G. 2000. How metonymic are metaphors?. *Topics in English Linguistics* 30: 93-108.
Thieme, P. 1929. *Das Plusquamperfectum im Veda*. Göttingen: Vandenhoeck and Ruprecht.
Viberg, Åke. 1984. The verbs of perception: a typological study. In Brian Butterworth, Bernard Comrie and Östen Dahl (eds.), *Explanations for language universals*, 123-162. The Hague etc.: Mouton.
Verhoeven, E. 2007. *Experiential Constructions in Yucatec Maya. A typologically based analysis of a functional domain in a Mayan language.* Amsterdam: John Benjamins.
Wierzbicka, A. 2007. Bodies and their Parts: A NSM Approach to Semantic Typology. *Language Sciences* 29: 14-65.
Zanchi, C. 2019. *Multiple Preverbs in Ancient Indo-European Languages: A Comparative Study on Vedic, Homeric Greek, Old Church Slavic, and Old Irish*. Narr Francke Attempto Verlag.
Zanchi, C., S. Luraghi, and C. R. Combei. 2022. PaVeDa-Pavia Verbs Database: Challenges and Perspectives. *Proceedings of the 4th workshop on research in computational linguistic typology and multilingual NLP*.

# The PaVeDa Dataset for Modern Standard Russian: Some Observations on the Revision Process of the ValPaL Russian Dataset

*Erica Pinelli*[*]

## 1. Introduction

The Valency Patterns Leipzig Online Database (ValPaL) already includes a dataset for Modern Standard Russian, compiled by Andrej Malchukov and Alexander Jahraus[1]. The ValPaL Russian dataset consists of 90 verb meanings with their corresponding Russian verbs, basic coding frames, and 15 language specific alternations.

As stated on the website, the Pavia Verbs Database (PaVeDa, https://paveda.unipv.it/) aims to expand and enhance the ValPaL database by adding more languages and further features. For what Russian language is concerned, the main goal of PaVeDa is to add information, in particular, real examples from corpora. However, before collecting examples for each verb form and their alternations, we revised the Russian dataset in order to correct mistakes and solve inconsistences. We revised the ValPaL Russian dataset in all its parts: verb meanings selection, identification of corresponding Russian verbs, description of basic or derived coding frames, and alternations. We made changes at all different levels of the dataset, sometimes adding elements, sometimes deleting them or modifying the existing ones, in order to obtain a complete and consistent dataset, not only from an intra-linguistic perspective, but also with respect to datasets for other languages in the database.

In this contribution, we focus on the revision process of the existing ValPaL Russian dataset. In particular, in Section 2 we discuss some of the inconsistences we found and solved at different levels of the dataset: at the

[*] University of Pavia

1. https://valpal.info/contributions/russ1263

level of verb meaning (2.1), verb form (2.2), basic coding frame (2.3), and alternations (2.4). In Section 3 we present some conclusions.

## 2. Inconsistencies in the ValPaL Russian dataset and possible solutions

### *2.1. Adding verb meanings*

The first step of the revision process of the ValPaL Russian dataset involved the level of the verb meaning. In some cases we decided to add verb meanings to the original dataset. Let us make some examples.

The ValPaL Russian dataset already includes the verb meaning ASK FOR and its corresponding Russian verb *prosit'*. Looking at ValPaL datasets for other languages, we can note that the corresponding verb form for ASK FOR usually covers also other meanings, such as "ask about" or "ask a question". For example, in the Italian and English datasets, the verb *chiedere* and the verb *to ask* respectively can be used not only to "ask for something", but also to "ask an information", "ask about something/someone" or "ask a question". Differently from Italian or English, the Russian verb *prosit'* can only be used to express the meaning ASK FOR. To request information or ask a question in Russian the verb *sprašivat'* is required. Thus, we decided to add two more verb meanings, ASK (about) and ASK (a question), to the Russian dataset (as shown in Table 1).

| VERB MEANING | VERB FORM | BASIC CODING FRAME |
|---|---|---|
| ASK (about) | *sprašivat'* | 1-nom V.subj[1] (o+2-loc) 3-acc |
| ASK (a question) | *sprašivat'* | 1-nom V.subj[1] (UTT2) (u+3-gen) |
| ASK FOR | *prosit'* | 1-nom V.subj[1] 2-acc (u+3-gen) |

Table 10. *The verb meaning ASK FOR (already given in the ValPaL Russian dataset) and the two verb meanings added to the PaVeDa Russian dataset: ASK (about) and ASK (a question).*

As we can see from Table 1 and from examples (1)-(3), the three verb meanings ASK (about), ASK (a question), and ASK FOR not only involve different verb forms, but they also require different coding frames.

| | | | | |
|---|---|---|---|---|
| (1) | *Kogda* | *ja* | *eë* | *sprašiva-l* |
| | When | 1SG.NOM | 3SG.F.ACC | ask-PST.SG.M |

*o* *čëm-nibud',* *ona* *ne* *chotela otvečat'*
about something.LOC, she not wanted to_answer
'Whenever I asked her about something, she didn't want to answer (...).' (V. Pelevin. *Nika*, 1992)

(2) *Ja* *spraši̇va-l* *u* *prokuror-a*
1SG.NOM ask-PST.SG.M from procurator-GEN
*Tversk-ogo* *rajon-a* *za* *čto* *že* *rebjat*
of_Tver'-GEN District-GEN for what PTC guys.ACC
*nakaza-l-i* *tak* *žestoko?*
punish-PST-PL so cruelly?
'I asked the procurator of the Tverskoj district, why were the guys punished so cruelly?' (A. Andreev. *Buduščee prinadležit nam!*, «Zavtra», 2003.08.22)

(3) *Ran'še* *Svetlan-a* *prosi-l-a* *den'gi*
Before Svetlana-NOM ask_for-PST-SG.F money.PL.ACC
*u* *nich* *tak*
from 3PL.GEN that_way
'Previously Svetlana asked them for money like that: (…).' (A. Slapovskij. *Gibel' gitarista*, 1994-1995)

Considering the relevant differences among these three verb meanings at the level of verb forms and coding frames, the information added to the PaVeDa Russian dataset not only provides a more complete overview of inquiry verbs in Russian, but also enables a more fruitful cross-linguistic comparison.

Similarly, other verb meanings have been added to the PaVeDa Russian dataset for the sake of completeness. Let us make another example. In the ValPaL Russian dataset the verb meaning FEEL PAIN corresponds to the verb form *bolet'*, as exemplified in (4).

(4) FEEL PAIN *bolet'*
BASIC CODING FRAME: 1-nom V.subj[1] u+2-gen
*I* *u* *menja* *golov-a* *segodnja*
And at 1SG.GEN head-NOM today
*bol-it* *nemnogo.*
hurt-PRES.3SG slightly
'And I have a little headache today.' (A. Gelasimov. *Džanna*, 2001)

However, it is interesting to note that the infinitive form of the verb of the second conjugation *bolet'* 'to feel pain' (example 4) is homonymous with the infinitive form of the verb of the first conjugation *bolet'* that means 'to be ill'

(example 5). In order to avoid misinterpretations or mistakes, it is worth to make the homonymy of these two infinitive forms explicit in the dataset: in this way we clarify that in Russian the verb meanings FEEL PAIN and BE ILL correspond to two different verbs that show homonymy in the infinitive form and have different coding frames. For this reason, we decided to add the verb meaning BE ILL to the PaVeDa Russian dataset.

(5) BE ILL *bolet'*
BASIC CODING FRAME: nom V.subj[1] (2-instr)

| | | | | |
|---|---|---|---|---|
| *Poėtomu* | | *det-i,* | *kotor-ye* | *rosli* |
| For_this_reason | | children-NOM | who-NOM.PL | grew |
| *na* | *korov'em* | *moloke,* | *často* | |
| on | cow's | milk | frequently | |
| *bole-jut* | | *rachit-om* | | |
| be_ill-PRS.3PL | | rickets-INST | | |

'For this reason children who grew on cow's milk frequently suffer from rickets.' (A. Razakova. *Pervyj god i vsja žizn'*, «100% zdorov'ja», 2002.11.11)

The addition of this verb meaning has a twofold motivation. On the one hand, it clarifies the structure of the verb *bolet'* from an intra-linguistic perspective; on the other hand, since the verb meaning BE ILL is already given in ValPaL datasets for other languages, it is also important for a complete and precise interlinguistic analysis.

## *2.2. Adding, modifying and deleting verb forms*

The creation of the PaVeDa Russian dataset, starting from data contained in ValPaL, allowed us to solve come inconsistences at the level of the verb forms. An example of this kind of inconsistences concerns the verb meanings BE A HUNTER, BE DRY, and BE HUNGRY. These three verb meanings appear in the original ValPaL Russian dataset, but they are treated differently.

As we can see in Figure 1, in the ValPaL Russian dataset the corresponding verb form for the meaning BE HUNGRY is *byt' golodnym*, which consists of a copula *byt'* 'to be' and the adjective *golodnyj* 'hungry' in the instrumental case.

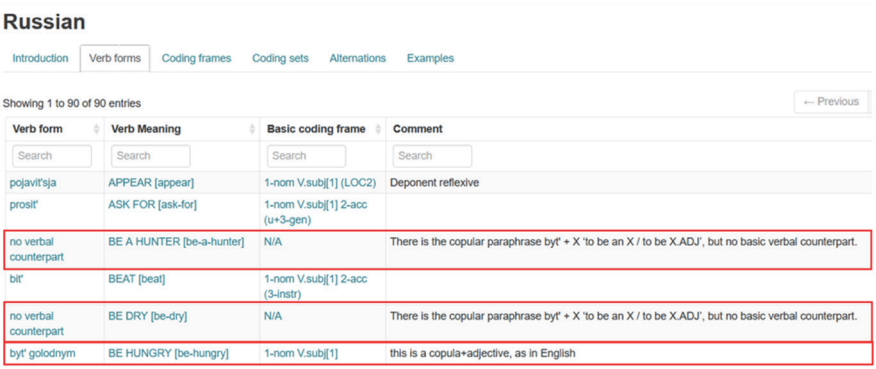


Figure 5. *Verb meanings BE A HUNTER, BE DRY, and BE HUNGRY in the ValPaL Russian dataset.*

It is interesting to note that, as made explicit in the comments, similar copular constructions could be used as verb forms for BE A HUNTER (*byt'* 'be' + *ochotnik-om* 'hunter-INS') and BE DRY (*byt'* 'be' + *such-im* 'dry-INS'). Nevertheless, the verb meanings BE A HUNTER and BE DRY remain without a verbal counterpart. In order to make data consistent, in the PaVeDa Russian dataset we decided to add the copular construction as verbal counterpart also for the verb meanings BE A HUNTER and BE DRY (see Table 2).

| VERB MEANING | VERB FORM | BASIC CODING FRAME |
|---|---|---|
| BE A HUNTER | *Byt' ochotnikom* | 1-nom V.subj[1] |
| BE DRY | *Byt' suchim* | 1-nom V.subj[1] |
| BE HUNGRY | *Byt' golodnym* | 1-nom V.subj[1] |

*Table 11. The verb meaning BE HUNGRY (already given in the ValPaL Russian dataset) and the verb meanings BE A HUNTER and BE DRY with corresponding verb forms and basic coding frames added to the PaVeDa Russian dataset.*

As just shown, in some cases, we decided to add a verbal counterpart to a verb meaning that was given in the dataset but was lacking of a corresponding Russian verb form. In some other cases, we decided to provide another verb form in addition to the one already given in the ValPaL dataset. For example, let us consider the verb meaning FEAR. In the ValPaL Russian dataset the meaning FEAR corresponds to the verb form *bojat'sja*, a reflexive verb that is used to convey prototypical fear, as shown in example (6).

(6) *V* *detstve* *ja* *boja-l-a-s'*
In childhood 1SG.NOM be_afraid-PST-SG.F-REFL
*grom-a* *boja-l-a-s'* *temnot-y*
thunder-GEN be_afraid-PST-SG.F-REFL darkness-GEN
'When I was a child, I was afraid of thunder and I was afraid of the dark.' (V. Grossman. *Žizn' i sud'ba*, č. 1, 1960)

However, a very frequently used construction to express occurrent fear (Apresjan 2004: 60) consists of a dative Experiencer and the predicative form *strašno* 'frightful', as exemplified in (7).

(7) *Ja* *ne* *mogu* *odna* *tam,*
1SG.NOM NEG can.PRS.1SG alone there
*mne* *strašno*
1SG.DAT frightfully
'I cannot stay there alone, I'm scared.'(A. Klepakov. *Opekun*, «Volga», 2016)

The choice to add a second verb form corresponding to the verb meaning FEAR is important both from an intra-linguistic and a cross-linguistic perspective. In Russian the verb *bojat'sja* and the predicative construction DAT+*strašno* differ in their semantics and their syntactic behavior. As we can see from example (6), *bojat'sja* can express a psychological state, but it can also be used to refer to dispositional fear, i.e. expectation or belief that something undesirable can happen, as shown in (8).

(8) a. *On* *bo-it-sja,* *čto*
3SG.M.NOM be_afraid-PRS.3SG-REFL that
*otec* *pried-et*
dad.NOM come-FUT.3SG
'He is afraid that his father will come.'
b. *On* *bo-it-sja* *za*
3SG.M.NOM be_afraid-PRS.3SG-REFL for
*syn-a*
son-ACC
'She is afraid for her son'.
(Adapted from Iordanskaja/Mel'čuk 1990: 307)

On the contrary, the construction DAT+*strašno* can only be used to refer to an occurrent emotional state, while it cannot refer to general feature (**emu strašno temnoty* 'he is afraid of the dark') or to dispositional fear (**emu*

*strašno, čto otec priedet* ‘he is afraid that his dad is coming’). Moreover, unlike the construction with the reflexive verb *bojat’sja*, in which the genitive-marked Stimulus is part of the verbal argumental structure, in the predicative construction DAT+*strašno* the Stimulus cannot be considered an argument and is usually not expressed (see example 7); in certain contexts, the Stimulus can be expressed by an infinitive verb (9a), a clause (9b) or a causal prepositional phrase (9c).

(9) a. *Ja znaju, tebe strašno echat’ tuda.*
I know, 2SG.DAT frightful go.INF there
‘I know you are scared to go there.’ (*Prichožan ždut na rodine*, «Žizn’ nacional’nostej», 2003.06.18)

b. *Mne strašno, čto my razluč-im-sja*.
1SG.DAT frightful that 1PL.NOM separate-PRS.1PL-REFL
‘I'm scared that we'll be separated.’ (Forum: *Obsuždenie fil’ma «Speši ljubit’»*, 2011)

c. *(…)Mne strašno ot ėt-oj tišin-y*.
(…)1SG.DAT frightful from this-GEN silence-GEN
‘(…) I am afraid of this silence.’ (E. Kulikov. *Bel’čonok*, «Dal’nij Vostok», 2019)

Considering the relevance of the two forms *bojat’sja* and *byt’ strašno* to express fear in Russian and their different argument structure, we decided to include both in the PaVeDa Russian dataset (see Table 3).

| VERB MEANING | VERB FORM | BASIC CODING FRAME |
|---|---|---|
| FEAR | *Bojat’sja* | 1-nom V.subj[1] 2-gen |
| FEAR | *Byt’ strašno*[2] | 1-dat V |

Table 12. *The verb form bojat’sja (already given in the in the ValPaL Russian dataset) and the verb form byt’ strašno (and its basic coding frame) added to the verb meaning FEAR in the PaVeDa Russian dataset.*

2. As we can see in Table 3, the predicative construction also requires the copula *byt’* ‘to be’ [Dative Experiencer+*byt’*+ *strašno*] that does not appear in the present tense, but is used in the past and future tense.

Moreover, if we look at the datasets for other languages we can observe that the addition of *byt' strašno* does not create an anomalous situation in the Russian dataset.

| | | |
|---|---|---|
| fall | FALL [fall] | 1-nom > V.subj[1] > LOC2 |
| fear | FEAR [fear] | 1-nom > V.subj[1] > 2-acc |
| be afraid | FEAR [fear] | 1-nom > V.subj[1] > of+2 |
| feel cold | FEEL COLD [feel-cold] | 1-nom > V.subj[1] |

Figure 6. *Two verb forms corresponding to the verb meaning FEAR in the ValPaL dataset for English.*

| | | |
|---|---|---|
| fallen | FALL [fall] | 1-nom V.subj[1] LOC2 |
| Angst haben | FEAR [fear] | 1-nom V.subj[1] vor+2-dat |
| fürchten | FEAR [fear] | 1-nom V.subj[1] 2-acc |
| kalt sein | FEEL COLD [feel-cold] | 1-dat V |

Figure 7. *Two verb forms corresponding to the verb meaning FEAR in the ValPaL dataset for German.*

In the ValPaL English dataset there are two verb forms corresponding to FEAR, i.e. *fear* and *be afraid* (see Figure 2); a similar situation can be observed in the German dataset in which both *fürchten* and *Angst haben* appear (see Figure 3). For this reason, the addition of *byt' strašno* makes the Russian dataset more homogeneous with respect to datasets for other languages contained in the database and therefore more complete and suitable for cross-linguistic analysis.

