



BRILL

# Object Marking in Western Eurasia

## *The Circum-Baltic Area Dissolves into the Broader Areal Background*

*Daria Alfimova* | ORCID: 0000-0002-1289-4537

PhD Student, Department of Slavonic Studies, University of Potsdam,  
Potsdam, Germany

*daria.alfimova@uni-potsdam.de*

Received 15 August 2022 | Accepted 5 August 2024 |

Published online 18 September 2024

### Abstract

This paper investigates object marking strategies in Circum-Baltic languages and beyond, using a sample of 103 predicates from 30 Western Eurasian languages from the BivalTyp database. The study aims to identify areal clusters in object marking and evaluate the relevance of the Circum-Baltic linguistic area in this context. It finds that while most Circum-Baltic languages dissolve into larger, genealogy-driven clusters, areal signals are present, particularly with Lithuanian merging with Slavic languages due to genitive-taking predicates. German deviates from the larger Germanic cluster, merging with Latvian and Hungarian without a specific marking strategy driving this alignment. The results suggest that the concept of a linguistic area is less effective for describing object marking relationships around the Baltic Sea, in contrast to the Balkan linguistic area, where cross-linguistic clusters of object marking strategies do align with the linguistic area. Additionally, the paper discusses large-scale trends, such as comitativity prominence and the comparison of observed object marking strategy distributions with the predicted Zipf distribution.

### Keywords

object marking – Circum-Baltic languages – linguistic area – cluster analysis

## 1 Introduction

The ways in which verbs and argument marking strategies may be copied in language contact situations have recently enjoyed much scholarly attention (among others, Wichmann and Wohlgemuth, 2008; Malchukov and Comrie, 2015; Say, 2014, 2018; Tirps and Stein, 2019; Michaelis, 2019; Grossman et al., 2019; Grossman, 2021; Gaszewski, 2020). In addition, several notable typological projects on the areal distribution of argument marking strategies have emerged. For example, the database of valency patterns ValPal (Hartmann et al., 2013); a project on semantic role clustering (Bickel et al., 2014); the World Atlas of Transitivity Pairs (WATP, 2014); a database of bivalent verbs and their encoding frames BivalTyp (Say, 2020); a database of Slavic argument flagging (Seržant et al., 2022).

This study aims to identify clusters of object marking strategies in Western Eurasia, with a primary focus on the Circum-Baltic (CB) area. The latter is considered to be a linguistic area (or Sprachbund) in which certain isoglosses are shared not only by adjacent languages but also by non-adjacent languages across the area (Koptjevskaja-Tamm and Wälchli, 2001; to appear). There are 27 CB isoglosses as, for example, differentiated spatial cases (Koptjevskaja-Tamm and Wälchli, 2001), comitative-instrumental syncretism (Stolz, 2001), dative-like experiencer constructions (Seržant, 2015a). Nevertheless, defining clear boundaries for the CB linguistic area is problematic because the isoglosses “hardly can be said to bundle” (Koptjevskaja-Tamm and Wälchli, 2001: 624). So, it is still disputable whether the notion of linguistic area is really necessary for the languages spoken around the Baltic Sea. However, it is useful when describing specific traits; for example, the independent partitive is proven to be an Eastern-Baltic isogloss (Seržant, 2015b).

Object marking strategies of various predicates have never been consistently compared within and outside the Baltic area, although intuitively, this is precisely where language interference effects should manifest. One can think of many examples when object marking strategies in the CB languages correspond to each other. For instance, Russian and Latvian employ similar spatial prepositions ‘on, onto’ for the object of the predicate ‘to look (at)’, see (1) and (2). In contrast, Lithuanian uses a different spatial concept, ‘in, into’ (3), and Estonian uses a partitive case for the same predicate (4).

(1) Russian

*Petja smotrit na mor-e*  
 PN look onto sea-ACC.SG  
 ‘Petja looks at the sea.’

- (2) Latvian  
*Pēteris skatās uz jūr-u*  
 PN look onto sea-ACC.SG  
 'Peteris looks at the sea.'
- (3) Lithuanian  
*Peteris žiūri į jūr-q*  
 PN look into sea-ACC.SG  
 'Peteris looks at the sea.'
- (4) Estonian  
*Peeter vaatab merd*  
 PN look sea.PART.SG  
 'Peeter looks at the sea.'

This is just one of the possible combinations of similarities and differences between the selected four languages. There are numerous other combinations to consider, and the questions to be answered in the current paper are: Do the CB languages cluster based on object marking of bivalent predicates? And, if they do not form a homogeneous cluster, what kind of language groupings are to be found when looking at the CB area from a broader geographical perspective?

To make probabilistically solid claims about correlations in object marking strategies across the CB area, this study will provide evidence not only for a large number of predicates but also for different languages within and outside the CB area. This distinction is crucial for disentangling macro-areal effects from local contact phenomena (see Campbell, 2006: 19). In this study, this approach is supported by the quantitative methods advocated by contemporary areal linguistics (Wiemer, 2019).

In this vein, I compare object marking strategies of 30 languages from Western Eurasia, thereby approaching CB languages from a broader areal perspective. Drawing on data from BivalTyp, I use my own comparative concepts in line with Haspelmath (2010) and in the spirit of Levinson et al. (2003). A clustering method is then applied, followed by an exploration of the results in geographical space through the analysis of the distribution of various strategies. Signals of language contact are identified as deviations from expected patterns within genealogical clusters of languages. The findings indicate that both genealogical relationships and areal influences – whether purely areal or a combination of areal and genealogical factors – contribute to similarities in object marking among languages.

I will argue that CB languages split into relatively independent, genealogically homogeneous sets (Slavic, Germanic, Finnic) with some clear traces of language contact: Lithuanian–Slavic (in particular, Polish and East Slavic) and Latvian–German. The diffusion of object marking strategies within and outside the area of interest (see Campbell, 2017: 27) provides evidence for binary contact situations rather than the convergence effect in the Baltic region.

I proceed as follows: in Section 2, I describe the data and explain the conventions adopted in this paper. The clustering results are presented in Section 3 and elaborated upon in Section 4. Subsequently, Section 5 briefly discusses some large-scale patterns in the distribution of object marking strategies around Western Eurasia. Section 6 concludes the paper with the final discussion.

## 2 Data and Method

I use the data from the BivalTyp database of bivalent verbs and their encoding frames (Say, 2020). As of March 2022, it counts 90 languages primarily spoken in Western Eurasia and is based on a questionnaire comprising 130 predicates in sentences (in some cases, even in a broader context clarifying the meaning of the predicate). All the argument marking devices except transitive ones are documented in language-specific labels, for example, *NOM\_naACC* for the Russian predicate ‘to look at’ in (1). A transitive pattern is defined in the database as the pattern of the predicates ‘break’ and ‘kill’ in a given language.

My sample comprises 30 languages from three families – Indo-European (26), Uralic (3), Basque (1) – and covers the territory with Norwegian (Bokmål), Swedish and Finnish in the North, Modern Greek in the South, Irish in the West and Russian in the East. It includes most of the languages associated with the so-called Standard Average European (SAE) area (Haspelmath, 2001; Van der Auwera, 2011), the CB area (see above), and the Balkan area (Joseph, 1983; Friedman, 2006).

Due to incomplete data in BivalTyp (as of March 2022), I have chosen 103 predicates with complete or easily fillable argument coding information out of a total of 130 predicates. This is important since gaps would artificially boost the differences among the languages. For this reason, I have amended 43 language-specific patterns that were missing in the original BivalTyp data based on various sources, including consultations with native speakers. The list of all selected predicates and their patterns can be found in the supplementary material.

TABLE 1 Types of comparative constructions replacing the language-specific object marking strategies

Formal constructions	A-form (A-like marker) O-form (O-like marker) INTRANSITIVE WITH, WITHOUT
Spatial constructions	ACROSS, ALONG, AROUND, AT/TO, BEHIND, FROM, IN/INTO, INFRONT, ON/ONTO, OVER, UNDER
Semantic constructions	ABOUT, AGAINST, FOR, POS (possessive)

While cross-linguistic comparison of transitive argument marking strategies is rather unproblematic (cf. Bickel et al., 2014; Haspelmath, 2015), the non-canonical, or non-transitive, strategies presents greater challenges. The problem is that the language-specific labels – such as instrumental or locative – do not cover the same functional space across languages. They are, however, “similar in the relevant respect,” as Haspelmath (2010: 666) puts it. Accordingly, I have classified all second arguments of the verbs according to the set of comparative constructions – understood in terms of Haspelmath’s (2010) comparative concepts – as given in Table 1.

The comparative constructions are assigned hierarchically: if a language-specific marker fits into more than one construction type, it is classified under the construction type that appears first in the table.