In some other cases, more consistent data were obtained by deleting verb forms from the original ValPaL Russian dataset. Let us look, for example, at the following verb meanings.

| VERB MEANING | VERB FORM | BASIC CODING FRAME |
|---|---|---|
| COME | *Prijti* | 1-nom V.subj[1] (LOC2) |
| GO | *Idti* | 1-nom V.subj[1] (LOC2) |
| LEAVE | *Ujti* | 1-nom V.subj[1] (iz/ot 2-gen) |
| LEAVE | *Uechat'* | 1-nom V.subj[1] (iz/ot 2-gen) |

Table 13. *Verb meanings COME, GO and LEAVE as given in the ValPaL Russian dataset.*

As we can see from Table 4, in the ValPaL Russian dataset the verb meanings COME and GO correspond to one verb form each, *prijti* and *idti*

respectively, while the verb meaning LEAVE has two corresponding verbs, *ujti* and *uechat'*. Being a particular class in Russian, verbs of motion deserve a detailed explanation. We should start by saying that in Russian different verbs are used to express different manners of the motion. For example, the verb *idti* is used for a motion on foot, while the verb *echat'* for a motion by a vehicle. Moreover, according to the categorization given by Talmy (2000), Russian is a satellite-framed language, i.e. a language in which the path of the motion is not expressed by a specific verb but by an element associated with the motion verb (see Slobin 2004). In particular, in Russian the 'satellites' used to specify the path are the prefixes: for example, the prefix *pri-* is used for arrival, *u-* for leaving, *vy-* for exiting, *v-* for entering, *ot-* for removal, and *pod-* for approaching. This means that all the verbs describing different manners of the motion, such as *idti* 'go (on foot)' or *echat'* 'go (by a vehicle)', can be prefixed in order to specify the path of the motion. Thus, Russian prefixed verbs, such as *prijti* 'arrive on foot', *priechat'* 'arrive by a vehicle', *ujti* 'go away on foot', *uechat'* 'go away by a vehicle', *vyjti* 'go out on foot', *vyechat'* 'go out by a vehicle' and so on and so forth, specify both the manner and the path of the motion.

Since *idti* is also used for general motion[3] (see Rachilina 2004), it is therefore absolutely justifiable the correspondence of *idti* to the meaning GO and of *prijti* to the meaning COME. However, the correspondence of two verb forms, *ujti* and *uechat'*, for LEAVE can be misleading: in fact, also for GO and COME other verb forms that imply a different manner can be used (*echat'* and *priechat'* respectively). For this reason, we decided to delete the second corresponding verb *uechat'* from the meaning LEAVE (see Table 5)[4].

| **Verb meaning** | **Verb form** | **Basic coding frame** |
|---|---|---|
| COME | *Prijti* | 1-nom V.subj[1] (LOC2) |
| GO | *Idti* | 1-nom V.subj[1] (LOC2) |
| LEAVE | *Ujti* | 1-nom V.subj[1] (iz/ot 2-gen) |

Table 14. *Verb meanings COME, GO and LEAVE in the Russian PaVeDa dataset*

3. Rachilina (2004: 3) shows that, in some specific contexts, the verb *idti* can be used as a "cover term for different types of movements".

4. For the sake of completeness, we added a comment to the verb meanings COME, GO, and LEAVE in order to explain that other verb forms expressing different manners of the motion can also be used. A similar comment has been added to the verb meanings BRING and CARRY that in Russian correspond to transitive motion verb forms (for example, *nesti* 'carry by hand' or *vezti* 'carry by a vehicle', and *prinesti* 'bring by hand' or *privezti* 'bring by a vehicle').

## *2.3. Changes at the level of basic coding frame*

The analysis of ValPaL Russian data allowed changes to be made also at the level of the basic coding frame. Let us make an example.

The already mentioned verb form *bojat'sja*, corresponding to the meaning FEAR, has the following basic coding frame in the ValPaL database: 1-nom V.subj[1] 2-acc. This coding frame shows that the verb *bojat'sja* requires a nominative Experiencer, while the Stimulus is coded by the accusative case. In general, Russian reflexive verbs in *-sja* do not normally combine with accusative objects; the verb *bojat'sja*, as other reflexive verbs expressing fear, such as *opasat'sja* 'be afraid of', *pugat'sja* 'be frightened by', *strašit'sja* 'be afraid of', but also other reflexive verbs, such as *kasat'sja* 'touch', *lišat'sja* 'be deprived of', and *slušat'sja* 'obey', usually combine with the genitive case (Apresjan 2004: 64; Kuznetsova-Nesset 2015: 255-256). Although genitive is the most neutral way to code the Stimulus, it is, however, true that *bojat'sja* admits an accusative object.

Iordanskaja and Mel'čuk (1990: 319) claim that the verb *bojat'sja* can occur with the accusative when the object itself, i.e. the Stimulus, is human and is in contact with the Experiencer, as exemplified in (10a); on the contrary, the genitive case is used to code a Stimulus that is not located next to the Experiencer, as in example (10b). This means that the sentences in (10a) and (10b) have the same meaning, but in (10a) the wife, i.e. the Stimulus, is present in the same location with the Experiencer, while in (10b) she is not necessarily present.[5]

(10) *a.* *On* *ne* *vojd-ët* *v* *zal:*
He NEG enter-FUT.3SG into hall.ACC
*bo-it-sja* *žen-u*
be_afraid-PRS.3SG-REFL wife-ACC
'He won't go into the room: he's afraid of his wife' [the wife is present here and now]

*b.* *On* *ne* *vojd-ët* *v* *zal:*
He NEG enter-FUT.3SG into hall.ACC
*bo-it-sja* *žen-y*
be_afraid-PRS.3SG-REFL wife-GEN

5. This example allows us to understand the semantic difference between the genitive and the accusative object of the verb *bojat'sja*. However, we should also consider that Russian shows a genitive-accusative syncretism for animate referents, with the exception of feminine singular referents. This means that in many cases it is not possible to distinguish the accusative from the genitive form.

‘He won’t go into the room: he’s afraid of his wife’ [the wife is not necessarily present]. (adapted from Iordanskaja/Mel’čuk 1990: 319)

However, there is no general agreement among linguists. The academy grammar (Švedova 1980: 26) indicates that the verb *bojat’sja* only admits genitive objects, while Iordanskaja and Mel’čuk (1990: 319) claim that the semantic distinction between accusative and genitive objects tends to disappear in Modern Russian and that younger generations do not admit the use of accusative objects. However, other linguists, such as Krys’ko (1997: 240-246), claim that the accusative is not only admitted but is massively used. A more recent corpus-based study by Kuznetsova and Nesset (2015) shows that accusative objects are relatively infrequent in Contemporary Russian, but they appear significantly more often in less restricted registers of language and, when used, they are more likely to code individuated objects (example 11). However, in the majority of cases[6] the verb *bojat’sja* occurs with a genitive-marked Stimulus, as shown in examples (6) and (12).

(11) *Ja* *bo-ju-s’* *sobak-u* *iz*
1SG.NOM fear-PRS.1SG-REFL dog-ACC from
*sosedn-ego* *dom-a* *(...).*
adjacent-GEN house-GEN
‘I’m scared of the dog from next door (…).’ [N. Tavobova. *Strach v detskich glazach* (2002), «Večernjaja Moskva», 09.12.2002]

(12) (…) *ja* *s* *detstv-a*
(…)1SG.NOM from childhood-GEN
*bo-ju-s’* *vod-y.*
fear-PRS.1SG-REFL water-GEN
‘I've been afraid of water since I was a child.’ [V. Spektr. *Face Control* (2002)]

Even if the verb *bojat’sja* in certain contexts can occur with an accusative object, we decided to provide the most semantically and stylistically neutral basic coding frame with the genitive object: 1-nom V.subj[1] 2-gen.[7]

6. Kuznetsova and Nesset (2015: 266) claim that the accusative appears only in 1% of examples in the main corpus, and in 7% of examples in the newspaper corpus. Although in newspapers the accusative case appears more frequently, the overall percentage of accusative objects is however low.

7. The choice of ValPaL contributors to select a basic coding frame with the accusative object for *bojat’sja* may also be motivated by the quest for consistency between the Accusative-

## 2.4. *Adding, modifying or deleting alternations*

The ValPaL database contains 15 alternations for Russian. Despite the fact that a definition of each alternation is given, several issues arise during the compilation of the dataset, in particular during the selection of examples from corpora. Addressing the definitions of alternations is beyond the aims of the present paper; nevertheless, we give some examples of the changes we made at this level of the dataset.

In the ValPaL Russian dataset only few examples are provided. In fact, one of the main advantages of PaVeDa is that the Russian dataset has been enriched with real examples mainly taken from the Russian National Corpus (www.ruscorpora.ru).[8] In PaVeDa, for each basic coding frame and each alternation, an example is provided to clarify the verb's argument structure and its meaning. The search for real examples in the Russian National Corpus gave us the possibility to check the existence of different structures in real language.

For some alternations no examples have been found. This is what happened with the verb meaning CARRY that corresponds to the Russian verb *nesti*. In the ValPaL Russian dataset the verb *nesti* is described as having three regularly occurring alternations, i.e. the Reflexive Passive, the Negative Accusative-Genitive, and the Participial Passive alternations, as well as a marginal one, i.e. the Reflexive Middle alternation. Searching for the reflexive form *nestis'* in the RNC, we found the Reflexive Passive alternation, as exemplified in (13).

(13)

| | | | | |
|---|---|---|---|---|
| *Ja* | *beremenna* | *tak* | *v* | *podarok* |
| I | pregnant | so | in | present.NOM |
| *nes-ut-sja* | | | *detsk-ie* | |
| carry.IPFV-PRS.3PL-REFL | | | children-PL.NOM | |
| *vešč-i* | | *iz* | *sekond-chend-a* | |
| cloth-PL.NOM | | from | second_hand-SG.GEN | |

'I'm pregnant, so children's clothes from a second-hand store are being carried as a gift.' (Ženščina + mužčina: Brak (forum), 2004)

Then, we searched for a Reflexive Middle alternation. According to the definition given in the ValPaL database, the Reflexive Middle, similarly to

Genitive alternation and the Negative Accusative-Genitive alternation (that have been completely removed from PaVeDa, see 2.4).

8. Only few examples have been taken from other sources (for example, GoogleBooks) or have been elicited from native speakers.

the English Middle (for example, "The glass easily breaks"), "requires a modification by means of an (evaluative) adverb" and "is used (only) with imperfective verbs". However, a Middle construction could not be found in Russian data.[9] For this reason, we decided to delete this alternation from the verb meaning CARRY.

Other alternations given in the ValPaL Russian dataset have been completely deleted in PaVeDa. In particular, we decided to delete those alternations that were not affecting the argument structure. Let us make an example. As we have already explained in 2.2. for *bojat'sja* 'fear, be afraid of', some Russian verbs (such as, *ždat'* 'wait for' or *iskat'* 'look for') admit both accusative or genitive direct objects. The use of a specific case mainly depends on the degree of individuation of the object: a determined and specific object is more likely to be coded by the accusative case (example 14a), while a less individuated object usually occurs in the genitive (example 14b).

(14) a. *Ja* *žd-u* *avtobus*
1SG.NOM wait-PRS.1SG autobus.ACC
*nomer* *2*
number two
'I am waiting the bus number 2 (that specific bus).'

b. *Ja* *žd-u* *avtobus-a*
1SG.NOM wait-PRS.1SG autobus-GEN
'I am waiting a bus (any bus).'

Another case in which the accusative and the genitive cases alternate is the Negative Accusative-Genitive alternation. In general, in Russian the object of a negated transitive verb occurs in the genitive (example 15a), while in the corresponding affirmative sentence the object is coded by the accusative (example 15b).

(15) a. *Ja* *ne* *čitala*
1SG.NOM NEG read-PST-F.SG
*gazet- ø.*
newspaper-GEN.PL
'I didn't read the newspapers.'

b. *Ja* *čitala* *gazet-y.*
1SG.NOM read-PST-F.SG newspaper-ACC.PL
'I read the newspapers.'

9. Moreover, no examples could be obtained by elicitation from native speakers of Russian.

However, similarly to the object of the Genitive-Accusative alternation, the object of a negated transitive verb can occur both in the genitive (16a) or the accusative (16b) depending on its degree of individuation.

(16) a. *Ja* *ne* *pi-l* *vod-y*
1SG.NOM NEG drink-PST.SG.M water-GEN
'I didn't drink (any) water.'
*b.* *Ja* *ne* *pi-l* *vod-u*
1SG.NOM NEG drink-PST.SG.M water-ACC
'I didn't drink (the) water.'(adapted from Kagan 2013: 5)

Even if this topic is rather more complex[10] and it has been addressed by a number of scholars (among many others, Korn 1967, Timberlake 2004, Partee et al. 2011, Kagan 2013), for our purpose it is enough to say that in Russian both the Accusative-Genitive and the Negative Accusative-Genitive alternations are used to mark specific features of the object and they do not affect the verbal argument structure, i.e., they do not lead to valency augmentation or reduction, nor to rearrangement or identification of the arguments (in terms of Malchukov 2015). For this reason, the Accusative-Genitive and the Negative Accusative-Genitive alternations have been deleted from the Russian PaVeDa dataset.

The process of selecting real examples from the RNC also allowed us to identify alternations that were not given in the ValPaL Russian dataset. For example, the verb meaning LOOK AT, that corresponds to the Russian verb *smotret'*, has no alternation in the ValPaL database. However, searching the RNC, we found examples for the Reflexive Middle alternation (example 17) and we decided to add it to the PaVeDa Russian dataset.

(17) LOOK AT: *smotret'* ALTERNATION: Reflexive Middle
ALTERNATION CODING FRAME: 2-nom V'sja.subj[2]
*Fil'm* *smotr-it-sja* *očen'* *legko*
Film.NOM look_at-PRS.3SG-REFL very easily
'The movie is very easy to watch.' (Forum: *kommentarii k fil'mu «Vse budet chorošo»*, 2008-2011)

10. For Example, for highly individuated human objects the accusative case is obligatory as in (a).

*a.* *Ja* *ne* *pomni-l-a* *Len-u/*Len-y*
1SG.NOM NEG remember-PST-SG.F Lena-ACC/GEN
'I didn't remember Lena.' (Adapted from Kagan 2013: 12)

In some other cases, we modified the alternations given in ValPaL. For example, in the ValPaL Russian dataset several alternations have been identified for the meaning BEAT, that corresponds to the Russian verb *bit'*. Let us focus our attention on the Reflexive Anticausative alternation.

| Alternation | Derived coding frame | Occurs | Comment |
|---|---|---|---|
| Search | Search | --any-- | Search |
| Negative Accusative-Genitive alternation | 1-nom negV'.subj[1] 2-gen (3-instr) | Regularly | |
| Participial Passive | 2-nom passV'.subj[2] (3-instr)/(1-instr) | Regularly | Requires prefixation on verb to make it perfective: po-bit' |
| Reflexive Reciprocal | | Marginally | |
| Reflexive Passive | | Marginally | |
| Instrumental subject alternation | | Marginally | |
| Reflexive Middle | | Marginally | |
| Reflexive Anticausative | | Marginally | Only for the perfective countepart raz-bit'sja |

Figure 8. *Alternations of the verb meaning BEAT (verb form bit') as given in the ValPaL Russian dataset.*

As we can see in Figure 4, the comment to the Reflexive Anticausative alternation specifies that the reflexive perfective verb *razbit'sja* is required. However, this makes some problems arise. In fact, the verb *razbit'sja* means 'to break' as we can see in (18).

(18) *Odin iz bokalov upal so stola*
One.NOM of wineglasses fell off table.GEN
*i razbi-l-sja.*
and broke-PST.3SG.M-REFL
'A wineglass fell off the table and broke.' (M.A. Bulgakov. *Master i Margarita*, č. 2, 1929-1940)

Prefixes are widely used in Russian and prefixation is one of the morphological processes to derive perfective verb forms. However, prefixes not only carry aspectual meaning, but they can also be carriers of lexical meaning, as in the case of *razbit'sja*, in which the prefix *raz-* adds the semantic element '(dividing) into different pieces or parts' that led to the meaning 'to break'. Since the meaning of the verb changes and does not correspond to BEAT anymore, we removed the comment from the dataset and we searched for a more appropriate example for the Reflexive Anticausative alternation.

(19) *Za oknom (...) bi-l-a-s' ob*
Outside window (...) beat.IPFV-PST-F.SG-REFL against
*sten-u na vetru avos'k-a s*
wall-ACC in wind string_bag-NOM.SG.F with
*pačkoj pel'menej*
pack of_dumplings

‘Outside the window (...), a string bag with a pack of dumplings was beating against the wall in the wind.’ (M. Traub. *Zamočnaja skvažina*, 2012)

Although the definition of the Reflexive Anticausative alternation in the ValPaL claims that the verb is “predominantly used in the perfective”, in the RNC we found examples in which the reflexive imperfective form *bit'sja* was used with an anticausative meaning, as shown in (19).