Within **the formal constructions**, which are the highest in the hierarchy, there are three types: transitivity, intransitivity and comitativity constructions. Transitivity constructions are defined according to the arguments of a transitive verb (Haspelmath, 2015: 136). Agent-like arguments (A-form) are those arguments that are coded like the ‘breaker’ argument of the ‘break’ predicate. Similarly, object-like arguments (O-form) are those markers that code the ‘broken thing’ of ‘to break’.<sup>1</sup>

Comitativity constructions include two subtypes with opposite meanings. Those language-specific markers that denote ‘being accompanied’ and are used in a comitative construction (see Arkhipov, 2009) are categorized as a

1 Estonian and Finnish may mark direct objects in two ways: the partitive case is used with unboundedness (of the object and/or of the action) and the accusative case otherwise (Kiparsky, 1998). For this reason, both the partitive and accusative cases are indistinguishably classified as O-form. Finnic languages also have lexical partitives, which are approached as direct objects too for the sake of simplicity.

WITH constructions. If a marker can be used for explicit marking of non-involvement of a participant in the situation ('to be without'), i.e., expresses a caritive meaning (Oskolskaya et al., 2020), it is classified as a WITHOUT construction.

The INTRANSITIVE type is employed for the only intransitive predicate in the sample – Hungarian *gitározik* 'to play the guitar' derived from the noun *gitár* with the suffix *-ozik* (cf. also *cigarettázik* 'to smoke cigarettes' < *cigaretta* 'cigarette' + *-zik*). However, this construction is not language-specific: there were no other language-specific markers corresponding to the introduced concept for intransitive predicates since all other predicates are bivalent.

If none of the construction types are applicable and a marker has a non-abstract spatial meaning, then one of the 12 **spatial constructions** is assigned to it according to its non-abstract spatial meaning (listed in Table 1).

Note that the same comparative construction often covers different markers in one language. For example, the Russian markers of goal *do* 'to, until' and *k* 'to' are both classified under the AT/TO construction (5)–(6).

(5) *Petja dotronulsja do sten-y*  
 PN touched to wall-GEN.SG  
 'Petja touched the wall.'

(6) *Petja podošel k sten-e*  
 PN approached to wall-DAT.SG  
 'Petja approached the wall.'

Frequently occurring in the data dative-marked recipients, as *učitelju* in (7), also belong to the AT/TO construction. Dedicated dative cases are found only in Slavic, Baltic, German and Basque (the significance of this feature is discussed in Section 4).

(7) *Petja otvetil učitel-ju*  
 PN answered teacher-DAT.SG  
 'Petja answered to the teacher.'

Another frequent language-specific category in the data is the instrumental case. All instrumentals in my sample can bear either comitative or spatial meaning. For example, Basque, Chuvash, and Kalderash Romani instrumentals are classified as a WITH construction since they can be used in a comitative construction. The Slavic and Lithuanian instrumentals are covered by the ALONG construction because they can indicate the corresponding spatial

meaning ‘along’. Bare instrumentals cannot be used in the comitative meaning in these languages. The presence or absence of instrumental-comitative syncretism is an important distinguishing feature in object marking, as shown in Sections 4.2 and 5.

Other semantic types (bottom row in Table 1) include those meanings that can refer to neither formal nor spatial concepts. The ABOUT construction is set when the marker is used to denote the topic (‘in regard to something’). FOR is assigned to the purpose markers (‘for the sake of, because of’). AGAINST is applied to the markers with the specific meaning ‘against something’.

POS (possessive) is a construction representing the marking strategy in the basic internal-possession construction. It is considered to be semantic in order to disentangle the dedicated possessive construction from the spatial one. Therefore, if a marker has a spatial meaning (like German *von*), it is marked with a spatial construction (FROM). If a marker has no spatial meaning but is used in possessive construction (like Polish genitive case), it is classified as POS.<sup>2</sup>

This approach unavoidably has some shortcomings. First, the sample does not contain valency alternations and dialectological data, which could help track the boundaries of different contact zones. Second, the list of bivalent predicates is arbitrary and not balanced in terms of frequency which is an important variable for discovering contact effects (Seržant, to appear). Finally, classification based on semantic comparative concepts may also introduce certain noise.

Nevertheless, applied consistently, comparative concepts are an adequate option for this study. While a “semantically blind” search for correlation patterns based on calculating entropy and mutual information (Say, 2014; Bickel et al., 2014) is a promising venue, it is not completely flawless because uncontrolled coincidences may boost similarity effects artificially. For instance, the Lithuanian instrumental case corresponds to the Russian dative case in several predicates, which is a coincidence that draws these languages together increasing their mutual predictability, cf. (8) and (9).

- (8) Russian  
*Petja verit / doverjaet / udivilsja Maš-e*  
 PN believes / trusts / was\_surprised PN-DAT.SG  
 ‘Petja believes / trusts / was surprised by Masha.’

<sup>2</sup> See the Supplementary Material for the full classification of language-specific markers into comparative constructions.

TABLE 2 Necessary conditions for the effects explaining clustering

Effect	Genealogical condition (same branch)	Geographical condition (adjacent territories <sup>a</sup> )
Genealogical	+	–
Areal	–	+
Genealogical and areal	+	+
Coincidence	–	–

a All languages in my list may be located on a map according to the territories where they are used the most.

(9) Lithuanian

*Petras tiki / pasitiki / nusistebėjo Marij-a*  
 PN believes / trusts / was\_surprised PN-INS.SG  
 ‘Petras believes / trusts / was surprised by Marija.’

In this paper, such coincidences are disentangled. Specifically, the Lithuanian instrumental case is replaced by the spatial comparative construction *ALONG*, and the Russian dative case is replaced by the *AT/TO* construction. Therefore, in contrast to the mutual information or entropy reduction-based approach, they are treated as different constructions. Comparative concepts are used as a *prima facie* tool to reveal large-scale patterns for further detailed consideration.

Methodologically, I will rely on the following distinctions to disentangle the areal signal from the genealogical one: Table 2 summarizes all theoretically possible cluster-forming effects.

For example, according to Table 2, if the languages of a cluster (subcluster) meet the condition of a close genealogical relationship, i.e., belong to the same subfamily of the second order (according to Glottolog) and are not geographically adjacent, I will refer to a cluster (or subcluster) as a genealogical one. If neither of the conditions is true, the tie between languages will be taken as accidental.<sup>3</sup>

For clarity, the quantitative methods employed in this study are discussed within the sections presenting the results: cluster analysis is detailed in Section

3 However, the possibility of historical or functional explanation for ‘unrelated’ languages is not excluded, it is just not elaborated further in order to focus on the bigger picture and save space.



3, while the comparison of observed frequencies of object marking strategies with the Zipf distribution is discussed in Section 5.

### 3 Cluster Analysis and Its Results

In this section, I compare the object marking strategies in the languages of Western Eurasia to track genealogical, contact-driven, and areal signals. I rely on several different clustering methods and multidimensional scaling applied to the distances between all possible language pairs of the sample. The code for this section can be found in the Supplementary Material.

The data analysis was conducted using R (R Core Team, 2020). After computing the distances between the languages (*asymmetric binary* method), I applied agglomerative clustering (R function *hclust*) and Sammon's non-linear multidimensional scaling (R package *MASS*, function *sammon*) to analyze the obtained distances. The optimal number of clusters was defined using the Silhouette method.

In order to explore all possible relations between languages, several clustering methods (*average*, *mcquitty*, *single*, *complete*, *ward.D2*, *ward.D*) were compared, see Figure 1. The clusters are visualized as phylogenetic trees using package *factoextra* (Kassambara and Mundt, 2020). The colors distinguish different clusters. Inside each cluster, the less edges separating two languages, the more similar they are. As clusters are built as incremental bottom-up mergers, the closest relations between languages are represented by the vertices most remote from the center of the tree.

First of all, the cluster trees are ordered in descending order from left to right according to their cophenetic correlation (R function *cophcor*; Brunet et al., 2004). The observed splits are regular enough to claim that Irish, Basque, Czech and Romani Kalderash demonstrate the lowest degree of similarity with the other languages, often building separate clusters. The other 26 languages merge into three to four larger clusters, showing an acceptable degree of variation.

As the *average* clustering method gives the highest cophenetic correlation coefficient of all available clustering methods (0.83), it is chosen for further analysis, i.e., the clusters will be referred to according to the splits obtained with this method. However, other cluster trees (cophenetic correlation from 0.82 to 0.48) certainly contribute to the understanding of the realistic picture of the relationship between languages.

Overall, Figure 2 shows that object marking strategies divide the languages of Western Eurasia into six clusters. For the sake of simplicity, clusters with more



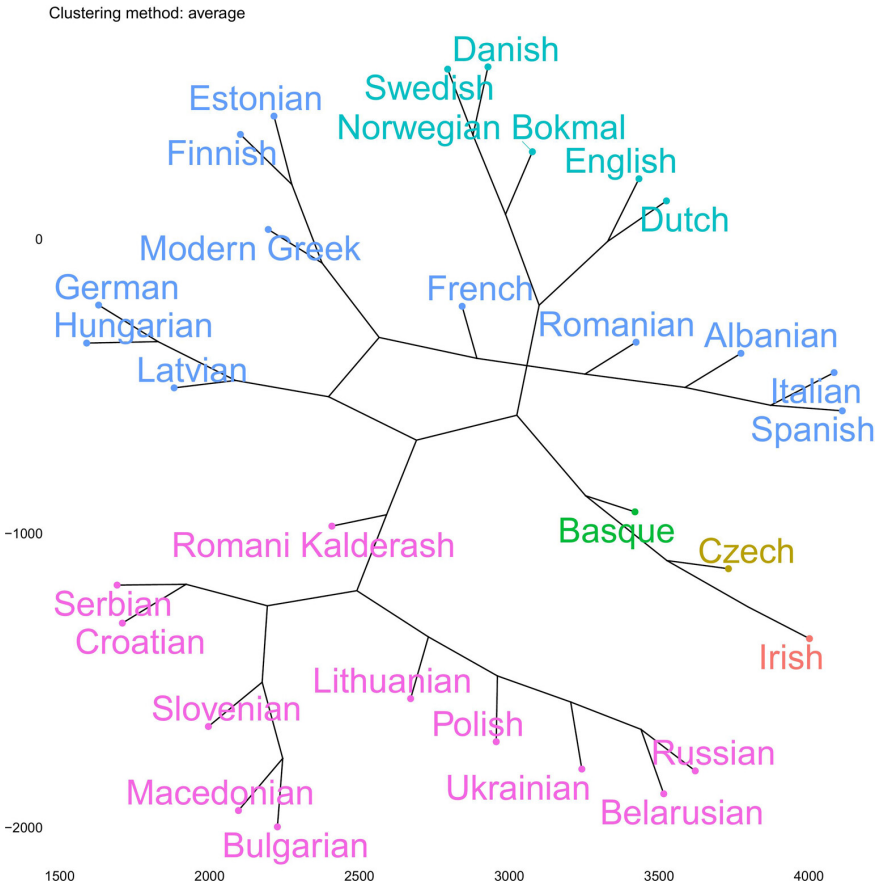


FIGURE 2 Language clusters (method: *average*)

than one language will be referred to by the dominant language group, where applicable (see Table 3).

The overall picture is complex but not random. It can be instantly noted that the largest clusters are genealogically homogeneous with minor embeddings.<sup>4</sup>

4 Most of the embeddings are explained below. The only coincidence with no genealogical and/or areal explanation is the embedment of the Finnic languages in the Mixed cluster. An important reminder shall be made here: Finnish and Estonian partitives were considered as direct objects (not like in the source database BivalTyp). I assume that the Finnic languages merged with Greek because of the considerable overlap in transitivity. The transitivity prominence (calculated as the percentage of transitive verbs out of a total of 103 verbs in the sample) is 65% for Estonian, 68% for Finnish, and 71% for Greek. In contrast, Albanian, Romanian, French, Italian, and Spanish have their transitivity prominence value in the 59–65% range. At the same time, Greek is inseparable from the Romance-based subcluster because

TABLE 3 Languages of the clusters in Figure 2

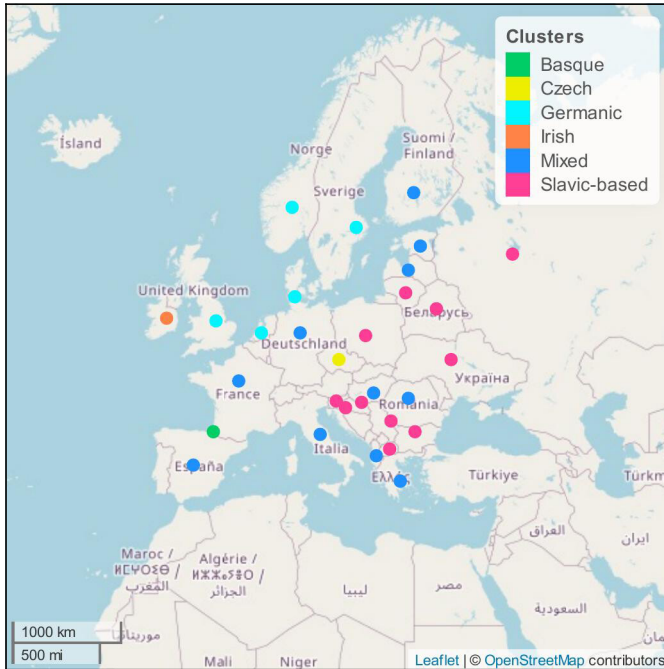
Cluster	Number of languages	Languages
Mixed	11	<u>IE</u> : Italo-Western Romance ( <b>French, Italian, Spanish</b> ); Eastern Romance ( <b>Romanian</b> ); <b>Albanian</b> ; <b>Modern Greek</b> ; West Germanic ( <b>German</b> ); Baltic ( <b>Latvian</b> ) <u>Uralic</u> : Finnic ( <b>Estonian, Finnish</b> ), <b>Hungarian</b>
Slavic-based	11	<u>IE</u> : South Slavic ( <b>Bulgarian, Croatian, Macedonian, Serbian, Slovenian</b> ); East Slavic ( <b>Belarusian, Russian, Ukrainian</b> ); West Slavic ( <b>Polish</b> ); Baltic ( <b>Lithuanian</b> ); Indo-Iranian ( <b>Kalderash Romani</b> )
Germanic	5	<u>IE</u> : North Germanic ( <b>Danish, Norwegian Bokmål, Swedish</b> ); West Germanic ( <b>Dutch, English</b> )
Basque	1	<u>Basque</u> : <b>Basque</b>
Czech	1	<u>IE</u> : West Slavic ( <b>Czech</b> )
Irish	1	<u>IE</u> : Celtic ( <b>Irish</b> )

Furthermore, as there are three dominating language groups in the sample, all non-Slavic, non-Romance, or non-Germanic languages end up somewhere in the middle of the dendrogram, distant from the larger genealogy-driven clusters and forming dangling nodes. This is precisely what we observe in Figure 1 for Basque and Irish which support the genealogical trend in clustering (they are unique representatives of their families or branches and therefore are not expected to cluster with other languages). In this sense, Albanian, Modern Greek and Hungarian are different from Basque and Irish since they end up in the larger clusters.

Two exceptions to the general genealogical trend are German (together with Latvian and Hungarian, presumably influenced by it) and Czech (form-

---

of its non-transitive object markers (presumably assimilated via intense and multidirectional language contact on the Mediterranean Sea). For example, the most widespread strategy for the meaning 'be angry' in Western Europe is ON/ONTO; however, some languages in the south (Spanish, Basque, Italian, Albanian, Greek) together with English and Irish employ the comitative strategy. Cf. the result in Say (2018: 595) with Modern Greek and Estonian clustered together in a NeighborNet dendrogram.



MAP 1 Clusters from Figure 1 on the map

Note: All the maps are created using the *lingtypology* R package (Moroz, 2017)

ing its own cluster). Unexpectedly, they do not merge with their relatives in the other larger clusters (Germanic and Slavic-based respectively). I suggest that this position is not accidental and, most likely, caused by intense multidirectional language contact (see 4.2.1 and 4.3). As Map 1 shows, German and Czech are geographically central in the area under consideration which might be an important prerequisite for multilateral contacts.

To summarize, the genealogical factor emerges as the most significant factor behind the clustering of object marking strategies in Western Eurasia. Areal effects manifest themselves as deviations from expected genealogical clustering: 1) Hungarian (Uralic) and Latvian (Baltic) merge with German (Germanic); 2) Albanian (Albanian) and Greek (Graeco-Phrygian) merge with the Romance-based cluster; 3) Lithuanian (Baltic) merges with the Slavic-based cluster. These cases are discussed in the following section. As for the solo clusters like Irish and Basque, no genealogical connections are expected since these languages are the only representatives of their groups in the sample.

These results, in general, support Say's (2014; 2018) conclusions based on mutual information between each language pair in BivalTyp: valency classes

are affected by both genealogical and areal factors. However, zooming in to a smaller geographical area (together with focusing on objects only) brings some new findings discussed further.

## 4 Clusters and the Circum-Baltic Area

This subsection elaborates on some of the conclusions drawn from Figures 1–2 and provides explanations of the observed clusters, mainly focusing on the languages relevant to the Circum-Baltic area.

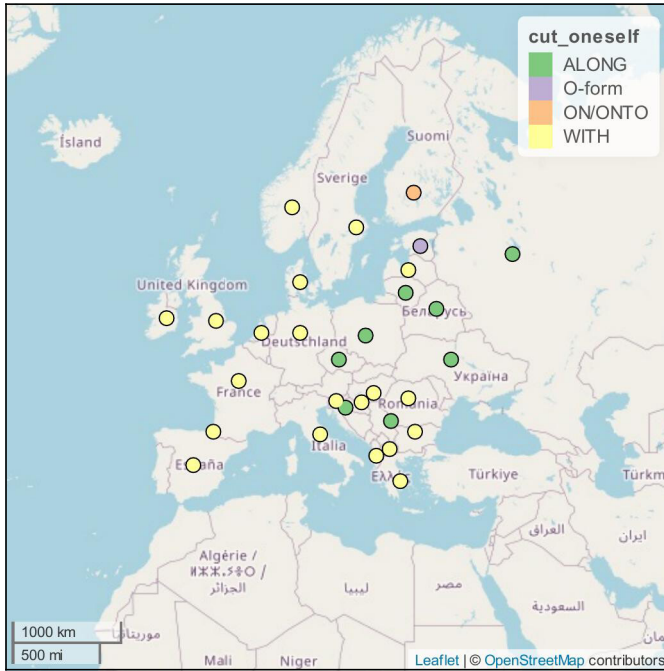
### 4.1 *Slavic-Based Cluster*

The Slavic-based cluster consists almost entirely of Slavic languages with the exceptions of Lithuanian (Baltic) and Kalderash Romani (Indo-Iranian).<sup>5</sup> One may argue that Lithuanian is influenced not only by areal factors but also by its genealogical heritage. Therefore, the clustering results may reflect that many Lithuanian object markers are inherited from the period before the Baltic and Slavic branches diverged. Indeed, many strategies are shared between Lithuanian and most of the Slavic languages (even those not in contact with Lithuanian), for instance, AT/TO with ‘lose a game’, ABOUT with ‘forget’, ALONG<sup>6</sup> with ‘be content’ and ‘cut oneself’. Map 2 shows the correspondence of the languages in the cluster and the distribution on the map.