## 3. Conclusions

In this paper we presented the revision process of the data contained in the ValPaL Russian dataset. The aim of the analysis is to create a more complete and consistent dataset for Modern Standard Russian to be included in the PaVeDa database. The revision of the dataset involved every level of the original ValPaL dataset, including the levels of the verb meaning, of the corresponding Russian verb form, of the basic coding frame and of alternations.
For each level of the dataset we provided examples of the changes we made in order to obtain a dataset that not only better describes the Russian valency patterns from an intra-linguistic perspective but, considering the main goal of PaVeDa, also ensures the most reliable cross-linguistic comparison.

We showed that, while in some cases we decided to add information, in others we removed data or we modified the existing ones. Although several other adjustments are still needed, the revision process we carried out represents the first step toward a more accurate and useful version of the Russian dataset.

## References


Apresjan, V. J. 2004. “Bojat'sja”. In Jurij D. Apresjan (ed.), *Novyj ob”jaznitel'nyj slovar' sinonimov russkogo jazyka*. Vtoroe izdanie, ispravlennoe i dopolnennoe. Vena: Jazyki slavjanskoj kul'tury, Venskij slavističeskij al'manach, Studia philologica, 60-66.

Iordanskaja, L. and Mel'čuk, I. 1990. Semantics of Two Emotion Verbs in Russian: Bojat'sja ‘to be Afraid’ & Nadejat'sja ‘to Hope. *Australian Journal of Linguistics* (10): 307-57.

Kagan, O. 2013. *Semantics of Genitive Objects in Russian : A Study of Genitive of Negation and Intensional Genitive Case*. Dordrecht: Springer.

Korn, D. 1967. Genitive or Accusative after Negation in Russian. *The Slavic and East European Journal* 11(4): 442-449.

Krys'ko, V. B. 1997. *Russkij istoričeskij sintaksis russkogo jazyka. Ob"ekt i perechodnost'*. Moskva: Indrik.
Kuznetsova, J, and Nesset, T. 2015. In Which Case Are Russians Afraid? Bojat'sja with Genitive and Accusative Objects. *Journal of Slavic Linguistics* 23(2): 255-283.
Malchukov, A. 2015. "Valency classes and alternations: parameters of variation". In A. Malchukov & B. Comrie (eds.), *Valency Classes in the World's Languages. Vol. 1*, 73-130. Berlin: Mouton de Gruyter.
Malchukov, A and Jahraus, A. 2013. *Russian*. In Hartmann, I. & Haspelmath, M. & Taylor, B. (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
Partee, B H., Borschev, V, Paducheva, E. V., Testelets, Y. and Yanovich, I. 2011. Russian Genitive of Negation Alternations: The Role of Verb Semantics. *Scando-Slavica* 57(2): 135–159.
Rachilina, E. V. 2004. There and back: the case of Russian 'go'. *Glossos* (5): 1-33.
Slobin, D. I. 2004. The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Strömqvist & L. Verhoeven (eds.), *Relating events in narrative: Vol. 2. Typological and contextual perspectives*, 219-257. Mahwah, NJ: Lawrence Erlbaum Associates.
Švedova, N. J. 1980. *Russkaja grammatika (2)*, Sintaksis, Moskva: Nauka.
Talmy, L. 2000. *Toward a cognitive semantics, vol. II: Typology and process in concept structuring*, Cambridge (MA): MIT Press.
Timberlake, A. 2004. *A reference grammar of Russian.* Cambridge: Cambridge University Press.

# Coding Issues in Turkish

*Zeynep Erk Emeksiz*[*]
*A. Sumru Özsoy*[**]
*Maria Chantal Seghedoni*[***]

## 1. Introduction

This paper discusses the coding issues regarding the Turkish basic counterparts of some verbs on the PaVeDa database and their alternations. As a morphologically rich agglutinative language, Turkish marks arguments for case and verbs for person and number agreement with the subject, and as such, employs both flagging and indexing as the basic argument-coding strategies. Our data set has revealed that Turkish differs from some other languages in a number of ways with respect to the nature of the basic coding frames (CF) as well as the alternations of some verbs, thus raised a number of issues regarding both the coding of the basic frame as well as of alternations.

Some of the issues Turkish data present regarding argument coding are:

1. Lexical entries versus alternations. Some pairs of verbs have one member morphologically derived from the other. Are they separate lexical entries or alternations? The transitive counterparts of the intranstive *korkmak* 'fear', *batmak* 'sink', *ölmek* 'die' are respectively *korkutmak* 'frighten', *batırmak* 'sink[tr]', *öldürmek* 'kill'. Considering them as alternations would entail thinking about a way to make the microroles consistent and rewriting the coding frames.
2. An unsure etymological reconstruction of the root in *yanmak* 'burn' and *yakmak* 'burn [tr]' posed problems regarding which of the two, if any, should be regarded as the basic form.

[*] Anadolu University
[**] Boğaziçi University
[***] University of Pavia

3. Coding complex verbs, making a distinction between compound verbs (with the light verbs *etmek*/*olmak*) and verbs with an incorporated argument (*yağmur yağmak* 'rain', *göz kırpmak* 'blink', *çığlık atmak* 'scream', *şarkı söylemek* 'sing'). Amongst these, a further differentiation, which in turn has determined the way we have decided to code them, was made between those verbs which can become transitive, their object argument having the possibility to behave independently by either taking a possessive or the accusative case suffix (*göz kırpmak* and *şarkı söylemek*), and those which do not tend to allow such behavior (*çığlık atmak*).
4. Valency decreasing operation passive and valency increasing operation causative are both coded and productive.
5. Other coded valency decreasing operations licensed are the impersonal, reflexive, and reciprocal alternations. Case marking of the P-argument of the base form of the verb determines the nature of the valency decreasing operation - passive vs impersonal.
6. Case marking of the second argument of the verb base form determines the case marker assigned to the initial 1-nom in causative.
7. Multiple occurrences of some valency alternation operations- double passive, double (and at a lesser frequency, triple) causative, have an effect on the valency of the derived verb as well as of the case marking of the intermediary agents. Examples of triple causative have not yet been attested in the present data set.
8. Causative precedes passive when both apply to the same verb stem.
9. Omission of subjects and contextually recoverable arguments is uncoded.

## 2. Basic valency patterns

### *2.1. Verb classes and argument-coding*

Morphologically, we distinguish between two types of verb stems in Turkish (i) mono-morphemic, *bilmek* 'know', *kırmak* 'break, (ii) polymorphemic, *korkutmak* 'frighten', *anlatmak* 'explain', *giydirmek* 'dress'. The distinction is based on the presence/absence of the causative morpheme on the verb stem where those bearing the causative suffix correspond to lexemes listed individually in the PaVeDa data set.

Complex verbs, e.g. ad *vermek* 'name', *şarkı söylemek* 'sing (a song)' are further distinguished as those (i) headed by a light verb *aç olmak* 'be hungry',

*takip etmek* ‘follow’, (ii) with a cognate argument *yağmur yağmak* ‘rain’, *yemek yemek* ‘eat’ and (iii) with an incorporated argument *çığlık atmak* ‘scream’, şarkı söylemek ‘sing (a song)’.

## *2.2. Flagging*

Of the 7 case markers; nominative, genitive, accusative, dative, and ablative, are considered as core cases by some researchers (Göksel & Kerslake 2004: 67, 154), while comitative marker (homophonous with instrumental) is not.

In the basic coding frame, A(gent) role is generally linked to the Nominative argument, P(atient) to the Acc(usative) argument. Instrument and Locative adjuncts marked [-lA] and [-DA] respectively occur with verbs of different coding frames.

The verbs are marked with the bound person marking associated with Nominative.

Predominantly a verb final language, Turkish licenses overtly case marked arguments to freely occur in any preverbal and postverbal position with no effect on valency. Indefinite/nonspecific arguments occur in bare form and are typically restricted to the immediately preverbal position.

## *2.3. Valency alternations*

Both valency decreasing and valency increasing operations. Four types of valency decreasing operations are licensed: passive [-Il][1], impersonal passive [-Il], reflexive [-(I)n][2], reciprocal [-(y)Iş].

The valency increasing operation is causative [-DIr][3]. Most valency alternations are coded. A verb can also be marked for multiple alternations such as multiple occurrences of passive and causative suffixes as well as suffixes of both valency increasing and decreasing alternations. In such cases, causative precedes the passive suffix.

1. The phonologically conditioned allomorphs of the passive morpheme are [-In] after a lateral-final stem, [-n] after a vowel-final stem, [-Il] elsewhere.
2. The allomorphic variation of the reflexive morpheme is [-n] after a vowel ending stem *yıkan* ‘wash oneself’, [-In] elsewhere.
3. The allomorphs of the causative are: [-t] after stems that end in a vowel, [-ar] after certain roots, *çık-ar* ‘take out’, [-ut] after *kork-ut* ‘frighten someone’, [-DIr] elsewhere.

## 3. Mono-morphemic verbs

### *3.1. Basic coding properties*

Based on P-argument marking, verbs fall into four classes: (i) Accusative [–(y)I] Ps, (ii) Dative [–(y)A] Ps, (iii) Ablative –Dan Ps, (iv) P Comitative –(y)lA. Classes (ii), (iii) and (iv) behave similarly with respect to the coding of the valency alternation operations.

*(i) Verbs with Accusative P-arguments.*

*bilmek* 'know' and *kırmak* 'break' are two such verbs which are linked by person and number indexing to the semantically agent/experiencer Nominative argument.

***bilmek*** 'know' – Basic CF: 1-nom 2-accV.subj[1] (1-knower; 2-known thing/person)

| | | | |
|---|---|---|---|
| *(1)* | *Sen* | *her şey-i* | *bil-iyor-du-n* |
| | 2SG.NOM | everything-ACC | know-PROG-PST-2SG |
| | 'You knew everything.' | | |

*kırmak* 'break' marks its P-argument Acc, licenses the Instrument adjunct *-lA*.

**kırmak** 'break' – Basic CF: 1-nom 2-acc (3-instr) V.subj[1] (1-breaker; 2-broken thing; 3-breaking instrument)

| | | | |
|---|---|---|---|
| (2) | *Ceviz-ler-i* | *el-in-le* | *kır-a-ma-z-sın.* |
| | walnut-PL-ACC | hand-POSS.2SG-INSTR | break-ABIL-NEG-AOR-2SG |
| | 'You cannot break the walnuts with your hands.' | | |

*(ii) Verbs with Dative marked P-argument*

*vurmak* 'hit', *bakmak* 'look' mark their P-argument Dative and are linked to the semantically agent Nominative argument.

***vurmak*** 'hit' – Basic CF: 1-nom 2-dat (3-instr) V.subj[1] (1-hitter; 2-hit thing; 3-hitting instrument)

| | | | |
|---|---|---|---|
| (3) | *Biri* | *kafa-m-a* | *çekiç-le* |
| | someone | head-POSS.1SG-DAT | hammer-INSTR |
| | *vur-du*. | | |

hit-PST(3SG)[4]
'Someone hit me in the head with a hammer.'

***bakmak*** 'look'
Basic CF: 1-nom 2-dat (3-instr) V.subj[1] (1-looker; 2-looked at entity; 3-look instrument)

(4) *Son kez o-na bak-tı-m.*
last time s/he/it-DAT look-PST-1SG
'I looked at him/her/it for the last time.'

*(iii) Verbs with Ablative marked P-argument*

*hoşlanmak* 'like' marks its M-argument Ablative -DAn and is linked to the semantically experiencer Nominative argument.

***hoşlanmak*** 'like' – Basic CF: 1-nom 2-abl V.subj[1] (1-liker; 2-liked entity)

(5) *Sen-den hoşlan-ıyor-um.*
2SG.ABL like-PROG-1SG
'I like you.'

*(iv) Comitative marked P-argument*

*görüşmek* 'meet' marks its P-argument Comitative -lA and is linked to the semantically agent Nominative argument.

***görüşmek*** 'meet' – Basic CF: 1-nom 2-com V.subj[1] (1-meeter; 2-met person)

(6) *11 parti temsilcisi-yle görüş-tü-m.*
11 party representative-COM meet-PST-1SG
'I met with 11 party representatives.'

## *3.2. Definiteness/specificity and P-marking*

Indefinite/nonspecific P-arguments of Accusative verbs occur in bare form, typically restricted to the immediately preverbal position. (7b) with *görmek* 'see', typically marking definite/specific P-argument ACC (7a), illustrates this.

***görmek*** 'see' – Basic CF: 1-nom 2-acc V.subj[1] (1-seer; 2-seen entity)

4. 3SG is zero marked in Turkish, indicated as (3SG) in verb final position in glosses.

(7) a. Danilo-yu film set-i bit-tiğ-i-nden
Danilo-ACC film set-CM end-NOM-3POSS-ABL
beri hiç gör-me-di-m.
since never see-NEG-PAST-1SG
'(…) Ever since the film set has ended, I have not seen Danilo.'

(7) b. Bakın gazete-ler-e, bir tane Izmirspor
look newspaper-PL-DAT one piece Izmirspor
haber-i gör-üyor mu-sunuz?
news-CM see-PROG Q-2PL
'Look at the newspapers, do you see a single piece of İzmirspor news?'

### *3.3. Valency alternations*

The P-marking of the verbs affects the coding properties of valency alternation operations. Verbs that mark P ACC conform to the properties of the derived coding frames of the valency decreasing operation Passive common cross-linguistically, while DAT, ABL, COM marking verbs do not. These form impersonal passive constructions where the verb is marked with -Il (homophonous with Passive), but the P argument retains its case marker. No *tarafından* 'by' phrase is licensed.

In the valency increasing operation CAUSE, the case marking of P of the base verb affects the case of its 1-nom argument. The ACC- P verbs mark the initial 1-NOM as 1-DAT, DAT/ABL/COM -P verbs mark it as 1-ACC.

#### 3.3.1. Valency decreasing operations

*(i) Passive*

The P-argument of the ACC verbs is marked 2-nom, the initial 1-nom optionally occurring as the object of the tarafından 'by' phrase (1-nom+*tarafından*).[5] The verb bears the passive -Il and the bound person marking associated with P marked Nominative.

***bil-in-mek*** 'be known' – Derived CF: 2-nom(1-nom+tarafından)(3-instr) V'pass.subj[2]

(8) *Merkez Banka-sı rezerv-ler-i-nden satış-ın*
central bank-CM reserve-PL-*CM-ABL* *sale-3GEN*
*kamu* banka-lar-i aracılığ-ıyla
public bank-PL-CM mediation-INSTR

5. Pronouns as the complement of *tarafından* 'by' are marked GENITIVE.

*yap-ıl-dığ-i* *bil-in-iyor*
do-PASS-NOM-POSS.3SG know-PASS-PROG(3SG)
'It is known that (…) the sale from the Central Bank reserves was done with the mediation of the public banks.'

*(ii) Impersonal passive*

Intransitive verbs, transitive verbs with an incorporated object as well as verbs with a second argument marked DAT, ABL, COM can occur in impersonal passive constructions where there is no overt subject and the verb is marked with the suffix *-Il*, homophonous with the personal passive. In the latter case, the second argument retains its case marker. The phonologically null single argument is interpreted as a 'generic, unspecified argument'. (Özkaragöz 1986)[6] Impersonal passive is a coded alternation for which we propose the derived coding frame [V'pass]. It differs from the personal passive coding frame in that there is no agentive phrase with the postposition *tarafından* and no agreement on the verb, which is 3SG by default.

(9) is an impersonal passive construction headed by the single argument verb *öl* 'die':

***ölmek*** 'die' – Derived CF: [V'pass]

(9) *Öl-meden* *öl-ün-ür* *mü* *bu* *dünya-da?*
die-CVB die-PASS-AORQ this world-LOC
'Does one die before dying in this world?'

(10) is the impersonal passive construction headed by the P-ABL verb *korkmak* 'fear'.

(10) *Marmara* *Ada-sı-nda* *hayat* *felç* *ol-du,*
Marmara Island-CM-LOC life paralysis be-PST
*salgın* *risk-i-nden* *kork-ul-uyor.*
epidemic risk-CM-ABL fear-PASS-PROG
'Life on the Marmara Island is paralyzed, everyone is afraid of the risk of epidemic.'

*(iii) Reflexivization*

6. Özkaragöz (1986) states that the verb is generally marked with the aorist marker. She also discusses different types of impersonal constructions in Turkish which are headed by a transitive verb: (i) double passive constructions headed by a monosyllabic transitive verb already marked with the passive suffix, (ii) those where the second passive marker occurs after the necessitative *-Abil* suffix preceded by the passive marker on the verbal stem. We have not come across any examples of these two types of double passive constructions in our data set yet, although both types are relatively frequent in daily use and literary texts.

A number of verbs marked with the suffix *-In* and are indexed with 1-nom as the subject omit P-argument. We propose the derived coding frame as 1-nom V'refl.subj[1].