Nevertheless, Polish (West Slavic) aligns most closely with Lithuanian, as briefly discussed in Section 3. This fact suggests that an areal effect is at work here in addition to the genealogical effect. For a better understanding of this connection, one should take into account individual marking strategies. One of the key similarities between Polish and Lithuanian is the use of the POS (possessive) construction, which involves genitives that can express various types of partitive meanings. “Possessively” marked objects under negation are obligatory in both languages whereas in Russian, Ukrainian, and Czech their application is much more restricted – for instance, by definiteness. These languages use accusative marking more readily (Miestamo, 2014; Seržant, 2015b).

5 Not discussed further because of the conclusion based on the Figure 1: Romani does not appear together with the languages of this cluster consistently. See the discussion after the Figure 1.

6 This construction applies when the marking device expresses ‘movement along the surface’. In Slavic and Baltic languages, this is typically the instrumental case (see Section 2).



MAP 2 Object marking strategies of the predicate 'cut oneself' in the sentence 'X cut oneself with a razor'

For example, the predicate 'hate' with the lexicalized negative particle is strictly non-transitive only in Lithuanian and Polish, cf. (10)–(14),<sup>7</sup> which opposes this pair to the rest of the continental languages of Western Eurasia.<sup>8</sup>

(10) Russian

*Petja nenavidit Mash-u*  
 PN hates PN-ACC.SG  
 'Petja hates Masha.' (BivalTyp)

(11) Ukrainian

*Petro nenavidit' Marij-u*  
 PN hates PN-ACC.SG  
 'Petro hates Marija.' (BivalTyp)

<sup>7</sup> The negative particles are given in bold. Synchronically, they are a part of the root.

<sup>8</sup> Irish object marking strategy for the predicate 'hate' is ON/ONTO.

## (12) Czech

*Petr nenávid-í Michal-a*

PN hates PN-ACC.SG

'Petr hates Michal.' (BivalTyp)

## (13) Polish

*Feliks nienawidz-i Ann-y*

PN hates PN-GEN.SG

'Feliks hates Anna.' (BivalTyp)

## (14) Lithuanian

*Petr-as nekenč-ia Marij-os*

PN hates PN-GEN.SG

'Petras hates Marija.' (BivalTyp)

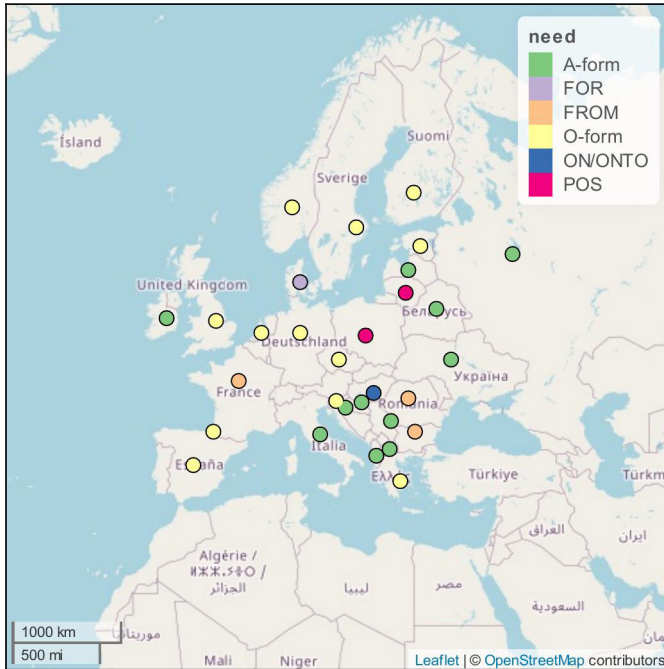
This example reflects a general Polish-Lithuanian-Finnic isogloss of mandatory partitive/partitive genitives under predicate negation (Arkadiev, 2017; Seržant, 2021).<sup>9</sup>

There are several instances where the possessive strategy aligns Lithuanian and Polish, setting them apart from other languages in their cluster. For example, the predicate 'need' (e.g., 'Someone needs money'), as shown in Map 3, illustrates this alignment. The correlations observed in Maps 2 and 3 are crucial to the merging of Lithuanian and Polish. I interpret this relationship as both genealogical and areal, based on the principles outlined in Table 2.

What I have shown above is evidence for a strong areal effect that explains why Lithuanian merges closer to Polish and not to any other Slavic language. As the possessive marking with negation is a common Balto-Slavic trait (see Seržant, 2015b; Arkadiev, 2017), this is most likely a case of language contact promoting retention of a trait. Polish is known to be influential and "attractive for Lithuanian and Ruthenian noblemen, and also for the townspeople" in the multi-ethnic Polish-Lithuanian state (Törnquist-Plewa, 2000: 194; also see Balode and Holvoet, 2001a: 45, on external influence on Lithuanian). Thus, Polish must have had a conservative effect on Lithuanian. At the same time, Lithuanian L2 speakers might have reinforced the trait in Polish (as opposed

9 Note that clustering of Polish, Lithuanian and the Finnic was not harmed by classifying Finnic partitives as an O-form construction: all BivalTyp stimuli contain real (rather than negated or hypothetical) situations (Say et al., 2020).





MAP 3 Object marking strategies of the predicate 'need' in the sentence 'X needs money'

to all other Slavic languages). If imperfect learning during language acquisition leads to language change (see, among others, Seifart, 2019), then “perfect” acquisition of grammatical patterns might reinforce the status quo, thereby promoting the retention of stable traits in languages in contact.

Another fascinating question that deserves to be answered in a separate case study is whether the non-negative predicates with the possessive strategy in both Lithuanian and Polish (see Table 4) represent retention or innovation.

Interestingly, all verbs except for ‘lack’ and ‘want’ are marked with participles in Finnish and Estonian, where it is the largest class of object marking strategies. Seven of eleven verbs employ the same strategy in Russian (though interchangeable with O-form in the cases of ‘want’ and ‘avoid’).

To sum up, with both negative and non-negative predicates, the possessive strategy in Polish and Lithuanian is inherited and likely preserved due to retention promoted by language contact. The results support Timberlake’s (2017) observation that “The Grand Duchy, then Commonwealth, became a small linguistic area in its own right”.

TABLE 4 Possessive construction usages for object marking in Lithuanian and Polish (compared to Russian and Finnish)

Lithuanian verb (Zaika, 2018)	Polish verb (Moroz, 2018)	Translation	Comparison
<i>trūkti</i>	<i>brakować</i>	‘lack’	Genitive in Russian (p.k.)  Partitive in Finnish (Shagal, 2021)
<i>norėti</i>	<i>chcieć</i>	‘want’ <sup>a</sup>	
<i>bijoti</i>	<i>bać się</i>	‘be afraid’	
<i>varžytis</i>	<i>wstydzić</i>	‘be shy’	
<i>pakakti</i>	<i>wystarczyć</i>	‘have enough’	
<i>vengti</i>	<i>unikać</i>	‘avoid’	
<i>ieškoti</i>	<i>szukać</i>	‘look for’ <sup>b</sup>	
<i>klausyti</i>	<i>stuchać</i>	‘listen’	
<i>klausytis</i>	<i>stuchać się</i>	‘obey’	
<i>gailėti</i>	<i>žalovaut</i>	‘sympathize’	
<i>reikėti</i>	<i>potrzebować</i>	‘need’	

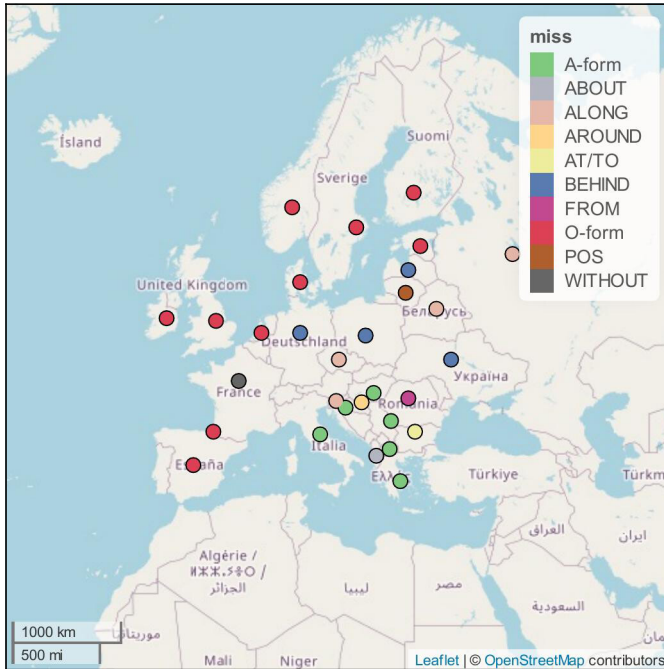
a For Polish in Russian verbs meaning ‘want’, only uncountable abstract nouns can be marked with genitives (e.g., pol. *wolność* ‘freedom’, rus. *spravedlivost’* ‘justice’, etc.). The same restriction is applicable to the Russian verbs *iskat’* ‘look for’ and *izbegat’* ‘avoid’. The Russian verb *slušať’sja* ‘obey’ is primarily transitive but still may be used with genitive objects, for example, in a set phrase *Poslušaj mojego soveta* ‘Take my advice’.

b See footnote 9.