***örtmek*** 'cover' – Derived CF: 1-nom V'refl.subj[1]

(11) *Zaten o dizi-de de biri*
in_fact that series-LOC EMPH someone
*baş-ı-nı ört-tü diye kimse*
head-3SG.POSS-ACC cover-PST because no_one
*ört-ün-me-ye kalk-ma-z-dı.*
cover-REFL-NMZ-DAT dare-NEG-AOR-PST
'In fact, no one would dare to cover themselves just because someone in that series covered their head.'

*(iv) Reciprocal*

Reciprocal functions as a valency reducing operation in a construction where the subject and the argument marked with the Comitative suffix *-lA* reciprocate the event expressed by the verb and where the Com argument is optionally missing. A relatively small number of verbs can be marked with the reciprocal suffix *–(y)Iş* which is indexed to agree with 1-nom as the subject. We propose the derived coding frame as [1-nom 2-com V'rec.subj[1].

***bakmak*** 'look' – Basic CF: 1-nom 2-dat V.subj[1]
Derived CG: [1-nom] 2-com V'rec.subj[1]

(12) *Ekran-da bir adet Redbullkutu-su*,
screen-LOC one piece Redbullbox-CM
*Badem-'le bak-ış-ıyor.*
Badem-COM look-REC-PROG
'On the screen, a box of Redbull and Badem are looking at each other.'

### 3.3.2. Valency increasing operations

The valency increasing operation coded and productive causative. The valency of the base verb as well as that of the class of verbs taking P-argument determines the marking of the 1-nom subject of the base verb: i. 1-nom of intransitive base verbs (*koşmak* 'run' (13)) as well as of DAT,ABL,COM transitive verbs (*bakmak* 'look' (14)) is marked ACC, ii. 1-nom subject of the ACC-P base verb is marked DAT (*yüklemek* 'load' (15)).

***koşmak*** 'run' – Basic CF: 1-nom (5-abl) (4-dat) V.subj[1]
Derived CF: 3-nom 1-acc (5-abl) (4-dat) V'caus.subj[3]

(13) *Çocuk-lar-ı* *dershane-den* *dershane-ye* *koş-tur-an*
child-PL-ACC class-ABL class-DAT run-CAUS-PTCP
*yeni* *bir* *sınav* *yap-ıl-ma-yacak*
new a test make-PASS-NEG-FUT
'No new tests will be given that makes children run from one classroom to another.'

***bakmak*** 'look' – Basic CF: 1-nom 2-dat V.subj[1]
Derived CF: 4-nom 1-acc 2-dat V'caus.subj[4]

(14) *Bütün* *sınıf-ı* *ot-lar-ın,* *çayır-lar-ın*
all class-ACC grass-PL-GEN meadow-PL-GEN
*üst-ü-ne* *yat-ır- ıp,* *gökyüzü-ne*
top-POSS.3SG-DAT lie-AOR-CVB sky-DAT
*bak-tır-ıyor-du-m.*
look-CAUS-PROG-PST-1SG
'Having the whole class lie down on the grass and meadows, I was making them look at the sky.'

***yüklemek*** 'load' – Basic CF: 1-nom 2-acc 3-dat V.subj[1]
Derived CF: 5-nom (1-dat) 2-acc 3-dat V'caus.subj[5]

(15) *Ora-dan* *yiyecek* *madde-ler-i* *yükle-t-ti-ler.*
there-ABL food supply-PL-ACC load-CAUS-PST-3PL
'They had food supplies loaded from there.'

### 3.3.3 Multiple alternations

Multiple alternation constructions are causative-passive, causative-causative, reciprocal-cause.

*(i) Causative-passive*
***görüşmek*** 'meet' – Basic CF: 1-nom 2-com V.subj(1)
Derived CF: 1-nom 2-com(3-nom+*tarafından*) V'caus.pass.subj[1]

(16) Ezel Başaran burada Ö.Y.
Ezel Başaran here.LOC Ö.Y.
ad-ı-nda-ki komiser ile
name-CM-LOC-PART chief_police with
görüş-tür-ül-dü.

meet-CAUSE-PASS-PST
'Ezel Başaran was made to meet with a police chief called Ö.Y. here'

*(ii) Multiple causative*

The verb can be marked with multiple causative suffixes.
***yüklemek*** 'load' – Basic CF: 1-nom 2-acc V.subj[1]
Derived CF: 5-nom (1-dat) 2-acc 3-dat V'caus.caus.subj[5]

(17) *Srirasmi* *tüm* *bu* *eşya-lar-ı* *kendisi-ne*
Srirasmi all this item-PL-ACC self-DA
*eşlik* *ed-en* *15* *otomobillik*
accompany do-PTCP 15 automobile
*konvoy-u-na* *yükle-t-tir-di*
convoy-POSS.3SG-DAT load-CAUS-CAUS-PST(3SG)
'Srirasmi, who went shopping in a limousine, had all these items loaded into the 15 cars convoy that accompanied her.'

*(iii) Reciprocal-causative*
***bakmak*** 'look' – Basic CF: 1-nom 2-dat V.subj[1]
Derived CF: 4-nom 1-acc 2-com V'rec.caus.subj[4]

(18) (…) *mânâ-yı* *harf-î* *ile* *bak-ış-tır. (…)*
meaning-ACC letter-POSS.3SG with look-REC-CAUS
'(…) have the meaning and the image look at each other (…).'

## 4. Polymorphemic verbs

Some verbs on the PaVeDa list are polymorphemic. These are (i) verbs with bound roots and (ii) V/ V+CAUSE pairs. The members of the latter exhibit differences in their basic coding frames and valency alternations.

### *4.1. Verbs with bound roots*

The verbs which share the same base form not attested as a free form are *yanmak* 'burn (inchoative)'/*yakmak* 'burn (tr)' and *öğretmek* 'teach'/(*öğrenmek* 'learn', not on the PaVeDa list).

#### 4.1.1. Basic coding frames

***yanmak*** ‘burn [intr]/***yakmak*** ‘burn [tr]’ [7]
The inchoative verb *yanmak* ‘burn’ is morphologically an unattested root **ya* and the reflexive *-n-*. The microrole can either be the fire, a burning thing (i.e. a torch, a candle) or a burnt thing. In the latter case, the meaning is often that of ‘burnt down’, reinforced by adverbs such as *tamamen* ‘completely’ or *cayır cayır* ‘furiously’.
For the verb BURN[tr] is *yakmak*, determining the origin of the non-productive suffix *-k-*, attested only in this verb, is problematic. From the stem yak-, the passive and causative passive coded alternations can be derived. At this stage, we tend to consider *yakmak* and *yanmak* two different lexical entries.
**yanmak** ‘burn’ – Basic CF: 1-nom V.subj[1] (1=fire, burning thing, burnt thing)

The semantic range of the single argument of the inchoative *yanmak* ‘burn’ is ‘fire’, burning thing’, and ‘burnt thing’.

(19) *Güzelim* *orman* *bir gece-de* *cayır*
beautiful forest(NOM) one night-LOC furiously
*ya-n-dı*
burn-REFL-PST(3SG)
‘The beautiful forest furiously burnt to the ground in one night.’

The transitive *yakmak* ‘burn’ licenses a burning *ateş* ‘fire’ as its P-argument, as its subject, and burn instrument, burnt thing as P-argument.
***yakmak*** ‘burn (tr)’
Basic CF: 1-nom 2-acc (3-instr) V.subj[1] (1-burner (tr), 2-burnt/burning thing (tr), 3-burn instrument (tr))

(20) *Yılan-ı* *öl-dür-mek* *için* *ateş* *yak-tı-m.*
snake-ACC die-CAUS-INF for fire burn-PST-1SG
‘I lit a fire to kill the snake.’

*öğretmek* ‘teach’ is the causative member of the bound root *öğre-* which, when marked with the reflexive suffix [*-(I)n*], yields the verb ‘learn’ (not on the PaVeDa list).
***öğretmek*** ‘teach’
Basic CF: 1-nom 2-dat (3-acc) V.subj[1] (1-teacher, 2-teachee, 3-taught content)

7. Determining the origin of the non-productive suffix *-k-*, attested only in this verb, is quite problematic. Clauson (1972: 941) regards it as emphatic.

(21) *Bu* *havuz-lar-da* *bugün-e* *kadar* *5milyon*
this pool-PL-LOC today-DAT until 5 million
*genç-imiz-e* *yüzme* *öğre-t-ti-k.*
youth-POSS.1PL-DAT swimming learn-CAUS-PST-1PL
'Until today, we taught 5 million young people how to swim.'

### 4.1.2. Valency alternations

***yanmak*** 'burn inchoative'/***yakmak*** 'burn (tr)'

Passive, causative and causative-passive coded alternations can be derived from the transitive stem *yak*-, but not the inchoative *yan*. Based on this crucial distinction, we consider *yakmak* and *yanmak* two different lexical entries.

*(i)* *Passive* (Derived CF: 2-nom (1-nom + tarafından) V'pass.subj[2])

(22) Akşam ateş yak-ıl-ır-dı
evening fire(NOM) burn-PASS-AOR-PST(3SG)
'A fire was lit in the evening.'

(23) Pilot (…) Bey tarafından yak-ıl-mış-tır
Pilot Mr by burn-PASS-REP-GMOD
'It was burned by Sergeant Pilot Vecihi (HÜRKUŞ) Bey (…)'

*(ii)* *Causative* (Derived CF: (4-nom) 2-acc (1-dat) V'caus.subj[4])

(24) 170 gemi haricinde-ki tüm gemi-ler-in-i
170 ship except-REL all ship-PL-POSS.3SG-ACC
yak-tır-dı
burn-CAUS-PST(3SG)
'She had all her ships except 170 burned (…) .'

*(iii)* *Causative passive* (Derived CF: 2-nom (4-nom+ *tarafından*) V'caus.pass'.subj[2])

(25) *Girolamo Savonarola* *tarafından* *yak-tır-ıl-mış*
Girolamo Savonarola by burn-CAUS-PASS-REP(3SG)
'(…) were caused to be burnt by Girolamo Savonarola.'

***öğret*** 'teach'

(i) *Passive* (Derived CF: 3-nom 2-dat (1-nom+tarafından) V'pass.subj[3])

(26) *Hasta-ya* *doğru* *duruş* *ve* *oturma* *pozisyon-u*
patient-DAT correct posture and sitting position-ACC
*öğret-il-ir.*
teach-PASS-AOR(3SG)
'The patient is taught correct posture and sitting position.'

## *4.2 V/V+CAUSE pairs*

Four pairs of verbs on the PaVeDa list correspond to one unmarked, the other member marked with the causative suffix: *korkmak* 'fear'/ *korkutmak* 'frighten', *batmak* 'sink (intr)/*batırmak* 'sink (tr)', *öl* 'die'/*öl-dürmek* 'kill', *anlamak* 'understand'/*anlatmak* 'tell' (*anlamak* is not listed in the initial verb list, hence we treat them as individual verbs in our data analysis. These exhibit differences in basic coding frames and valency alternations.

### 4.2.1. Basic coding frames

***korkmak*** 'fear'/***korkutmak*** 'frighten'
The two members differ with respect to the number of their arguments.
*korkmak* 'fear' – Basic CF: 1-nom 2-abl V.subj[1] (1-fearer; 2-fear stimulus)

(27) *Örümcek-ler-den* *neden* *kork-uyor-sun?*
spider-PL-ABL why fear-PROG-2SG
'Why are you afraid of spiders?'

*korkutmak* 'frighten' – Basic CF: 1-nom 2-acc (4-instr) V.subj[1] (1-frightener; 2-frightenee; 4-frightening thing/instrument)

(28) *Biz-i* *tehdit-ler-iniz-le*
1PL-ACC threat-PL-POSS.2PL-INS
*kork-ut-a-ma-z-sınız!*
fear-CAUS-ABIL-NEG-AOR-2PL
'You cannot frighten us with your threats'

***batmak*** 'sink'/ ***batırmak*** 'sink (tr)'
*batmak* 'sink' – Basic CF: 1-nom V.subj[1] (1-sunken entity)

(29) *Halbuki* *gemi* *bat-ıyor*
however ship(NOM) sink-PROG(3SG)
'However, the ship is sinking.'

*batırmak* ‘sink (tr)’ – Basic CF: 1-nom 2-acc (3-dat) V.subj[1] (1-sink-(caus)er; 2-sinking entity; 3-sink location)

Transitive *batırmak* ‘sink’ is morphologically marked as the causative of inchoative *batmak* ‘sink’.

(30) 20 *Mayıs: "Eğer kamara-sın-a*
20 May if cabin-POSS.3SG-DAT
*git-me-z-se-m gemi-yi*
go-ABIL-NEG-COND-1SG ship-ACC
*bat-ır-acağın-ı söylü-yor*
sink-CAUS-NMZ-ACC say-PROG(3SG)
‘May 20: "He says he'll sink the ship if I don't go to his cabin.’

***ölmek*** ‘die’/***öldürmek*** ‘kill’

*ölmek* ‘die’ – Basic CF: 1-nom (3-abl) (4-loc) V.subj[1] (1-dieer; 3-cause of death; 4-dying location)

(31) *Yine kadın-lar yine emekçi-ler*
again woman-PL(NOM) again labourer-PL(NOM)
*öl-dü*
die-PST(3PL)
‘Again women, again labourers died.’

*öldürmek* ‘kill’ – Basic CF: 1-nom 2-acc (3-instr) V.subj[1] (1-killer, 2-killee, 3-killing instrument)

*Öldürmek* ‘kill’ is the causative of the inchoative *ölmek* ‘die’. As such, its subject is semantically the agent of killing.

(32) *Sen-i kendi el-ler-im-le*
2SG-ACC own hand-PL-POSS.1SG-INSTR
*öl-dür-ür-üm!*
die-CAUS-AOR-1SG
‘I will kill you with my own hands!’

***anlatmak*** ‘tell’ – Basic CF: 1-nom 3-acc (2-dat) V.subj[1] (1-teller; 2-tellee; 3-told content)

(33) *Açık yüreklilikle Alanya-yı ve Alanyalı-yı*
open heartedly Alanya-ACC and Alanyalite-ACC
*anlat-ır-dı.*
tell-AOR-PST(3SG)

'S/He used to talk about Alanya and Alanyalites open heartedly.'

### 4.2.2 Valency alternations

***korkmak*** 'fear'/***korkutmak*** 'frighten' can both be marked with the passive suffix [-*Il*]. However, the passive of the inchoative *korkmak* 'fear' is impersonal passive (39) in contrast to (40):
*korkmak* 'fear' forms impersonal passive, i.e. can not occur with the agentive *tarafından* 'by' phrase.

*(i)* *Impersonal Passive* (Derived CF: 2-abl V'pass.subj)

(34) *salgın* *risk-in-den* *kork-ul-uyor*
epidemic risk-CM-ABL fear-PASS-PROG(3SG)
'there is fear of an epidemic risk'.

***korkutmak*** 'frighten'
*(ii)* *Passive* (Derived CF: (2-nom) 1-nom+tarafından V'pass.subj[2])

(35) *Kıbrıs-ta-ki* *ve* *Türkiye-de-ki* *militarist*
Cyprus-LOC-REL and Türkiye-LOC-REL militarist
*güç-ler* *tarafından* *kork-ut-ul-uyor-lar*
force-PL(NOM) by fear-CAUS-PASS-PROG-PL
'They are intimidated by militarist forces in Cyprus and Türkiye.'

Given that *korkutmak* 'frighten' is the causative of *korkmak* 'fear', the latter does not undergo further causativization

***batmak*** 'sink'/ ***batırmak*** 'sink (tr)'
No instance of the impersonal passive of the inchoative *batmak* 'sink' has been attested in the data yet.

*(i)* *Passive* (Derived CF: 2-nom (1-nom+ tarafından) V'pass.subj[2])

(36) *bir savaş gemi-miz* *kendi jet-ler-imiz*
One war ship-POSS.1PL(NOM) own jet-PL-POSS.1PL(NOM)
*tarafından* *bat-ır-ıl-mış-tı*
by sink-CAUS-PASS-PST(3SG)
'One of our warships was sunk by our own jets (…).'

***ölmek*** 'die'/ ***öldürmek*** 'kill'

The impersonal construction of the verb is licensed in Turkish as exemplified in (9). The transitive *öldürmek* 'kill' can undergo passivization as well as double and triple causativization.

*(i)* *Passive* (Derived CF: 2-nom (1-nom+*tarafından*) (3instr) V'pass.subj[2]

(37) *Askerlerin bir kısm-ı kaç-tı, bir*
soldier-PL-GEN one part-POSS.3SG escape-PST one
*kısmı direnişçi-ler tarafından öl-dür-ül-dü*
part-POSS.3SG resister-PL(NOM) by die-CAUS-PASS-PST(3PL)
'Some of the soldiers escaped, some were killed by the resistance.'