#### 4.2 Mixed Cluster

The Mixed cluster consists of three subclusters: German–Hungarian–Latvian, Estonian–Finnish–Modern Greek and Spanish–Italian–Albanian–Romanian–French (briefly, the Romance-based subcluster). Since only the first cluster is relevant to the context of the CB area, it is discussed in more detail in Section 4.2.1. However, the other two clusters also merit a brief comment.

The Romance-based subcluster (leaving Finnic languages aside) is a merger of the Romance languages – Spanish, Italian, Romanian, French – and non-Slavic Balkan languages, namely Albanian and Modern Greek. Noteworthy, there is a similar split in the use of ‘have’-perfects, lack of verbal negation, and the grammaticalized category of definiteness for the languages traditionally known as the Balkan Sprachbund (see Haspelmath, 2001; Joseph, 2010; Friedman and Joseph, 2017). This study demonstrates that object marking is a grammatical feature that underscores closer relations between Albanian, Greek, and the Romance languages, while excluding the Balkan Slavic languages. Unlike other distinguishing features, it also separates the Germanic languages from



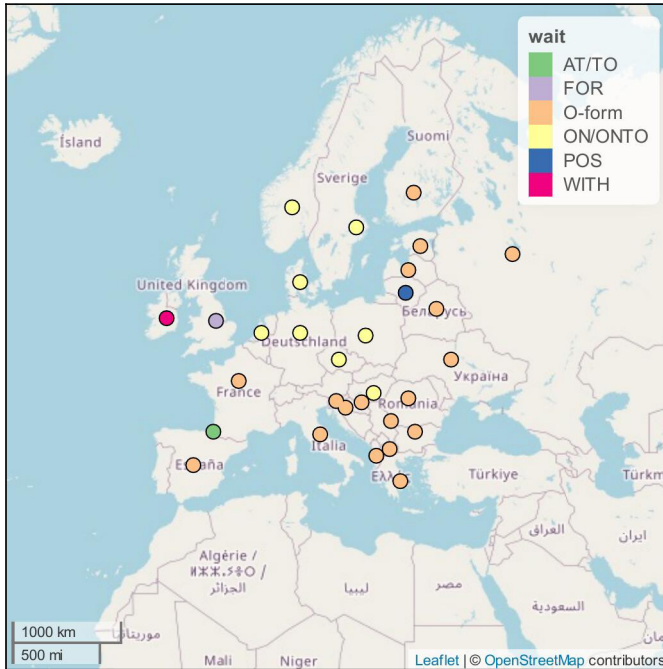
MAP 4 Object marking strategies of the predicate 'miss' in the sentence 'X misses Y'

the Romance and non-Slavic Balkan languages. In the Balkans, tracing the direction of interference is challenging. However, some evidence suggests an earlier Albanian-Romance interference (Rusakov, 2013; 2021), which is in line with the clustering results.

A rather unexpected clustering of the Finnic languages with Greek on the one hand and the Romance-based subcluster on the other hand is an artificial effect of applying comparative concepts. Finnish and Estonian partitives were all classified as O-form, which boosted the number of transitive patterns (see Footnote 1). A parallel but independent expansion of the transitive pattern in Finnic, Greek, and Romance may also be a factor affecting their clustering. This assumption, however, remains to be explored in further research.

#### 4.2.1 German–Hungarian–Latvian Subcluster

Historically and demographically, German influences Latvian and Hungarian, rather than the other way around. No singular object marking strategy simultaneously isolates German with Hungarian and Latvian to the same extent as the possessive strategy isolates Lithuanian and Polish. Instead, some strategies are shared between German and Hungarian, while others are shared between German and Latvian (see Maps 4–5).



MAP 5 Object marking strategies of the predicate 'wait' in the sentence 'X is waiting for Y'

Besides the distribution shown in Maps 4–5, where the variability of the marking strategies is very high, there are other cases when German merges with the Eastern half of the languages (mainly Slavic) in contrast to the Western half. For example, the verb 'help' has only three object marking strategies in the sample: AT/TO (all the Slavic, German, Hungarian, Romani, Basque), O-form (all the Germanic, Finnic, Romance including Romanian, Albanian, Greek) and WITH (Irish). So, unlike the other Germanic languages in the sample, German marks a recipient with a dedicated case (dative) as opposed to the preposition of goal *zu* 'to', which is a common trait of German and the Balto-Slavic.

German is, therefore, so to say, at the crossroads: some of its marking strategies coincide with the ones typical for the Germanic (Map 5), while some others – with the Balto-Slavic (Map 4) and the Romance ('help'). German–Hungarian and German–Latvian marking strategies do not coincide for intransitive predicates.

Looking at the sociolinguistic situation clarifies this issue. In historical Hungary, Hungarian was not the mother tongue of the majority of the population. Even in the nineteenth century, during the heyday of the national idea, less

than 50% of the population in Hungary spoke Hungarian (Törnquist-Plewa, 2000: 191). In 1790, when Latin was abolished as the official language, it was initially replaced by German. However, this was short-lived as the Hungarian nobility intervened immediately, promoting the Magyar identity, culture, and language against the German elite and the Slavic peasantry. During the Language Reform Movement, borrowings from other languages spoken in the Habsburg monarchy, especially German, were shunned (Benkő, 1972; Deme, 1972). This clearly indicates that the Hungarian language codifiers of that period were bilingual, as they could identify and avoid German words. Morphological rules and syntactic patterns, however, are harder to identify as calques, allowing them to penetrate the standard language. This may be the case for object marking strategies but not exclusively; there is more evidence from the verbal domain. For instance, the use of local adverbs as postpositional modifiers of the verb (see 15a–b) is “an important contribution to the development of the Hungarian preverbal system in the period of intensive contacts between Hungarian and German” (Kurzová, 2019: 273). Moreover, there are auxiliaries *ist* vs. *wird* in German and *van* vs. *lesz* in Hungarian, distinguishing the stative/resultative and processual meaning and being used similarly in these languages (Kurzová, 2019).

## (15) a. German

*komm doch hier*  
 come.IMP yet here  
 ‘Come here!’ (Kramer, 1981: 135)

## b. German

*jöjj hát ide*  
 come.IMP yet here  
 ‘Come here!’ (Kramer, 1981: 135)

The German-Latvian sociolinguistic situation is quite similar. German is known to be the dominant language and lingua franca in the Latvian territory from the Hansa period (Timberlake, 2017; Vanags, 2019) until at least the mid-19th century (Vanags, 2011). It is also noted that German was native language for almost all authors (translators) of religious Latvian texts, who had learned the Latvian language quite incompletely, already in adulthood (Vanags, 2019). Although the Latvian peasantry remained illiterate and tied to elements of the old feudal system with a German-speaking nobility even after the Emancipation of the Serfs in 1861 (Carmichael, 2000: 268), written Latvian became a language of the educated class, influenced by the language of the codifiers and luminaries. For

instance, Pastor K.F. Watson, the creator and editor of the first Latvian newspaper “Latweeschu Awises,” and J. Ch. Wolter, who initiated the foundation of the Curonian primary schools, were key figures for the initial steps of Latvian linguistic development.

As for German-Latvian bilingualism at the time, there are two points of view. Some authors argue that it was restricted because the Baltic German nobility “tended to isolate itself from the native population rather than to assimilate to it” (Balode and Holvoet, 2001b: 10). Others say that the bilingualism of the literate population had to be high since German was a means of raising social status for native Latvians (Wolter, 1908: 374–375). So, the scenario in which the literary norm spreads through the flourishing national literature and partially downgrades to the conversational norm is also probable.

Whatever the preconditions for language change (oral communication, standardization processes, or both), there are many instances of the structural influence of German on the languages of the Baltics. For example, the comitative-instrumental syncretism found in German has been transferred into the local languages: Latvian, Estonian, and Livonian (Stolz, 2001: 607), see (16a–b).