*(ii)* *Causative* (DerivedCF:(5-nom)2-acc(1-dat)(3-instr) V'caus.subj[5])

(38) binlerce sakin-in-i öl-dür-t-tü
thousands inhabitant-CM-ACC die-CAUS-CAUS-PST(3SG)
'(…) he (…) and had thousands of its inhabitants killed.'

***anlat*** 'tell'

*(i)* *Passive* (Derived CF 3-nom (1-tarafından) V'pass.subj[3])

(39) *Açıklama-da hastalığ-ın geçmiş-i de*
explanation-LOC illness-GEN past-POSS.3SG too
*anlat-ıl-dı.*
explain-PASS-PST(3SG)
'In the presentation the past of the illness was also explained.'

*(ii)* *Causative* (Derived CF: 4-nom 1-dat 3-acc V'caus.subj[4])

(40) *Ben bu şiddet-i niye çocuk-lar-ım-a*
I this violence-ACC why child-PL-POSS.1SG-DAT
*anlat-tır-ayım?*
tell-CAUS-SUBJ
'Why should I have this violence told to my children?'

## 5. Complex verbs

With respect to the nature of the head verb and their arguments, we analyze complex verbs in three types: (i) headed by light verbs: *kuru olmak* 'be dry', *traş etmek* 'shave', (ii) with cognate arguments: *yemek yemek* 'eat', *yağmur yağmak* 'rain', (iii) with incorporated arguments: *çığlık atmak* 'scream', *şarkı söylemek* 'sing'.

### *5.1. Verbs headed by light verbs*

Complex verbs headed by *ol-* are stative (47), by *et-* are dynamic (48).

#### 5.1.1 Basic coding frames

***kuru olmak*** 'be dry' – Basic CF: 1-nom V.subj[1] (1-dry thing)

(41) *Çamaşır-lar-ın* *çoğ-u* *kuru* *ama*
laundry-PL-GEN much-POSS.3SG dry but
*gene* *de* *topla-mak* *için* *biraz*
still EMP collect-INF in_order_to a_little
*bekle-mek gerek.*
wait-INF necessary
'Most of the laundry is dry, but it is still necessary to wait a while in order to collect it.'

The argument of the verbs headed by light verb olmak do not take the accusative case.

***tıraş etmek*** 'shave' (Basic CF: 1-nom 2-acc (3-instr) V.subj[1] (1-shaver; 2-shaved body part; 3-shaving instrument)

The reflexive alternation of the compound verb *traş etmek* is *traş olmak*.

(42) *Kaze',* *çocuğ-un* *baş-ın-ı* *tıraş*
Kaze'(NOM) child-GEN head-POSS.3SG-ACC shave
*ed-ip* *saç-lar-ın-dan* *bir*
do-CVB hair-PL-POSS.3SG-ABL one
*kısm-ın-ı* *bırak-mak-tır*
part-POSS.3SG-ACC leave-INF-GMOD
Kaze'means shaving the child's head and leaving some of her hair.

*(i)* *Passive* (Derived CF: 5-nom (1-nom + tarafından) V'pass.subj[5])

(43) *Güvey* *taraf-ın-a* *yastık* *gel-dik-ten*
Bridegroom side-POSS.3SG-DAT pillow come-NMZABL
*sonra* *güvey* *ve* *Mısayiv* *genç*
after bridegroom(NOM) and Mısayev(NOM) young
*erkek* *arkadaş-lar-ı* *tarafından* *tıraş* *edilir*
male friend-PL-POSS.3PL(NOM) by shavedo-PASS-AOR(3PL)
After the pillow comes to the bridegroom's side, the bridegroom and Misayev are shaved by their young male friends.'

*(ii)* *Reflexive* (Derived CF: 1-nom V.subj[1])

The reflexive form of traş etmek is formed by replacing the light verb etmek with olmak yielding traş olmak. The meaning corresponds also to the new meaning in the list SHAVE (self) [shave-self], reenforcing the question as to whether to treat certain verbs as alternations or as separate lexical entry.

(44) *Erkek-ler* *işlem-den* *hemen* *önce* *tıraş*
man-PL(NOM) procedure-ABL immediately before shave
*ol-ma-malı*
be-NEG-NEC(3PL)
'Men should not shave immediately before the procedure.'

## *5.2. Compound verbs with cognate arguments*

*yağmur yağmak* 'rain' and *yemek yemek* 'eat' take an indefinite/nonspecific cognate argument (subject of *yağmur yağmak* 'rain' and object of *yemek yemek* 'eat'), unmarked for case and typically incorporated into the verb. When the cognate object of *yemek yemek* 'eat' is definite, i.e. marked with possessive and/or ACC, the argument is not incorporated. We treat only the weather verb *yağmur yağmak* 'rain' here.

### 5.2.1 Basic coding frame

*yağmur yağmak* 'rain' – Basic CF: 1-nom (2-dat) V.subj[1] (1-rain, 2-rain goal)

The verbal element of *yağmur yağmak* 'rain' can be used with nouns, such as *kar* 'snow', *dolu* 'hail' to refer to other meteorological phenomena.

(45) *Her gün kar ve yağmur yağ-ıyor*
every day snow and rain rain-PROG
'It snows and rains every day.'

### 5.2.2 Valency alternation

We have come across only the causative alternation of *yağmur yağmak* 'rain' until now, although multiple causative and causative-passive alternations are also possible.

*(i)* *Causative* (Derived CF: 3-nom (2-dat) 1-acc V'caus.subj[3])

(46) *yani yağmur yağ-dır-arak (...)*
in.other.words rain rain-CAUS-CVB (…)
*ürün ver-me-sin-i sağla-r-dı*
produce give-NMZ-POSS.3SG-ACC ensure-AOR-PST(3SG)
'In other words, (…) by making it rain, (he would) enable (the plants) to produce.'

## *5.3. Verbs with incorporated arguments*

### 5.3.1 Basic coding frame

***çığlık atmak*** 'scream' – Basic CF: 1-nom V.subj[1] (1-screamer)

(47) Kız-lar çığlık at-ıyor-lar ama
girl-PL(NOM) scream throw-PROG-PL but
*ilgilen-en yok*
care-PTCP not.exist
'Girls are screaming but no one cares.'

It is possible for the incorporated argument to be preceded by the indefinite article *bir*, but not to take the accusative case:

(48) *Bekle-di-m ki biz-im rehber bir çığlık*
wait-PST-1SG COMP we-POSS.1PL guide a scream
*da biz-im için at-sın*
PART we-POSS.1PL for throw-SUBJ
'I waited for our guide to scream for us, (…).'

***göz kırpmak*** ‘blink’ – Basic CF: 1-nom (3-dat) 4(-acc) V.subj[1] (1-blinker; 3-winked at person; 4-eye)

(49) *Akşam-a* *kadar* *göz* *kırp-ıp* *dur-du*
evening-DAT until eye blink-CVB stop-PST(3SG)
‘S/he kept blinking until the evening.’

Only when marked by a possessive suffix which makes a noun definite, the internal argument is marked with the accusative case.

(50) *saatlerce* *göz-ün-ü* *kırp-madan*
for hours eye-POSS.3SG-ACC blink-CVB
*ufk-a* *bak-ar-dı”* *de-r*
horizon-DAT look-AOR-PST(3SG) say-AOR
‘He used to (…), looking at the horizon without blinking’, he says.

***şarkı söylemek*** ‘sing’ – Basic CF: 1-nom 2(-acc) V.subj[1] (1-singer; 2-song)

In reference to a specific song, P-argument is marked ACC (60). In this case too, therefore, we have put the argument in the accusative in brackets.

(51) *Billie* *5* *yaş-ın-da* *şarkı*
Billie(NOM) 5 age-POSS.3SG-LOC song
*söyle-me-ye* *başla-dı*
say-INF-DAT start-PST(3SG)
‘Billie started singing at the age of 5.’

(52) *Ferah* *ise* *sahne-nin.* *kenar-ın-da.* *şarkı-yı*
Ferah(NOM) instead scene-GEN edge-POSS.3SG song-ACC
*söyl-üyor*
sing-PROG(3SG)
‘Ferah is singing the song at the edge of the stage.’

### 5.3.2 Valency alternations

***çığlık atmak*** ‘scream’

*(i)* *Impersonal passive* (Derived CF: V’pass)

(53) *Bu* *süre-de* *Denizli 1-0* *öne* *geç-ince*
this period-LOC Denizli 1-0 ahead pass-CVB

*sevinç* *çığlık-lar-ı* *at-ıl-dı*
joy scream-PL-CM throw-PASS-PST(3SG)
'During this period, screams of joy were emitted when Denizli took the lead 1-0.'

***şarkı söylemek*** 'sing (a song)'

*(i)* *Passive* (Derived CF: 2-nom (1-nom + tarafından) V'pass.subj[2])

(54) *Çalışma* *zevk* *ve* *isyan-la* *coş-an*
work pleasure and rebellion-INS exult-PTCP
*kalabalık-lar-ın* *şarkı-sı* *söyle-n-meli-dir*
multitude-PL-GEN song-CM sing-PASS-NEC-GMOD
'The song of the multitudes excited by work, pleasure and rebellion must be sung.'

*(ii)* *Impersonal passive* (Derived CF: V'pass)

(55) *Bayan-lar-dan* *kurulu* *özel* *bir* *müzik* *grub-u*
woman-PL-ABL composed special a music group CM
*eşliğ-in-de* *şarkı-lar* *söyle-n-ir,*
accompaniment-CM-LOC song-PL say-PASS-AOR
*dans-lar* *ed-il-ir-di*
dance-PL do-PASS-AOR-PST
'Songs were sung and dances were performed accompanied by a special musical group made up of women.'

*(iii)* *Causative* (Derived CF: 3-nom 1-dat (2-acc) V'caus.subj[3])

(56) *öğrenci-ler-in-e* *şarkı* *söyle-t-erek*
student-PL-POSS.3SG-DAT song say-CAUS-CVB
*rahatla-ma-lar-ın-ı* *sağla-yan (…)* *kahraman*
comfort-NMZ-PL-POSS3-ACC provide-PTCP hero
*ol-du*
become-PST(3SG)
'(A teacher) who made his students relax by singing (…) became a hero.'

*(iv)* *Benefactive (uncoded)* (Derived CF: 1-nom 4-dat (2-acc) V.subj[1])

(57) *bebeğ-e* *şarkı-lar* *söyl-üyor-Ø*
baby-DAT song-PL sing-PROG(3SG)
'She was (…) singing songs to the baby...'

## 6. Concluding comments

The above discussion is based on the data in the ValPal list for Turkish and as such reflects the analysis of the valency properties and the coding issues in Turkish as attested in that data set up to this point. There are a number of other alternations (middle, antipassive) of some of the verbs that are not reflected in the above discussion. This is partly due to the fact that we have not yet come across them in that data set yet or that such constructions are not reflected in that data set but are more frequent in other texts. Our search and analysis continue.

## References


Biktimir, T. 1986. Impersonal Passives and the -ArAk construction in Turkish. In D. I. Slobin and K. Zimmer (eds), *Studies in Turkish Linguistics*, 53-76. Amsterdam: John Benjamins.

Csato, E. and L. Johanson, 2021. Turkish. In L. Johanson & E. Csato (eds.), *The Turkic Langıuages*, 203-235. London and Cambridge: Routledge.

Dikmen, F., S. Karakaş, A. S. Özsoy. 2023. Türkçe'deki Orta ve Nedensi Yapılar Üzerine bir İnceleme. In Hale Işık-Güler, Duygu Özge, Pınar Turan (eds.), *Dilbilimde Güncel Tartışmalar*, 121-140. Ankara: Dilbilim Derneği Yayınları.

Erk-Emeksiz, Z. 2023. Valency Patterns of Polysemous Verbs in Turkish. Paper presented at 21st International Conference on Turkish Linguistics at Johannes Gutenberg Universitat in Mainz.

Göksel, A. and C., Kerslake. 2004. *Turkish A Comprehensive Grammar*. London and Cambridge: Routledge.

Knecht, L. 1986. Lexical Causatives in Turkish. In D. I. Slobin and K. Zimmer (eds.), *Studies in Turkish Linguistics*, 93-122. Amsterdam: John Benjamins.

Nakipoğlu, M. Türkçe'deki ayrık geçişsiz eylemlerin olay yapısal bir incelemesi. *Dilbilim Araştırmaları*, 1-16. İstanbul: Boğaziçi Üniversitesi Yayınları.

Özkaragöz, İ. 1986. Monoclausal Double Passives in Turkish. In D. I. Slobin and K. Zimmer (eds.), *Studies in Turkish Linguistics*, 77-92. Amsterdam: John Benjamins.

Özsoy, A. S (ed.). 2018. Derivation of Fronting in Turkish. In S. A. Özsoy (ed.), *Word Order in Turkish*, 139-177. Amsterdam: Springer.

Özsoy, A. S. 2009. Argument Structure, Animacy, Syntax and Semantics of Passivization in Turkish: A corpus-based approach. In Y. Kawaguchi, M. Minegishi &

J. Durand (eds.), *Corpus Analysis and Variation in Linguistic*s, 259-279. Amsterdam: John Benjamins.
Öztürk, B. 2005. Case, Referentiality and Phrase Structure. *Linguistik Aktuell 77*, Amsterdam: John Benjamins.

# Argument Coding Phenomena in Hungarian: Issues and Challenges in Building a Database for the PaVeDa Project

*Rákosi György*[*]

## 1. Introduction

This paper reports on the ongoing work in creating the Hungarian database for the PaVeDa project, with special attention to a selected array of language specific issues that may also have relevance beyond Hungarian. Hungarian has relatively free word order, but its morphology is rich. Valency alternations are typically coded on the verb, and non-subject arguments receive designated case markers. This rich morphology opens a window onto a world of fine-grained variation in argument coding that may be gently disguised in languages with less articulated morphology. In section 2 of this paper, I comment on some of the ensuing issues and challenges that arise during the study of argument structure phenomena in general and the building of the database in particular. In section 3, I discuss another characteristic feature of Hungarian, the frequent licensing of implicit arguments of different types. The paper focuses on the issue of how much of the variation discussed in the two central chapters is to be stored in the database in view of the overall aims of the PaVeDa project. Section 4 concludes and rounds up this discussion.

## 2. Fine-grained variation in flagging and microrole assignment: A case study

Hungarian has 3 major types of adpositional elements: case markers (1a), case-like postpositions (1b) and case assigning postpositions (1c):[1]

[*] University of Debrecen

1. The abbreviations used in the glosses are as follows:

(1)

| | | | |
|---|---|---|---|
| a. | *a* | *híd-on* | |
| | the | bridge-SUP | |
| | 'on the bridge' | | |
| b. | *a* | *híd* | *mellett* |
| | the | bridge | next.to |
| | 'next to the bridge' | | |
| c. | *a* | *híd-on* | *át* |
| | the | bridge-SUP | across |
| | 'across the bridge' | | |

Case-like postpositions take a noun phrase complement that bears no case morphology, whereas case-assigning postpositions govern a particular case marker on their complement. The postposition *át* 'across', for example, requires superessive case ('on') on the noun phrase. As expected, superessive case is also licensed on its own (1a). A recent overview takes the number of case markers in Hungarian to be 17 (Dékány & Hegedűs 2021), though this number may vary according to what definitional criteria are assumed.

Rákosi (2021) observes that flagging on core arguments is dominantly achieved via case markers, and postpositions are rarely selected for this purpose. This is especially true of the core verbal vocabulary addressed in the PaVeDa project and the Leipzig Valency classes project. Since the number of case markers available for coding arguments is relatively large in Hungarian, flagging may show considerable versatility, also in terms of the mapping onto microroles. I discuss here the illustrative case of *segít* 'help'.[2]

The microroles associated with the verbal meaning HELP in the PaVeDa database are *helper*, *helpee*, *help causer*, and *helping instrument*. Most languages included in the database employ one case marker or an adposition to code the *helpee*, typically accusative or dative case. The Hungarian *helpee*, however, can receive either accusative, dative or superessive case. I have collected the forthcoming examples from the Hunglish Corpus, a parallel Hungarian-English corpus.[3] I quote the Hungarian examples together with their English equivalents as they appear in the corpus.

1: first person, 2: second person, 3: third person, ACC: accusative case, DAT: dative case, DEFOBJ: definite object conjugation, DEL: delative case ('from'), INE: inessive case ('in'), INF: infinitival suffix, IMP: imperative, INST: instrumental case ('with'), NOM: nominative case, PAST: past tense, POS: possibility suffix, POSS: possessive morphology on the possessum, SUP: superessive case ('on'), SUBL: sublative case ('onto').

2. The citation form of Hungarian verbs is the 3SG present tense form, given that this is usually the unmarked slot in the paradigm.

3. The Hunglish Corpus is available at http://hunglish.hu.

The accusative-marked *helpee* has to be an animate entity (2a and 2c), though action nouns also appear occasionally with this case (2b). The pronominal objects are *pro*-dropped in the original examples, an option licensed by definite object indexing morphology on the verb. I add the missing pronouns in brackets for expository purposes.

(2) a. *Saul* *pedig* *segít-i*
Saul.NOM and help-DEFOBJ.3SG
*(ő-t)* *a* *szökés-ben.*
he-ACC the escape-INE
'Saul's helping him escape.'

b. *A* *sósav* *segít-i* *az*
the HCL.NOM help-DEFOBJ.3SG the
*emésztés-t,...*
digestion-ACC
'The hydrochloric acid is good for the digestion; ...' (lit.: 'it helps digestion')

c. *A* *férfi* *fel-segít-et-te*
the man up-help-PAST-DEFOBJ.3SG
*(ő-t)* *a* *föld-ről.*
her-ACC the ground-DEL
'He drew her onto her feet.' (lit. 'helped her (rise) up from the ground')

Accusative-marked *helpees* identify participants that are involved in the core event denoted by the verb, and which are typically affected in the sense of undergoing some observable change of state. (2a), for example, describes an event where the subject Agent (*Saul*) provides direct and continuous support to the object argument. This is typical of accusative *helpees* in general, that is, they require a *helper* subject argument that causally interacts with them during the helping event. We may consequently describe this participant an *affected helpee*.

The *helpee* can also receive dative case, as happens in the examples below:

(3) a. *Akkor* *egy* *kicsi-t* *segít-ek* *nek-ik.*
then a little-ACC help-1SG DAT-3PL
'Then I will help them.'

b. *Dors* *segít* *nekem* *a*
Dors.NOM help.3SG DAT.1SG the

*felesleges* *adat-ok* *kiszűrés-é-ben.*
redundant datum-PL filtering-POSS-INE
'Dors is helpin' me filter out the garbage.'

c. *...segít-ek* *neki* *egy* *ügy-ben.*
help-1SG DAT.3SG a case-INE
'...I'm doing a little work for him.' (lit. 'I help him in a case.')

d. *Lát-t-am,* *mi-vel* *segít-het-ek* *neki.*
see-PAST-1SG what-with help-POS-1SG DAT.3SG
'I saw the way to help her.' (lit. 'I saw what I could help her with.')

The dative *helpee* is always animate, and is preferred in events where the helping relation between *helper* and *helpee* is less direct. (3d), for example, may be true if the *helpee* is not even aware of the fact that (s)he is being helped. I call this type of *helpee* a *recipient helpee*.

A third option is also available via the superessive case marker -*Vn* 'on':[4]

(4) a. *Kérlek, segít-s* *rajtam.*
please help-IMP.2SG on.me
'Please help me.'

b. *De* *ez-en* *segít-het-ünk.*
but this-on help-POS-1PL
'We're going to fix that.'

c. *Rajtam* *már* *nem* *segít-het-ett* .
on.me yet not help-POS-PAST.3SG
*senki*
nobody.NOM
'There was no help for me.'

There is no animacy constraint on this type of *helpee* (see 4b). In general, a superessive-marked *helpee* tends be less involved in and less affected by the core event denoted by the verb. Let us call this type the *beneficiary helpee*.

As the Hunglish translations from the corpus illustrate, the English *help+object* construction may be used in each case, but a *helpee* in Hungarian can be either of the three types we have just identified: *affected helpee* (accusative case), *recipient helpee* (dative case) or *beneficiary helpee* (superessive case). It is not immediately obvious which coding frame is the basic

4. The pronominal paradigm involves a suppletive form all through, that is why the case marker itself does not appear on *rajtam* 'on me'.

one, given that each seems to be relatively frequent. The frequency data in Table 1. below pertain to the frequency of the case markers that appear within the local domain of the embedding clause in the Hungarian National Corpus. These data are quoted from the valency dictionary of Sass et al. (2010: 153), who argue that this measure is a good approximator of valency patterns in Hungarian.

| Accusative | Dative | Superessive | Inessive | Instrumental | Frequency |
|---|---|---|---|---|---|
| + | - | - | - | - | 4874 |
| - | - | + | - | - | 2312 |
| - | - | - | + | - | 2235 |
| - | - | - | - | - | 1831 |
| - | + | - | - | - | 1510 |
| + | - | - | + | - | 1139 |
| + | - | - | - | + | 950 |
| - | + | - | + | - | 425 |
| + | - | - | + | + | 288 |

Table 15. *The raw frequency of valency patterns of the verb* segít *'help' in the Hungarian National Corpus (Sass et al. 2010: 153).*

Table 1 shows that the accusative is most frequent on dependents of *segít* 'help', but superessive and dative are also relatively frequent. There are no truly significant differences between the three types of *helpees* in terms of frequency of use.

Therefore, taking any one of these *helpees* to be representative of the basic coding frame seems to be a somewhat arbitrary decision if frequency is taken to be the defining measure. It is, nevertheless, important that information concerning this variation *is* coded in the database – what is not immediately obvious is how this information is to be stored. We may regard the variation among these three *helpee* types as realisational variation within the same coding frame, identifying only sub-types of the microrole *helpee*, or as alternation among different coding frames, possibly resulting in distinct microroles. Note in this respect that the 3 types of *helpees* cannot be co-realised, as the data in Table 1 also suggest. This could be possibly be interpreted as a (weak) argument against treating this variation in case marking as a change in valency. But the semantic differences among the three *helpee* types do seem to be substantial enough to assume that these patterns are distinct coding frames.

Notice that this morphological variation is not arbitrary, but follows lexicalization patterns identified by Dowty (1991) and subsequent research. The accusative marked *helpee* is characterised by what Dowty calls Proto-Patient

entailments (being an affected participant, undergoing a change of state), whereas the non-accusative *helpees* do not have these properties, or not necessarily.[5] Similar alternations are also attested with other verbs in the database. *Beat*, for example, either takes an accusative marked *beatee*, or one that bears sublative case ('onto'). The following two examples are from the Hungarian Gigaword Corpus (Oravecz et al. 2014).

(5) a. *Magamat* *üt-ött-em.*
myself-ACC beat-PAST-1SG
'I was beating myself.'

b. *... teljes* *erő-vel* *a* *karfára*
full strenght-INST the handrail-SUBL
*üt-ött-em*
beat-PAST-1SG
'I beat on/at the hand-rail with my full strength.'

This is an instance of what is generally known as the conative alternation, where the accusative variant entails that the *beatee* is affected and undergoes a change of state. The sublative-marked *beatee* is not necessarily affected in this sense, as it does not have to undergo any change of state. This variation therefore represents the same phenomenon that *segít* 'help' instantiates, and we may treat it as an alternation among different coding frames. It is certainly useful for future users of the database if they can learn about data concerning variation of this type, not only because it helps us understand cross-linguistic variation in how the respective verbal meanings are coded, but also because this may prove to be useful when we study diachronic changes affecting valency patterns of pertinent verb classes.

Coming back to the data in Table 1, notice that helpees can co-occur with inessive and instrumental case-marked PPs. I repeat (3b) and (3d) as (6).

(6) a. *Dors* *segít* *nekem* *a*
Dors.NOM help.3SG DAT.1SG the

5. Dowty's Corollary 2 to his Argument Selection Principle (1991: 576) runs as follows:

"With a three-place predicate, the nonsubject argument having the greater number of entailed Proto-Patient properties will be lexicalized as the direct object and the nonsubject argument having fewer entailed Proto-Patient properties will be lexicalized as an oblique or prepositional object ..."

This applies not only to the two internal arguments of three-place predicates, but also to the single VP-internal argument of verbs like *help*, if they show pertinent differences in argument selection.

*felesleges* *adat-ok* *kiszűrés-é-ben.*
redundant datum-PL filtering-POSS-INE
'Dors is helpin' me filter out the garbage.'

b. *Lát-t-am,* *mi-vel* *segít-het-ek* *neki.*
see-past-1SG what-with help-POS-1SG DAT.3SG
'I saw the way to help her.'

The instrumental PP is an expression of the *helping instrument* microrole, but the inessive PP-type is not yet included in the PaVeDa database. We may name the pertinent microrole as *help domain* if we want to treat it as an argument. It is certainly frequent enough, at least with *affected* and *recipient helpees* (see Table 1). That the distribution of *helping instruments* and *help domains* is restricted to two types of *helpees* is yet another argument for treating the three types of *helpees* as distinct microroles, and as representatives of different coding frames. What is perhaps less obvious is why we should treat PPs bearing the *helping instrument* and *help domain* roles as arguments. I turn now to the discussion of this and related issues in the next section.

## 3. Argument omission and implicit arguments

Hungarian is a language where subject *pro*-drop is licensed by rich agreement morphology on the verb. And while this is not a valency related phenomenon, I still find it useful to include a related change as an alternation in the case one of the verbs in the database. This verb is *esik* 'rains', which means 'fall' otherwise.[6] If this verb appears without an overt subject (7a), then it can only mean 'it rains' (or it may also mean 'it falls', but we focus on the weather verb here). With an overt subject, however, the verb can denote the fall of any other type of precipitation (7b).

(7) a. *Esik.*
fall.3SG
'It rains.'

b. *Esik* *az* *eső* / *a* *hó.*
fall.3SG the rain.NOM the snow.NOM
'It is raining/snowing.'

6. There are also other languages in the PaVeDa database where the verb for RAIN means 'fall', Xârâcùù being one example.

Since this variation affects the selectional restrictions concerning the interpretation of the subject, it can be regarded as a change in coding frames, and an alternation that is worth registering. I treat it as a specific type of impersonal alternation, one that restricts the meaning of the subject if it is not overt.

Hungarian also has object *pro*-drop, licensed by the definite object conjugation paradigm of the verb. This roughly targets definite objects.[7] Consider the following examples with the verb *énekel* 'sing' for illustration.

(8) a. *Kati* *énekel.*
Kati.NOM sing.3SG
'Kati is singing.'

b. *Kati* *énekel* *egy* *dal-t.*
Kati.NOM sing.3SG a song-ACC
'Kati is singing a song.'

(9) a. *Kati* *énekl-i* *a* *dal-t.*
Kati.NOM sing-DEFOBJ.3SG the song-ACC
'Kati is singing the song.'

b. *Szép* *a* *himnusz,* *ha* *Kati*
beautiful the Anthem.NOM if Kati.NOM
*énekl-i.*
sing-DEFOBJ.3SG
'The Anthem is beautiful when Kati sings it.'

The default verbal paradigm is selected if the verb is intransitive (8a), or if the object is indefinite (8b). Definite objects trigger the definite object conjugation (9a), which licenses object *pro*-drop. Whether the pronoun object is spelled out in examples like (9b) is a matter of information structure, and thus object *pro*-drop represents a discourse-governed phenomenon that has no direct bearing on valency.

The issue, however, is somewhat more complex as certain types of changes in the expression of object arguments are treated as valency changes in the PaVeDa database, which includes different types of object or oblique omissions for certain languages. The switch from (8b) to (8a), for example, could be regarded as *generic object omission*. The English example (10a) from the database is described as a construction containing an *understood*

7. Object *pro*-drop is in fact a more complex phenomenon, but this rough description is sufficient for our purposes. See Coppock (2022) for a recent in-depth discussion, and an overview of the pertinent literature.

*omitted object*, and the relation between (10a) and (10b) can therefore be considered an alternation in valency.

(10) a. Can you help?
b. Can you help me?

The most likely Hungarian equivalent will include a dative *recipient helpee*, which is frequently omitted in Hungarian in first and second person.

(11) *Tudsz* *(nekem)* *segít-eni?*
can.2SG DAT.1SG help-INF
'Can you help (me)?'

This sort of omission of obliques that refer to prominent discourse participants is ubiquitous among different classes of verbs in Hungarian, including many examples where the relevant argument needs to be spelled out in English. The verb *ad* 'give', for example, has the basic coding frame in (12a), but the dative *giving recipient* argument is regularly omitted in first and second person (12b).

(12) a. *ad* 'give' : 1-nom V.subj[1] 2-acc 3-dat
b. Ad-j (nekem) egy kanal-at!
give-IMP.2SG DAT.1SG a spoon-ACC
'Give me a spoon.'

The dative pronoun in examples like (12b) is in fact more frequently dropped than not. The valency dictionary of Sass et al. (2010: 19-20) quotes the following raw frequency figures for the dative and accusative arguments of *ad* 'give':

(13) lemma *ad* 'give' +acc: 10.921
lemma *ad* 'give' +dat +acc: 6.985
lemma *ad* 'give +dat: 1.209

The simultaneous spellout of the accusative and the dative argument is not the dominant pattern in statistical terms, as these data show. It is also an option to have a dative argument alone, with the accusative object omitted. Variation among these realisational options is not coded on the verb, and it does not affect the meaning of the verb itself. This variation also does not seem to affect the interpretation of the pertinent arguments. Given these facts,

and given the fact the discourse-licensed argument omission is very frequent in Hungarian, it is not motivated or useful to regard these changes as valency alternations in general, or in the specific context of the PaVeDa project in particular.

Optional non-core arguments represent a different, but also slightly related issue. Take the case of instrument PPs, which are prominently represented in the Leipzig Valency classes project. The following two examples are from the Database Questionnaire Manual:[8]

(14) a. The boy touched the snake with a stick.
b. The woman cut the bread with a sharp knife.

The expression of the instrument PP is optional in both cases, but while *cut* entails the existence of the instrument as a participant of the core event denoted by the verb, *touch* does not. In a technical sense, the change from (15a) to (15b) can be considered a valency augmenting alternation, and then we treat the *with*-PP in (15b) as a regular argument.

(15) a. The boy touched the snake.
b. The boy touched the snake with a stick.

But the relation between (16a) and (16b) is less obviously an instance of the same phenomenon, given that the existence of an instrument is entailed even in the case of (16a).

(16) a. The woman cut the bread.
b. The woman cut the bread with a sharp knife.

We may consequently regard (16a) as version of (16b) with the instrument argument omitted, and treat it as a case of a valency decreasing type of alternation. In other words, (16a) may be an instance of generic argument omission, while such a treatment does not seem to motivated for (15a). And though the authors of the language-specific databases represented in the PaVeDa project do not treat (16a) and similar examples as instances of argument omission, it is not immediately obvious why this approach should not be entertained for languages where the alternation between (16a) and (16b) correlates with no changes on the verb, i.e., where it is not coded. The same considerations apply to the Hungarian equivalents of (15) and (16).

8. Available at https://valpal.info/static/ValencyDBQuestionnaireManual.pdf.

Whether instruments are entailed by the verb (*cut*) or not (*touch*), they typically occur only rarely. I present corpus data in Rákosi (2013) to show that there are no significant differences in this respect among *cut*-verbs and other agentive verbs that may take instruments: verbs do not frequently take a *with*-instrument in either class. Consequently, examples of this type may be best restricted to the core cases specified in the Leipzig manual in languages like English or Hungarian, where these changes do not induce morphological changes on the verb. Similar considerations may apply to other non-core arguments, such as optional beneficiaries or comitatives.

## 4. Summary

In this paper, I have investigated argument coding phenomena with a focus on Hungarian, and with the aim of commenting on what pertinent changes may be worth representing in the PaVeDa database. Hungarian has rich case morphology, and I have shown via the example of *segít* 'give' that what seems to be the same *helpee* argument across languages may in fact be the instantiation of different types of *helpees*, each receiving a different morphological case marker in Hungarian. Understanding this variation and other similar data may help us better understand cross-linguistic variation in coding frames as well as diachronic changes in argument realisation. It is therefore useful to include variation of this type in the database, perhaps as instances of designated valency alternations.

Another characteristic feature of Hungarian is the frequent employment of different types of argument omission and the concomitant licensing of implicit arguments. What connects these cases is that they tend to be discourse driven, and they induce no specific morphological changes on the verb. While similar phenomena are occasionally included in the PaVeDa database as instances of valency alternations, I have argued, in line with the general spirit of the project, that it may be worth providing some representative examples only in specific cases, and avoid a more comprehensive coverage in general.

## References

Dékány, É. & V. Hegedűs. 2021. Postpositions: formal and semantic classification. In Katalin É. Kiss & Veronika Hegedűs eds. *Syntax of Hungarian. Postpositions and postpositional phrases*. Amsterdam: Amsterdam University Press. 11-192.

Dowty, D. 1991. Thematic Proto-roles and argument selection. Language 67 (3). 547-619.
Coppock, E. 2022. Object agreement in Hungarian. *Journal of Uralic Linguistics* 1 (1). 121-148.
Sass, B; T. Váradi, J. Pajzs & M. Kiss. 2010. Magyar igei szerkezetek. A leggyakoribb vonzatok és szókapcsolatok szótára. [Hungarian verbal constructions. A dictionary of the most frequent arguments and collocations.] Budapest: Tinta Könyvkiadó.
Oravecz, C.; T. Váradi & B. Sass. 2014. *The Hungarian Gigaword Corpus*. In: Proceedings of LREC 2014.
Rákosi, G. 2013. Down with obliques? In: Tracy, Holloway King; Valeria, de Paiva (eds.) *From quirky case to representing space: Papers in honor of Annie Zaenen*. Stanford (CA): CSLI Publications. 127-138.
Rákosi, G. 2021. PPs used as arguments. In. Katalin É. Kiss & Veronika Hegedűs (eds.) *Syntax of Hungarian: Postpositions and postpositional phrases*. Amsterdam: Amsterdam University Press. 285-354.