(16) a. German

*Zuerst notiert man die Erzählungen der*  
 first take\_notes:3SG man DEF.ACC.PL report:PL DEF.GEN.PL  
*Forscher mit Bleistift*  
 explorer:PL with pencil  
 ‘First the reports of the explorers are written down with a pencil.’ (Stolz, 2001: 596)

b. Latvian

*Pētnieku stāstus vispirms pieraksta ar*  
 explorer:GEN.PL report:ACC.PL first write\_down:3(SG) with  
*zīmuli*  
 pencil:ACC  
 ‘First the reports of the explorers are written down with a pencil.’ (Stolz, 2001: 596)

Another example is a significant German–Latvian interference in the use of prepositions noted in early translations from German to Latvian. For instance, prepositional constructions involving *iekšān* ‘in, into’ often replace the (unmarked) constructions with a dedicated locative case (Vanags, 2019: 285;

see also Vanags, 1992). In modern Latvian, the use of *iekšā* ‘in, into, inside’ is rather restricted to the contexts of being or moving inside the “container”.

In short, object marking strategies demonstrate traces of language contact providing evidence for the binary nature of interlinguistic influence between German and Hungarian as well as German and Latvian. At the same time, German shows some convergence effects with the Romance languages, which is likely determined by a geographical factor, as reflected in the diversity of the Mixed cluster.

### 4.3 *Other Clusters*

The only remaining cluster that contains CB languages is the Germanic one. According to the principles in Table 1, it has been influenced by both genealogical and areal pressures. Firstly, relations in the cluster reflect the genealogical division into North and West Germanic. Secondly, the Germanic languages form a geographical continuum (see Map 1). For object marking strategies, there is no evidence of significant Germanic influence on the languages of the Eastern Baltic coast beyond the impact of German discussed above.

The other clusters are not directly related to the CB languages. The fact that the Czech object marking deviates from the Slavic-based cluster is rather surprising. One might expect Czech to cluster more closely with the languages of the Mixed cluster, given its prolonged contact with German within the Habsburg monarchy and Austria-Hungary. Timberlake (2017) notes that Czech underwent a reduction in case functions due to German influence, and it shows a preference for the allative preposition *do* (‘up to, into the vicinity of’) over the illative *v* (‘in, into’), with *do* functionally overlapping more with the German *nach* (‘up to’) than with *v*. Additionally, Czech’s central position on the map of Western Eurasia, surrounded by diverse language groups, suggests varied and frequent language contact situations.

## 5 Large-Scale Patterns

This section provides an overview of object marking strategies distributed by frequencies. The distribution observed is compared with the predicted Zipf distribution, which is a principle derived from Zipf’s Law. Zipf’s Law suggests that in many types of data, including linguistic data, the frequency of any element is inversely proportional to its rank in a frequency table. In other words, a small number of items occur very frequently, while most items occur infrequently. This distribution is useful for understanding how common and rare strategies are in relation to one another.

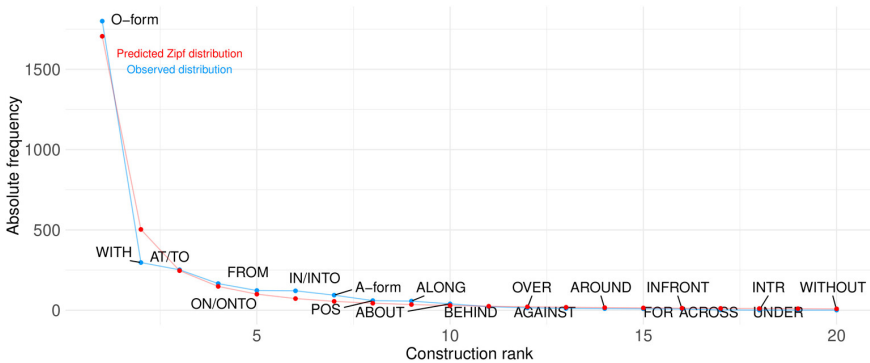


FIGURE 3 Observed and predicted Zipf distribution of object marking strategies in the sample

Additionally, I briefly comment on the predicates that employ marking strategies shared by the overwhelming majority of languages in the sample. These include not only transitive predicates but also those using the comitative strategy.

Across the entire dataset, O-form occurrences constitute 58% of all instances, which is six times more frequent than the second most common strategy, WITH (10%). The third most frequent strategy is AT/TO (8%), while the strategies FROM, ON/ONTO, and IN/INTO each account for no more than 4%–5%, as shown in Figure 3.

Taken together, the languages of Western Eurasia demonstrate an extremely high dominance of the transitive strategy (O-form) followed by the comitative strategy (WITH). The concepts of goal/recipient/location (AT/TO) and source (FROM) turn out to be more frequent than the other spatial strategies, which is probably a sign of their relative stability in the languages under consideration.

Although the observed distribution differs significantly from the predicted Zipf distribution (Pearson's Chi-squared test with simulated p-value was applied;  $\chi^2 = 125.01$ , p-value  $< 0.0005$ ), it resembles the Zipfian curve: the larger the ordinal number of an element, the smoother its difference from the subsequent element. The most dramatic deviation from Zipf's curve corresponds to the second rank (WITH). This deviation may result from a limited verb set; specifically, the proportion of WITH-marked objects is higher in the statistical population of all bivalent verbs in the languages of Western Eurasia.

The alternative explanation comes from the nature of bivalent verbs: the O-form strategy dominates over the other strategies because it is efficient to employ the same strategy for the situations with two participants and assign the roles according to the position of the argument. The data demonstrate a



TABLE 5 Transitives in all languages of the sample

---

bend	milk
break	open
cover	paint
drink	plough
drive	pour
eat	put on
find	read
fry	see
hear	sing
kill	take
lose	upset
make	wash
melt	write

---

steady trend: 26 out of 103 do not vary in object marking. They are transitive in all 30 languages of my sample; see Table 5.<sup>10</sup>

As expected, most predicates in Table 5 imply a high affectedness of the patient. However, perception predicates like ‘hear’, ‘see’, ‘read’ are the highest in the transitivity prominence scale, too (cf. Tsunoda, 1985, and the discussion of a broader list of BivalTyp predicates in comparison to Tsunoda’s hierarchy in Say, 2018).

The comitative strategy ranks second in frequency (Figure 3). There is only one predicate coding the object with the comitative strategy in all languages of the sample – ‘have a quarrel’ with the ‘opponent’ participant coded by comitative. Similar to Haspelmath’s (2015) scale of transitivity prominence, comitativity may also be represented on a scale. Comitativity prominence can be calculated similarly to transitivity prominence; namely, by dividing the number of comitatives for a given predicate by the total number of languages in the sample. Table 6 lists the predicates ordered by their comitativity prominence.

---

<sup>10</sup> These 26 exclusively transitive predicates were not excluded from the sample because their similarity does not impact the analysis given the chosen method. Ultimately, only 77 predicates (= 103–126) contribute to the overall analysis.

TABLE 6 Comitativity prominence

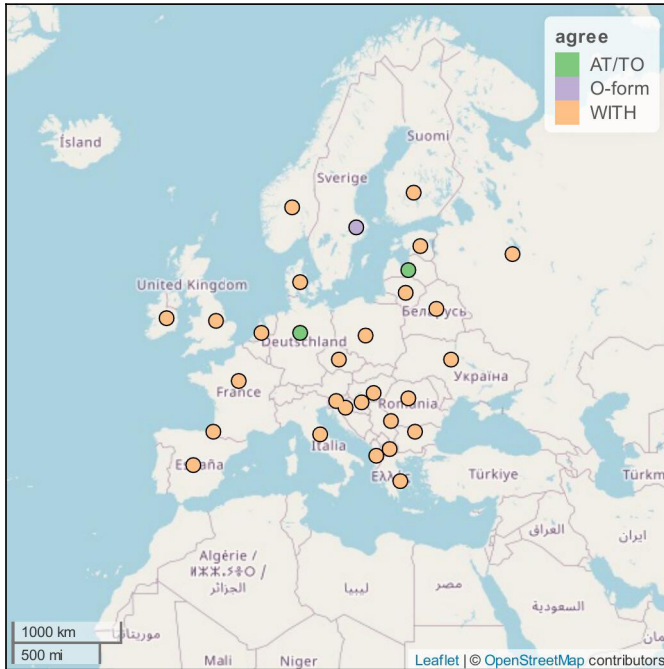
Predicate meaning	N of occurrences	Comitativity prominence
have a quarrel	30	1.00
speak	29	0.97
fight	28	0.93
mix	28	0.93
agree	27	0.90
cut oneself	20	0.67
get to know	19	0.63
be content	16	0.53
fill (as in 'The bucket filled with water')	16	0.53
encounter	13	0.43
wave	12	0.40

Some of the predicates ('cut oneself', 'wave', 'fill') reflect the tendency to comitative-instrumental syncretism (see Stolz, 2001). However, most of the meanings – especially those with the comitative prominence of 0.90 and higher – require the object to be a human companion. Deviations from these relatively stable comitative patterns may point to some non-accident correlations. For instance, Map 6 opposes German and Latvian to the rest of the languages because of their AT/TO-strategy of the 'agree' predicate, which is another German-Latvian isogloss.

There are also other predicates, which are neither transitive nor comitative, that exhibit low variability in object marking. For example, the predicate 'answer' is primarily associated with the AT/TO strategy, while 'sink' predominantly uses the IN/INTO strategy. However, these predicates are relatively sporadic and do not appear to form a larger class like the transitive or comitative ones.