# Decreasing and Rearranging Alternations in Hindi

*Lucrezia Carnesale*[*]
*Andrea Drocco*[**]

## 1. Introduction

This paper focuses on coded valency-decreasing alternations (Malchukov 2015) in Hindi and the methodological issues they raise in their annotation within the Pavia Verbs Database (PaVeDa, Zanchi et al. 2022).

Hindi displays two decreasing alternations coded on the verb, the passive and the anticausative alternation. While both alternations reduce the number of arguments, they differ in how they profile the event. The passive alternation backgrounds the Agent but keeps it conceptually present, demoting it to an oblique adjunct. The anticausative alternation removes the Agent entirely and allows the event to be reinterpreted as spontaneous (Saksena 1982). Hindi displays some constructions that, while based on valency decreasing morphology, actually rearrange the argument-structure rather than decreasing it. The result is a construction in which, the verb morphology indicates a decrease in valency, but the first argument is not removed and appears in an oblique case maintaining a salient syntactic status.

Consider the following examples: in 1 the verb *dekhnā* 'look at, see' occurs in its basic coding frame (transitive), with the first argument marked by the ergative and the second by the accusative. In 2, the verb appears in its anticausative form *dikhnā* meaning 'appear, be seen, be visible'. Notably, the first argument is not removed and appears in the dative case. The overall semantics of the construction is not 'appear' but 'see' and yields a non-agentive reading (Carnesale 2024).

(1) *jamādār=ne* *use* *sir=se*

[*] University of Pavia
[**] University of Ca' Foscari

jamadar=ERG 3SG.ACC head(M.SG.OBL)=INS
*pāṁv=tak* *dekh-ā.*
foot(M.SG.OBL)=to look_at-PRF.M.SG
"Jamadar looked at him from head to foot." (Lit. Corp)

(2) *mujhe* *to* *tumhāre* *svāsthy=meṁ* *ab=hī*
1SG.DAT then 2PL.GEN health=in now=EMPH
*koī* *tabdīlī* *nahīṁ* *dikh-t-ī.*
IDF.ADJ change(F.SG.NOM) not appear-IPRF-F.SG
"I don't see any change in your health right know." (Lit. Corp)

The structure of this paper is as follows. Section 2 provides preliminary remarks on the Hindi verbal lexicon, discussing its division into simple verbs and complex predicates (CPs). Section 3 introduces the valency-decreasing alternations attested in the language, with section 3.1 examining the passive alternation and section 3.2 focusing on the anticausative alternation. In section 4, we turn to constructions that employ decreasing morphology in valency-rearranging rather than valency-reducing strategies. Section 4.1 is devoted to the so-called *passive of reluctance*, while section 4.2 discusses rearranging alternations based on the anticausative morphology, focusing on what we refer to as the *instrumental Agent alternation* and the *dative Experiencer alternation*. Lastly, in Section 5 we address the methodological challenges posed by the annotation of these alternations within PaVeDa and outline our proposed solutions. We focus on two main issues: first, the mismatches between Hindi valency orientation and the PaVeDa structure (sec. 5.1); and second, the under-representation of Hindi CPs in the PaVeDa meaning list, along with the implications this has for an accurate representation of valency patterns and alternations in the language (sec. 5.2).

## 2. Some remarks on Hindi verbal lexicon

Hindi verbal lexicon is divided into two groups: simple verbs and noun-verb CPs. In this study, we use the term *complex predicate* (CPs) to indicate predicates composed by a sequence of a non-verbal element and a verb (Masica 1991, Mohanan 1994, Butt 1995, Kulkarni 2011, Montaut 2016). These constructions are commonly referred to also as "conjunct verbs" in the literature on South Asian languages (Shapiro 1989, Mohanan 1994, Kachru 2006, Kulkarni 2011). The phenomenon is widespread in Hindi and encompasses a wide range of semantic classes of verbs. These predicates consist of a nominal host and a light verb that act as a single predicate unit, whose semantics is different from that of the light verb alone. The semantic core is

expressed by the nominal, while the verb provides information on TAM properties and voice. Consider sentences in 3: sentence a) shows the complex predicate *dikhāī denā* 'see' formed by the light verb *denā* 'give' and the noun *dikhāī* 'seeing', while b) shows the verb *denā* when used as a simple verb.

(3) a. *pannā =ko* *cārom̐_or* *andher-ā*
panna=DAT all_four_directions darkness(M)-SG.NOM
*hī* *dikhāī* *de-t-ā* *th-ā*
EMPH seeing(F)SG.NOM give-IPRF-M.SG be.PST-M.SG
"Panna could see only darkness all around." (Lit. Corp)

b. *karūnā* *gau-om̐=ko* *pānī*
karuna.NOM cow(F)-PL.OBL=DAT water(M.SG.NOM)
*de-t-ī* *hai*
give-IPRF-F.SG be.PRS.3SG
"Karuna gives water to the cows." (Lit. Corp)

The class of nominals that can function as hosts in Hindi is highly productive: any noun or adjective – and even some adverbs – can be used to form a CP. The class of verbs that can be used as light verbs is closed and only a dozen verbs can occur in this function. Among these, the verb *honā* 'be' and the verb *karnā* 'do' are the most frequent: the first is used to form intransitive predicates and the second to form transitive predicates. Among the other verbs that can be used as light verbs in CPs there are *ānā* 'come', *lagnā* 'adhere', *rahnā* 'stay', *paṛnā* 'fall', *uṭhnā* 'rise', which form intransitive predicates, and *denā* 'give', *lenā* 'take', *mārnā* 'hit', *ḍālnā* 'throw', *lagānā* "attach", which form transitive predicates. Both the nominal host and the verb can be used independently. In this paper, we will focus on CPs formed with the light verbs *karnā* 'do' and *honā* 'be', whose alternation typically serves as the primary means of realizing the anticausative alternation.

CPs constitute a significant part of the Hindi verbal lexicon (Mohanan 1994, Kulkarni 2011): they do not only occur in huge numbers, but they also express basic notions that do not have a corresponding simple verb. For example, notions like 'work' (*kām karnā*), 'love' (*pyār karnā*), 'wait' (*intazār karnā*) and 'remember' (*yād honā/karnā*) are expressed by CPs in Hindi. The high productivity of this phenomenon is also evident in the incorporation of verb borrowings into the lexicon. When a verb is borrowed from another language (such as English, Persian, Arabic, and Sanskrit), it becomes a part of the Hindi lexicon by forming a CP consisting of the borrowed term functioning as the nominal host, followed by a Hindi light verb. For instance, the verb *print karnā*, meaning 'print', is composed by the English borrowed noun *print* and the Hindi verb *karnā* 'do'. Similarly, the verb *mālūm honā* 'know' is formed by the Arabic adjective *mālūm*

'known' and the Hindi verb *honā* 'be' (Kachru 2006, Montaut 2004). Given the relevance of CPs in Hindi, it is crucial to consider the dual structure of the verbal lexicon while analyzing any phenomenon related to Hindi verbal morphology.

## 3. Coded decreasing alternations in Hindi

Hindi displays two valency-decreasing alternations coded on the verb, the passive alternation and the anticausative alternation. As mentioned above, the passive alternation maintains the causative nature of the event and allows the Agent to remain in the background (demoted to an adjunct status). The anticausative alternation, on the other hand, does not involve an Agent, and reinterprets the event as spontaneous: the action unfolds without external intervention. Compare 5 and 6, showing the passive (5) and the anticausative (6) form of the verb *toṛnā* 'break' with the transitive basic coding frame given in 4. In the following sections, we will discuss the passive and the anticausative alternations focusing on their form and their functions in the language.

(4) *maiṁ=ne* *cūḍiyā* *nahīṁ* *toḍ-ī,*
1SG=ERG bangle not break-PRF.F.SG […]
"I did not break the bangles […]' (Lit. Corp)

(5) *somnāth* *mandir* *mugal* *ākrāntā-oṁ* *dvārā*
somnath temple mughal invader.PL.OBL from
*kaī* *bār* *toḍ-ā* *ga-yā* *th-ā*
several time break-PRF.M.SG go.PRF-M.SG be.PST-M.SG
"The Somnath temple was destroyed several times by Mughal invaders." (hiTenTen21)

(6) *maiṁ=ne* *samjh-ā,* *ab* *us=se*
1SG=ERG understand-PRF.M.SG, now 3SG.OBL=COM
*sadā=ke_lie* *nātā* *ṭūṭ*
always=for relationship(M.SG.NOM) break(anticaus)
*rah-ā* *hai.*
PRGR-M.SG be.3SG.PRS
'I see that my relationship with him is breaking forever.' (Lit. Corp.)

### *3.1. Passive Alternation*

In the passive construction, the event is profiled as caused by an external force or agent that is backgrounded but remains conceptually present. The passive is typically used when the Agent is unknown, unspecified or

irrelevant. Hindi passives are formed using the verb's perfective participle, followed by the auxiliary verb *jānā* 'go', which agrees with the subject in gender and number. The demoted Agent may appear in the form of an oblique adjunct marked either by the compound postposition (*-ke*) *dvārā* – the postposition specifically used to mark the demoted Agent in standard passive– or, very commonly, by the postposition *-se*. The latter has a broader functional range in the language, serving also as an ablative, instrumental and comitative marker. In 8, an example of the passive alternation with the simple verb *jalānā* 'burn' is given in contrast with the basic coding frame in 7. While 10 illustrates the passive alternation with the complex predicate *śurū karnā* 'begin' compared with the basic coding frame in 9. In Hindi CPs, the passive alternation follows the same principles as with simple verbs: the light verb *karnā* 'do' takes the perfective participle form and is followed by the auxiliary verb *jānā* 'go'. As the examples show, the passive retains the causative reading and allows the Agent to remain implicit/demoted to an adjunct.

(7) *use* *vaidhavy=kī* *āg=meṁ* *na*
3SG.ACC widowhood=GEN fire=in not
*jalā-ūṁ-g-ā.*
burn-1SG-FUT-M.SG
"I will not burn her in the fire of widowhood." (Lit. Corp)

(8) *āg=se* *jalā-yā* *jā-ye-g-ā.*
fire=INS burn-PRF.M.SG go-3SG-FUT-M.SG
"It will be burnt by the fire." (Lit. Corp)

(9) *maiṁ=ne* *saphar=kī* *taiyārī* *śurū* *k-ī*
1SG=ERG journey=GEN preparation(F) start do-PRF.F.SG
'I started the preparation for the journey.' (Lit. Corp)

(10) *bhārtīy* *taknīkī* *evaṁ* *ārthik* *sahayog*
indian technical and economic cooperation
*kāryakram* *bhārtīy* *mantrimaṇḍal=ke* *nirṇay*
programme indian cabinet=GEN decision
*dvārā* *1964=se* *śurū* *ki-yā* *ga-yā.*
by 1964=from start do-PRF.M.SG go.PRF-M.SG
'The Indian Technical and Economic Cooperation Program was initiated by a decision of the Indian Cabinet in 1964.' (hiTenTen21)

### *3.2. Anticausative Alternation*

The anticausative alternation reduces valency by entirely removing the Agent, transforming a causal action into a spontaneous agentless event (see Malchukov 2015, Zúñiga and Kittilä 2019). Unlike the passive, where the

Agent is backgrounded but still conceptually present, the anticausative profiles the event as unfolding independently, without any implied external cause or initiator.

Compare 11, showing the anticausative form of the verb *jālnā* 'burn (tr)', with the basic coding frame and the passive form given in 7 and 8 in the previous section.

(11) *maiṁ* *āg=ko* *jal-t-e*
1SG.NOM fire=ACC burn(ITR)-IPRF-M.OB
*hu-e* *dekh-t-ā* *hūṁ,* *par* *nahīṁ*
be.PRF-M.OB see-IPRF-M.SG 1SG.PRS but not
*jān-t-ā* *ki* *kaise* *jal-t-ī*
know-IPRF-M.SG that how burn(ITR)-IPRF-F.SG
*hai?*
be.3SG.PRS
'I see the fire burning, but I don't know how it burns.' (Lit. Corp)

The realization of the anticausative alternation varies between simple verbs and CPs. In simple verbs, the alternation is marked by changes in the verb root, according to the class the verb belongs to (class (a) vs class (b)). In class (a), verbs mark the anticausative alternation via ablaut, whereby a contrast in vowel length or quality signals the difference between intransitive anticausative and transitive causative forms. In these cases, intransitive anticausative verbs typically exhibit a lax or short root vowel, while their transitive causative counterparts show a tensed or long vowel (as shown in the first column of Table 1).

When the root vowel is *-a*, the alternation is realized through vowel lengthening, as in *marnā* 'die' vs *mārnā* 'kill'. While in roots containing vowels other than *-a*, the alternation is marked by a change in vowel quality – sometimes accompanied by consonantal changes (see the alternation between the intransitive anticausative *biknā* 'be sold' vs the transitive causative *becnā* 'sell' or the intransitive anticausative *phūtnā* 'bursh' vs the transitive causative verb *phoṛnā* 'make burst' in Table 1, Class a).

In verbs belonging to Class b) the anticausative alternation is encoded through suffixation: transitive causative verbs contrast with intransitive anticausative verbs by means of a suffix *-ā*, which may be accompanied by the shortening of the root vowel, as in the verb pairs *hilnā* 'move (intransitive)' vs *hilānā* 'move (transitive)' (see verbs belonging to Class b in Table 1).

| Class a) ABLAUT | | Class b) SUFFIXATION | |
|---|---|---|---|
| *marnā* 'die' | *mārnā* 'kill' | *hilnā* 'move I' | *hilānā* 'move T' |
| *ṭūṭnā* 'break I' | *toṛnā* 'break T' | *baiṭhnā* 'sit' | *baiṭhānā* 'seat' |
| *ruknā* 'stop I' | *roknā* 'stop T' | *girnā* fall' | *girānā* 'drop' |
| *khulnā* 'open I' | *kholnā* 'open T' | *paṛhnā* study' | *paṛhānā* 'teach' |

Table 16. *The anticausative alternation of Hindi simple verbs.*

As the examples in Table 1 show, Hindi simple verbs are clearly valency-oriented and exhibit transitivizing strategies. This pattern has been pointed out in previous research, both in language-specific studies (Saksena 1982, Butt 2002) and in typological studies (Masica 1976, Haspelmath 1993). In transitivizing languages, intransitive verbs serve as the morphologically basic forms, typically simpler and less marked, while transitive verbs are morphologically derived from them (Nichols et al. 2004). Both series in Hindi show a transitivizing pattern. In Series b) the suffixed causative transitive form is morphologically more complex than the anticausative intransitive counterpart. In Series a) roots showing the vowel *-a* are clearly transitivizing, as the causative form is phonologically heavier, showing a vowel lengthening. Verb pairs where the root is other than *-a-* in series (a) raise the question of whether the vowel alternation should be interpreted as an oriented alternation as well, as in the alternation between *khulnā* 'open (intransitive)' vs. *kholnā* 'open (transitive)'. However, the directionality can be traced diachronically to Old Indo-Aryan root vowel lengthening, where middle/passive forms exhibited a weaker root vowel, while causatives showed an augmented grade through the addition of the vowel *-a-* (i.e., *guṇa, vṛddhi*, see Montaut 2011). In verbs such as *khulnā* 'open (itr)' vs *kholnā* 'open (tr)' or *dikhnā* 'appear' vs. *dekhnā* 'see', the transitive form is marked by a vowel resulting from the monophthongization of a diphthong that emerged via the addition of *-a-* to the intransitive root (e.g., $\bar{u} + a > o$, $i + a > e$).

In CPs, the anticausative alternation is realized through alternating the light verb *honā* 'be' with the light verb *karnā* 'do' (Table 2).

| Causative transitive | Anticausative intransitive |
|---|---|
| *śurū honā* 'start (I)' | *śurū karnā* 'start (T)' |
| *garam honā* 'be hot' | *garam karnā* 'warm' |
| *band honā* 'close(I)/be closed' | *band karnā* 'close(T)' |

Table 17. *Anticausative alternation of Hindi complex predicates.*

The light verb *karnā* 'do' forms causative transitives encoding events caused by external agents, while the light verb *honā* 'be' forms anticausative intransitives, encoding events occurring independently of any external causation. In 12,

an example of the verb *śurū honā* 'start (itr)' is given in contrast with the transitive causative (*śurū karnā* 'start (tr)') in 9 above.

(12) *dūsre din=se boāī śurū hu-ī.*
second day=from sowing start(F.SG.NOM) be.PRF-F.SG
'The sowing started from the second day.' (Lit. Corp)

The Hindi verbal lexicon thus displays two distinct valency orientations: whereas Hindi simple verbs show oriented transitivizing strategies, CPs are not oriented and exhibit indeterminate strategies (Nichols et al. 2004), in which neither member of the verb pair is morphologically derived from the other and alternations are realized via auxiliary change.