## 6 Conclusions

This study used clustering methods to examine object marking strategies in Circum-Baltic languages from a broader areal perspective. No clusters of more than two languages were confined solely to the CB area. Instead, the identified clusters extend well beyond the CB area and are primarily driven by



MAP 6 Object marking strategies of the predicate 'agree' in the sentence 'X agrees with Y'

genealogical affiliations. Deviations from this pattern typically result from specific contact between two languages. For example, German detaches from the Germanic cluster and forms a subcluster with Hungarian and Latvian within the Mixed cluster (German–Hungarian–Latvian, Estonian–Finnish–Modern Greek, Spanish–Italian–Albanian–Romanian–French). Consequently, there are no grounds for considering the German-Latvian relationship to be of a different nature than the German-Hungarian one. This result supports the claims about the dominant role of binary contact relationships (Dahl, 2001; Campbell, 2006).

One notable exception is a single object marking strategy unique to the CB area, which can be referred to as the CB isogloss. This is evident in the correlation of genitive/partitive predicates across Lithuanian, Polish, Russian, Finnish, and Estonian – languages from three different genealogical clades that form a geographical continuum (Seržant, 2015b). However, the infrequency of these predicates was insufficient to override the genealogical signals in the sample. Other correlations involving more than two CB languages from different clades are rare and not tied to a specific object marking strategy.

According to the clustering analysis, the areal explanation is particularly useful for describing the Mixed cluster, which overlaps significantly with core SAE languages (cf. Haspelmath, 2001: 1505). This overlap is largely due to the prevalence of transitive marking strategies in these languages (see also Say, 2014; Seržant et al., 2022), although specific spatial concepts used for object marking also play a role (Section 4.2).

The data reveal that languages located at the confluence of different genealogical clades have a higher likelihood of deviating from their genealogical clusters compared to those surrounded by closely related languages. For instance, German diverges notably from other Germanic languages, and Czech falls outside the Slavic-based cluster.

Additionally, the comitative strategy, following the transitive strategy, represents the most frequent class. A significant number of bivalent predicates, particularly those involving a human companion, use the comitative strategy. This trend may have broader cross-linguistic implications.

Future research could address the study's limitations by incorporating dialectological data and a broader range of moderately transitive bivalent and trivalent predicates to better understand convergence effects in object marking.

### Supplementary Materials

<https://doi.org/10.17605/OSF.IO/2ZAXN>

### Abbreviations

3	third person
ACC	accusative
DAT	dative
DEF	definite
GEN	genitive
IMP	imperative
INS	instrumental
PART	partitive
PL	plural
PN	person name
REFL	reflexive
SG	singular
p.k.	personal knowledge

## References

- Arkhipov, Alexander. 2009. Comitative as a cross-linguistically valid category. In Patience Epps and Alexander Arkhipov (eds.), *New Challenges in Typology: Transcending the Borders and Refining the Distinctions*, 223–246. Berlin, New York: De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110219067.4.223>.
- Arkadiev, Peter M. 2017. Объектный партитив отрицания: ареально-типологическая перспектива [Об"ektnyj partitiv otricanija: areal'no-tipologičeskaja perspektiva]. *Acta Linguistica Petropolitana. Trudy Instituta Lingvističeskix Issledovanij*, 13(3): 194–243.
- Balode, Laura and Axel Holvoet. 2001. The Lithuanian language and its dialects. *Circum-Baltic Languages* (1): 3–40; 41–79.
- Benkő, Loránd. 1972. The lexical stock of Hungarian. In Loránd Benkő and Samu Imre (eds.), *The Hungarian Language*, 171–226. The Hague, Paris: Mouton.
- Bickel, Balthasar, Taras Zakharko, Lennart Bierkandt, and Alena Witzlack-Makarevich. 2014. Semantic role clustering: An empirical assessment of semantic role types in non-default case assignment. *Studies in Language* 38(3): 485–511.
- Brunet, Jean-Philippe, Pablo Tamayo, Todd R. Golub, and Jill P. Mesirov. 2004. Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the National Academy of Sciences of the United States of America* 101(12): 4164–4169.
- Campbell, Lyle. 2006. Areal Linguistics: A Closer Scrutiny. In Yaron Matras, April McMahon, and Nigel Vincent (eds.), *Linguistic Areas*, 1–31. London: Palgrave Macmillan. DOI: [https://doi.org/10.1057/9780230287617\\_1](https://doi.org/10.1057/9780230287617_1).
- Campbell, Lyle. 2017. Why is it so hard to define a linguistic area. In Raymond Hickey (ed.), *The Cambridge Handbook of Areal Linguistics*, 19–39. Cambridge: Cambridge University Press.
- Carmichael, Cathie. 2000. Coming to Terms with the Past: Language and Nationalism in Russia and its Neighbours. In Stephen Barbour and Cathie Carmichael (eds.), *Language and Nationalism in Europe*, 264–279. Oxford: Oxford University Press.
- Dahl, Östen. 2001. Principles of areal typology. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible (eds.), *Language Typology and Language Universals: An International Handbook* 2, 1456–1470. Berlin: Mouton de Gruyter.
- Deme, László. 1972. Standard Hungarian. In Loránd Benkő and Samu Imre (eds.), *The Hungarian Language*, 255–298. Berlin, Boston: De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110880236-007>.
- Friedman, Victor A. 2006. Balkanizing the Balkan sprachbund: A closer look at grammatical permeability and feature distribution. In Alexandra Y. Aikhenvald and Robert M.W. Dixon (eds.), *Grammars in Contact: A Cross-Linguistic Typology*, 201–219. Oxford: Oxford University Press.

- Friedman, Victor A. and Brian D. Joseph. 2006. *The Balkan Languages*. Cambridge: Cambridge University.
- Friedman, Victor A. and Brian D. Joseph. 2017. Reassessing Sprachbunds: A view from the Balkans. In Raymond Hickey (ed.), *The Cambridge Handbook of Areal Linguistics*, 55–87. Cambridge: Cambridge University Press.
- Hammarström, Harald, Robert Forkel, Martin Haspelmath, and Sebastian Bank. 2022. *Glottolog 4.6*. Leipzig: Max Planck Institute for Evolutionary Anthropology. DOI: <https://doi.org/10.5281/zenodo.6578297>. Available online at <http://glottolog.org> (accessed August 15, 2022).
- Gaszewski, Jerzy. 2020. Does Verb Valency Pattern Areally in Central Europe? A First Look. In Luka Szucsich, Agnes Kim, and Uliana Yazhinova (eds.), *Areal Convergence in Eastern Central European Languages and Beyond*, 13–53. Berlin et al.: Peter Lang.
- Grossman, Eitan, Alena Witzlack-Makarevich, and Ilja A. Seržant (eds.). 2019. Valency and Transitivity in Contact: Theoretical and Empirical Issues. *Journal of Language Contact* 12(1).
- Grossman, Eitan. 2021. Transitivity, diachrony, and language contact. In Silvia Luraghi and Elisa Roma (eds.), *Valency over Time: Diachronic Perspectives on Valency Patterns and Valency Orientation*, 13–30. Berlin, Boston: De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110755657>.
- Hartmann, Iren, Martin Haspelmath, and Bradley Taylor (eds.). 2013. *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <http://valpal.info> (accessed August 15, 2022).
- Haspelmath, Martin. 2001. The European linguistic area: Standard Average European. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible (eds.), *Language Typology and Language Universals*, vol. 2, 1492–1551. Berlin: Mouton de Gruyter.
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86(3): 663–687.
- Haspelmath, Martin. 2015. Transitivity prominence. In Andrej Malchukov and Bernard Comrie (eds.), *Valency Classes in the World's Languages*, 131–148. Berlin/Boston: De Gruyter Mouton.
- Joseph, Brian D. 1983. *The Balkan Infinitive*. Cambridge: Cambridge University Press.
- Joseph, Brian D. 2010. Language contact in the Balkans. In Raymond Hickey (ed.), *The Handbook of Language Contact*, 618–633. Oxford: Wiley-Blackwell.
- Kassambara, Alboukadel, and Mundt, Fabian. 2020. *factoextra*. Available online at <https://CRAN.R-project.org/package=factoextra> (accessed August 15, 2022).
- Kiparsky, Paul. 1998. Partitive case and aspect. In Miriam Butt and Wilhelm Geuder (eds.), *The Projection of Arguments: Lexical and Compositional Factors*, 265–307. Stanford, CA: CSLI Publications.
- Koptjevskaja-Tamm, Maria and Bernhard Wälchli. 2001. The Circum-Baltic languages.