Both strategies play a crucial role in Hindi valency profile. The transitivizing patterns found in simple verbs represent a historically older system, inherited from Old Indo-Aryan. In contrast, CPs' indeterminate strategy represents a later innovation in the language, which gained prominence between the 14th and 16th centuries, after an intense Persian influence occurred during the Mughal Empire (Montaut 2016). This contact with Persian restructured the Hindi verbal system, leading to a significant lexical borrowing via CPs, and thus contributing to the rise of indeterminate strategies, marking a strong shift away from the oriented transitivizing patterns of Old Indo-Aryan.

## 4. Decreasing morphology in re-arranging strategies

Hindi exhibits certain constructions that, while based on valency-decreasing morphology, do not reduce valency in the way that standard passives or standard anticausatives do. Instead, these constructions rearrange the arguments, retaining them both, but with a marking different from that in the basic frame. This allows the argument structure to remain intact while re-profiling the event. Such alternations convey nuances of unintentionality, inefficacy, or reduced Agent control. In the following sections, we discuss these constructions, distinguishing between those based on passive morphology (sec. 4.1) and those based on anticausative morphology (sec. 4.2).

### *4.1. The passive of incapacity or reluctance*

The Hindi passive of reluctance is used to express actions that could not be performed due to the Agent's incapacity or unwillingness, typically arising from internal constraints such as emotional resistance or psychological

discomfort, rather than from external circumstances. See sentence 13, adapted from Montaut (2004: 208), in which the verb stands in the passive form and the Agent appears in the oblique case, marked with the postposition *se*. The interpretation of 13 is not that of a standard passive. Rather, it expresses the emotional and psychological inability and reluctance of the main participant to carry out the action. As Montaut (2014) points out, the implication is that the participant cannot bring themselves to observe the sorrow due to emotional incapacity.

| | | | | |
|---|---|---|---|---|
| (13) | *us=se* | *merā* | *gam* | *nahīṁ* |
| | 3SG.OBL=INS | my | sorrow(M.SG.NOM) | not |
| | *dekh-ā* | *ga-yā* | | |
| | see-PRF.M.SG | go.PRF-M.SG | | |
| | 'He/She couldn't bear to see my sorrow.' | | | |

This construction morphologically differs from the passive construction in two main ways. First, unlike in standard passives, where the Agent is backgrounded or omitted, here the Agent remains explicitly expressed, and the construction highlights his/her unwillingness to perform the action. Second, this construction occurs with both transitive and intransitive verbs (as in 14, adapted from Montaut 2004: 208), indicating that its primary function is not to promote the object, but to highlight the participant's internal struggle or inability to complete the action.

| | | | | | |
|---|---|---|---|---|---|
| (14) | *mujh=se* | *yahāṁ* | *baiṭh-ā* | *nahīṁ* | *ga-yā* |
| | 1SG.OBL=INS | here | sit-PRF.M.SG | not | go.PRF-M.SG |
| | 'I couldn't sit here (was unable to).' | | | | |

### *4.2. The anticausative in re-arrenging constructions*

Hindi exhibits two constructions that, although formed with anticausative morphology, do not actually reduce the verb's valency. These constructions profile the event as one that happens rather than one that is actively performed by an Agent. Nevertheless, the first argument remains syntactically present, typically appearing in the oblique case and marked with a postposition, implying that the event was not fully under the control of the participant, and that it happened unintentionally or with limited effectiveness. Which postposition marks the oblique argument depends on the semantic class of the verb, namely effective action verbs vs experiential verbs.

With effective action verbs**,** the first argument is marked with the instrumental *se*, indicating a non-intentional involvement of the Agent. This construction has been referred to as the *anticausative of inadvertence* (Montaut 2004). See 15, where the verb appears in the anticausative form *ṭūtnā* 'break (itr)' and the Agent, expressed by the first-person pronoun, is followed by the postposition *se* (*mujh=se* '1SG.OBL=INS'). The sentence describes the event as something that happens by mistake, because of the participant but unintentionally, rather than a fully agentless event.

(15) *galtī=se* *koī* *kāṁc=kī* *cīj*
mistake=INS INDF glass=GEN thing(F.SG.NOM)
*mujh=se* *ṭūṭ* *ga-ī*
1SG.OBL=INS break go.PRF-F […]
'If I clean (the house) and I break a glass by mistake, you will scold me.' (hiTenTen21)

A similar construction, sometimes referred to as the *anticausative of inefficiency* (Montaut 2004) is given in 16. The verb appears in its anticausative form, and the Agent is explicitly present, marked with the postposition *-se*. A key feature of this construction is the obligatory presence of negation. Despite being the negation of the *anticausative of inadvertence*, this construction takes on a completely different interpretation and indicated that the participant attempts the action but fails to carry it out effectively. The focus is on the Agent's incapacity and inability to complete the action, rather than on the removal of an intentional causal participant from the scene (as in 16).

(16) *baḍ-ā* *golak* *to* *us=se*
big-M.SG.NOM box(M)SG.NOM then 3SG.OBL=INS
*ṭūṭ-ā* *nahīṁ*
break(INTR)-PRF.M.SG not
'He could not break the bigger box. *He did not break the box by mistake' (hiTenTen21)

With experiential verbs, the first argument is marked with the postposition *ko*, that is the prototypical case marking Experiencers in Hindi (compare 17 with 18). In this construction, the first argument is construed as the Beneficiary/Recipient of the event expressed by the anticausative verb. Notably, that the Experiencer is an argument in this construction, and not an adjunct, is shown by the fact that it behaves as a non-canonical subject (Carnesale 2024). This construction serves to construe the experiential situation as spontaneous and profiles the Experiencer as non-volitional.

(17) *rameś* *is=kā* *koī* *javāb*
ramesh.NOM this.OBL=GEN INDF answer(M.SG.NOM)
*soc* *rah-e* *th-e*
think PRGR-PRF.M.PL be.PST-M.PL
'Ramesh was thinking of an answer to this.' (Lit. Corp)

(18) *use* *na* *kuch*
3SG.DAT not something(M.SG.NOM)
*sūjh-t-ā* *th-ā,* *na kuch sunāī detā thā.*
be_perceived-IPRF-M.SG be.PST-M.SG nor hear anything
'He could neither perceive/think nor hear anything.' (Lit. Corp)

In Table 3, we give a schematic summary of these constructions according to the semantic class they occur with.

| | | 1° arg | 2° arg | Verb |
|---|---|---|---|---|
| **Basic CF** | | 1-nom/erg | 2-nom/acc | V.subj[1] |
| **Derived CF** | Effective Action | 1-ins | 2-nom | anticausV'.subj[2] |
| | Experience | 1-dat | 2-nom | anticausV'.subj[2] |

Table 18. *Re-arranging constructions that use the anticausative morphology in Hindi.*

In sum, Hindi employs anticausative morphology not solely to mark standard valency-reducing alternations, but also to encode events that diverge from the transitive prototype. These constructions are used to profile the event as a state or as a spontaneously happening event in which the main participant is not a prototypical Agent. The anticausative intransitive construction is extended by an additional argument which is marked according to its semantic properties and may take subjecthood properties. Crucially, the use of anticausative morphology in rearranging strategies is especially productive with experiential verbs. Hindi exhibits a wide range of cognate verb pairs that distinguish between agentive and non-agentive construals of experiential events. This pattern applies to a wide spectrum of experiential predicates, as illustrated by pairs such as *yād karnā* vs *yād honā* 'remember', *dekhnā* vs *dikhnā* 'see', *socnā* vs *sūjhnā* 'think' *vicār karnā* vs *vicār honā* 'think'. Interestingly, these rearranging alternations are significantly more productive with CPs than with simple verbs in Hindi. A large number of experiential predicates participate in alternations involving the light verbs *honā* 'be' and *karnā* 'do', signaling a contrast between non-agentive and agentive construals of the same underlying event (see 19). In contrast, there seemes to be only two verbs participating in anticausative alternation involving experiential simple verbs (*dikhnā* "appear/see" and *sūjhnā* "be

perceived/feel"). This asymmetry suggests that the anticausative morphology in rearranging strategies is restricted to CPs.

(19) a. *maiṁ* *asāmi-yoṁ=kā*
1SG.NOM tenant-PL.OB=GEN
*satāyā_jānā* *pasand* *nahīṁ* *kar-t-ā*
persecution.M.SG.NOM liking not do-IPRF-M.SG
'I do not like the harassment of my tenants.' (Lit. Corp)

b. *mālī=kā* *kām* *use* *pasand* *th-ā.*
gardening=GEN job(M)NOM 3SG.DAT liking be.PST-M.SG
'She liked gardening.' (Lit. Corp)

## 5. Methodological issues

### *5.1. Mismatches between Hindi orientation and PaVeDa structure*

One key challenge in annotating Hindi valency patterns and alternations within PaVeDa arises from structural mismatches between the database's assumed valency orientation, which is based on English verbal lexicon, and the orientation we find in Hindi. As discussed in section 3, in the case of simple verbs Hindi consistently exhibits a transitivizing orientation, while in the case of CPs the language shows an indeterminate strategy, as the alternation is realized via auxiliary changes. The orientation of Hindi simple verbs generates some methodological issues while annotating the data, as PaVeDa does not always reflect the language-specific orientation. This situation results in the three main types of mismatches we discuss below.

1. Alignment of Hindi valency orientation with database structure: The basic meaning given in the database is intransitive anticausative and it matches the orientation of Hindi simple verbs. An example is the meaning BURN (intransitive in the database). In Hindi BURN (itr) is expressed by *jalnā* 'burn (itr)', while the causative *jalānā* 'burn (tr)' is morphologically derived by suffixation. In this case, we treated the intransitive *jalnā* as basic and the causative *jalānā* as derived.
2. Mismatch of Hindi valency direction and database structure: The PaVeDa meaning is causal, but in Hindi the transitive form is morphologically derived from an intransitive base. For example, the meaning CUT in PaVeDa has a basic causal reading, but in Hindi, the transitive causal verb *kāṭnā* 'cut' is derived via the lengthening of

the root vowel from the intransitive verb *kaṭnā* 'be cut'. In such cases, we decided to treat the intransitive form *kaṭnā* 'be cut' as the basic form even though the basic meaning in PaVeDa is causal.

3. Lexically related entries with shared roots in Hindi: In some cases, two distinct English verbs correspond to morphologically related forms in Hindi. The pair DIE and KILL is a clear example: *marnā* 'die' is the basic form, while *mārnā* 'kill' is derived through vowel lengthening. In such cases, we maintain two distinct lexical entries and we then cross-reference them.

### *5.2. Under-representation of Hindi CPs in PaVeDa meanings list*

While the PaVeDa meanings selection provides valuable insights for typological generalizations, it may not fully account for language-specific phenomena, as is the case with Hindi CPs. The choice of verb meanings largely influences our understanding of a language's basic valency profile. A limitation in the database arises from the under-representation of CPs, which as we discussed in section 2, are a core component of the verbal system in Hindi. Only 15 of the 80 meanings in the PaVeDa list are expressed by Hindi CPs. This gap raises two major issues:

1. The indeterminate strategy that characterizes CPs is not adequately represented, leading to an oversimplified picture of Hindi valency orientation and an over-representation of transitivizing strategies.
2. Productivity of valency-decreasing morphology in rearranging strategies (e.g., experiential verbs) is not well represented, since it seems to correlate with experiential CPs (sec. 4) which are under-represented.

To address these issues, we propose two complementary strategies. First, we explicitly acknowledge this limitation in the introductory section, highlighting the constraints imposed by the meanings list. Second, we supplement the database by including additional verb meanings that are expressed by CPs in Hindi. These include: (i) The effective action verbs START, CLOSE and HEAT addressing problem 1 and (ii) the experiential verbs REMEMBER, HOPE and DECIDE addressing problem 2.

## References

Butt, M. 1995. *The Structure of Complex Predicates in Urdu*. Stanford: CSLI.

Butt, M. 2002. *The morpheme that wouldn't go away*, talk presented at the Workshop on Pertinacity, Shloss Freudental.
Carnesale, L. 2024. Experiential constructions in Hindi: the encoding of perceptions and bodily sensations. *JSALL* 11: 155-187.
Haspelmath, M. 1993. More on the typology of inchoative/causative verb alternations. In Comrie B. and M. Polinky (eds.), *Causatives and Transitivity*, 87-120. Amsterdam: John Benjamins.
Kachru, Y. 2006. *Hindi*. Amsterdam: John Benjamin.
Kulkarni, A. 2011. Agreement in Hindi Conjunct Verbs. In *Proceedings of ICON-2011: 9th International Conference on Natural Language Processing*. India: Macmillan Publishers.
Malchukov, A. 2015. Valency classes and alternations: parameters of variation. In Malchukov A. Comrie B. (eds.), *Valency Classes in World's Languages, Vol. 1*, Berlin: De Gruyter Mouton.
Masica, C.P. 1991. *The Indo-Aryan languages.* Cambridge: CUP.
Mohanan, T. 1994. *Argument Structure in Hindi*. Stanford: CSLI.
Montaut, A. 2004. *A Hindi Grammar*. LINCOM EUROPA.
Montaut, A. 2011. Basic Intransitivity: A Typologically Relevant Feature for Indo-Aryan. In Koul O.N. (ed.), *Indo-Aryan Linguistics*, 29-46. Mysore: Central Institute of Indian Languages.
Montaut, A. 2016. Noun-Verb Complex Predicates in Hindi and the rise of non-canonical subjects. In P. Samvelian & L. Nash (eds), *Approaches to Complex Predicates*, 142-174. Amsterdam: John Benjamins.
Nichols, J., D. Peterson & J. Barnes. 2004. Transitivising and detransitivising languages. *Linguistic Typology* 8 (2): 149–211.
Saksena, A. 1982. *Topics in the Analysis of Causatives, with an Account of Hindi Paradigms*. Los Angeles: University of California Press.
Shapiro, M.C. 1989. *A Primer of Modern Standard Hindi*. Delhi: Motilal.
Zanchi, C., S. Luraghi & C.R. Combei. 2022. PaVeDa-Pavia Verbs Database: Challenges and Perspectives. In *Proceedings of the 4th Workshop on Research in Computational Linguistic Typology and Multilingual NLP* (SIGTYP 2022), 99-102. Seattle, Washington.
Zúñiga, F. & Kittilä, S. 2019. Changing Semantic Valency: Causatives, Applicatives, and Related Constructions. In F., Zúñiga & S., Kittilä (eds.), *Grammatical Voice*, 12-81. Cambridge: CUP.

 ISBN 9788835183631

# Vi aspettiamo su:

**www.francoangeli.it**

**per scaricare (gratuitamente) i cataloghi delle nostre pubblicazioni**

DIVISI PER ARGOMENTI E CENTINAIA DI VOCI: PER FACILITARE
LE VOSTRE RICERCHE.


**Management, finanza, marketing, operations, HR**
**Psicologia e psicoterapia: teorie e tecniche**
**Didattica, scienze della formazione**
**Economia, economia aziendale**
**Sociologia**
**Antropologia**
**Comunicazione e media**
**Medicina, sanità**



**Architettura, design, territorio**
**Informatica, ingegneria**
**Scienze**
**Filosofia, letteratura, linguistica, storia**
**Politica, diritto**
**Psicologia, benessere, autoaiuto**
**Efficacia personale**
**Politiche e servizi sociali**

FrancoAngeli

La passione per le conoscenze

 ISBN 9788835183631

FrancoAngeli
La passione per le conoscenze

 ISBN 9788835183631

## Materiali Linguistici

This volume presents the current state of the Pavia Verbs Database (PaVeDa), an open-source typological resource for the investigation of verb valency patterns and alternations from a cross-linguistic and diachronic perspective. The volume collects papers written by Italian and international early-career researchers who contribute to PaVeDa by developing datasets for new languages. Some papers in the volume discuss theoretical issues, focusing on argument structures and alternations in specific languages, while others address methodological challenges encountered during the annotation process and propose ways to overcome them. The selection of papers highlights the wide array of languages represented in PaVeDa, covering several linguistic families, including Uralic (Hungarian), Turkic (Turkish), and Indo-European languages – both modern (Hindi and Modern Standard Russian) and ancient (Vedic Sanskrit, Latin, Early Italo-Romance and the varieties of the British Isles). The volume seeks to offer an overview of the current results of the PaVeDa project and to inspire further typological and diachronic research on verb valency.

*Lucrezia Carnesale* is a postdoctoral researcher in Linguistics at the University of Pavia, where she works on the creation of the *Pavia-Verbs-Database*. She mainly works on New Indo-Aryan languages, focusing on Hindi argument structure constructions from a cognitive linguistics perspective.

*Martina Giuliani* is a postdoctoral researcher in linguistics at the University of Pavia, where she is contributing to the construction of the *Pavia-Verbs-Database* annotating data from Latin. Her research interests include Indo-European linguistics with a focus on Latin morphosyntax, cognitive linguistics, and language typology.

FrancoAngeli
La passione per le conoscenze