- An areal-typological approach. In Östen Dahl and Maria Koptjevskaja-Tamm (eds.), *Circum-Baltic Languages* Vol. 2., 615–750. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Kramer, Johannes. 1981. Die Übernahme der deutschen und der niederländischen Konstruktion Verb + Verbzusatz durch die Nachbarsprachen. In Wolfgang Meid and Karin Heller (eds.), *Sprachkontakt als Ursache von Veränderungen der Sprach- und Bewusstseinsstruktur. Eine Sammlung von Studien zur sprachlichen Interferenz*, 129–140. (Innsbrucker Beiträge zur Sprachwissenschaft 34.) Innsbruck: Institut für Sprachwissenschaft.
- Kurzová, Helena. 2019. Defining the Central European convergence area. In Alexandra Y. Aikhenvald and Robert M.W. Dixon (eds.), *Slavic on the Language Map of Europe*, 261–290. De Gruyter Mouton.
- Levinson, Stephen C., Meira, Sérgio, and The Language and Cognition Group. 2003. ‘Natural concepts’ in the spatial topological domain – Adpositional meanings in crosslinguistic perspective: An exercise in semantic typology. *Language* 485–516.
- Malchukov, Andrej and Bernard Comrie (eds.). 2015. *Valency Classes in the World's Languages*. Berlin / Boston: De Gruyter Mouton.
- Michaelis, Susanne Maria. 2019. World-Wide Comparative Evidence for Calquing of Valency Patterns in Creoles. *Journal of Language Contact* 12(1): 191–231.
- Miestamo, Matti. 2014. Partitives and negation: A cross-linguistic survey. In Silvia Luraghi and Tuomas Huomo (eds.), *Partitive Cases and Related Categories (Empirical Approaches to Language Typology 54)*, 63–86. Berlin and Boston: De Gruyter Mouton.
- Moroz, George. 2017. *lingtypology: easy mapping for Linguistic Typology*. Available online at <https://CRAN.R-project.org/package=lingtypology> (accessed August 15, 2022).
- Moroz, George. 2018. Маркирование актантов двухместных предикатов в польском языке [Markirovanie aktantov dvuxmestnyx predikatov v pol'skom jazyke]. In Sergey S. Say (ed.), *Валентностные классы двухместных предикатов в разноструктурных языках [Valentnostnye klassy dvuxmestnyx predikatov v raznostrukturnyx jazykax]*, 183–197. Saint Petersburg: ILS RAN.
- Oskolskaya, Sofia A., Natalia M. Zaika, Sergey B. Klimenko, and Maksim L. Fedotov. 2020. Defining caritive as a comparative concept. *Voprosy Jazykoznanija* 3, 7–25.
- R Core Team 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available online at <https://www.R-project.org/> (accessed August 15, 2022).
- Rusakov, Alexander Yu. 2013. Некоторые изоглоссы на албанской диалектной карте (К вопросу о возникновении и распространении балканизмов албанского языка) [Nekotorye izoglossy na albanskoj dialektnoj karte (K voprosu o vznikenii i rasprostranienii balkanizmov albanskogo jazyka)]. In Vyacheslav Ivanov

- (ed.), *Issledovanija po tipologii slavjanskix, baltijskix i balkanskix jazykov (preimuščestvenno v svete jazykovyx kontaktov)*, 113–174. St. Petersburg: Aleteyya.
- Rusakov, Alexander Yu. 2021. Contemporary language contacts in the Balkans: Situations and outcomes. *Between Separation and Symbiosis. South Eastern European Languages and Cultures in Contact*, 1–26. Boston, Berlin: De Gruyter Mouton.
- Say, Sergey, Sergey Dmitrenko, Dmitry Gerasimov, Viktor Khrakovskij, and Natalia Zaika. 2020. Instructions for contributors. In Sergey Say (ed.), *BivalTyp: Typological Database of Bivalent Verbs and Their Encoding Frames*. St. Petersburg: Institute for Linguistic Studies, RAS. Available online at <https://www.bivaltyp.info> (accessed August 15, 2022).
- Say, Sergey. 2014. Bivalent verb classes in the languages of Europe: A quantitative typological study. *Language Dynamics and Change* 4(1): 116–166.
- Say, Sergey. 2018. Маркирование актантов двухместных предикатов: предварительные итоги типологического исследования [Markirovanie aktantov dvухmestnyx predikatov: predvaritel'nye itogi tipologičeskogo issledovanija]. In Sergey S. Say (ed.), *Валентностные классы двухместных предикатов в разноструктурных языках* [Valentnostnye klassy dvухmestnyx predikatov v raznostrukturnyx jazykax], 557–616. Saint Petersburg: ILS RAN.
- Say, Sergey (ed.). 2020. *BivalTyp: Typological Database of Bivalent Verbs and Their Encoding Frames*. St. Petersburg: Institute for Linguistic Studies, RAS.
- Seifart, Frank. 2019. Contact-induced change. In Jeroen Darquennes, Joe Salmons, and Wim Vandenbussche (eds.), *Language Contact: An International Handbook*, 13–23.
- Seržant, Ilja A. 2015a. Dative experiencer constructions as a Circum-Baltic isogloss. In Peter Arkadiev, Axel Holvoet, and Björn Wiemer (eds.), *Contemporary Approaches to Baltic Linguistics*, 325–348. Berlin/München/Boston: De Gruyter Mouton.
- Seržant, Ilja A. 2015b. Independent partitive as a Circum-Baltic isogloss. *Journal of Language Contact* 8: 341–418.
- Seržant, Ilja A. 2021. Typology of partitives. *Linguistics* 59(4): 881–947.
- Seržant, Ilja A., Björn Wiemer, Eleni Bužarovska, Martina Ivanová, Maxim Makartsev, Stefan Savić, Dmitry Sitchinava, Karolina Skwarska, and Mladen Uhlik. 2022. Areal and diachronic trends in argument flagging across Slavic. In Eystein Dahl (ed.), *Alignment and Alignment Change in the Indo-European Family*.
- Seržant, Ilja A. (To appear). The Circum-Baltic Area. An Overview. In Jan Fellerer and Neil Bermel (eds.), *Oxford Guides to the World's Languages: The Slavonic Languages*. Oxford: Oxford University Press.
- Shagal, Ksenia. 2021. Bivalent patterns in Finnish. In Sergey Say (ed.), *BivalTyp: Typological Database of Bivalent Verbs and Their Encoding Frames*. St. Petersburg: Institute for Linguistic Studies, RAS. Available online at <https://www.bivaltyp.info> (accessed August 15, 2022).



- Stolz, Thomas. 2001. On Circum-Baltic instrumentals and comitatives: To and fro coherence. *Circum-Baltic Languages* 2: 591–612.
- Timberlake, Alan. 2017. Slavic Languages. In Raymond Hickey (ed.), *The Cambridge Handbook of Areal Linguistics (Cambridge Handbooks in Language and Linguistics)*, 331–355. Cambridge: Cambridge University Press.
- Törnquist-Plewa, Barbara. 2000. Contrasting ethnic nationalisms: Eastern Central Europe. *Language and Nationalism in Europe*, 183–220. Oxford University Press, Oxford.
- Trips, Carola, and Achim Stein. 2019. Contact-induced changes in the argument structure of Middle English verbs on the model of Old French. *Journal of Language Contact* 12(1): 232–267.
- Tsunoda, Tasaku. 1985. Remarks on transitivity. *Journal of Linguistics* 21(2): 385–396.
- van der Auwera, Johan. 2011. Standard Average European. In Bernd Kortmann and Johan van der Auwera (eds.), *The Languages and Linguistics of Europe: A Comprehensive Guide*, 291–306. Berlin/Boston: De Gruyter Mouton. DOI: <https://doi.org/10.1515/9783110220261>.
- Vanags, Pēteris. 1992. Locative in the Earliest Latvian Writings. *Journal of Baltic Studies* 23(4): 387–394.
- Vanags, Pēteris. 2011. *Problems of Standard Latvian in the 19th Century: Changes in the Language Planning and Changes in Planners, from Baltic-Germans to Latvians*.
- Vanags, Pēteris. 2019. German Influence on the Christian Discourse of Early Written Latvian. In Kauko, Mikko, Miika Norro, Kirsi-Maria Nummila, Tanja Toropainen, and Tuomo Fonsén (eds.), *Languages in the Lutheran Reformation: Textual Networks and the Spread of Ideas*, 273–302. Amsterdam: Amsterdam University Press. DOI: <https://doi.org/10.2307/j.ctvfrxqjc>.
- WATP. 2014. *The World Atlas of Transitivity Pairs (2014)*. Tokyo: National Institute for Japanese Language and Linguistics. Available online at: <http://verbpairmap.ninjal.ac.jp> (accessed August 15, 2022).
- Wichmann, Søren and Jan Wohlgemuth. 2008. Loan verbs in a typological perspective. In Thomas Stolz, Dik Bakker, and Rosa Salas Palomo (eds.), *Aspects of Language Contact. New Theoretical, Methodological and Empirical Findings with Special Focus on Romancisation Processes*, 89–122. Berlin/New York: Mouton de Gruyter. DOI: <http://doi.org/10.1515/9783110206043.89>.
- Wiemer, Björn. 2019. Matrěška and areal clusters involving varieties of Slavic: On methodology and data treatment. In Alexandra Y. Aikhenvald and Robert M.W. Dixon (eds.), *Slavic on the Language Map of Europe*, 1st ed., 21–61. Berlin and Boston: De Gruyter Mouton.
- Wolter, Ernst. 1908. Die lettische Literatur. In Adalbert Bezzenberger (ed.), *Die osteuropäischen Literaturen und die slawischen Sprachen* vol. 9, 372–377. BG Teubner.
- Zaika, Natalia M. 2018. Маркирование актантов двухместных предикатов в литов-

ском языке [Markirovanie aktantov dvухmestnyх predikatoв v litovskom jazyke]. In Sergey S. Say (ed.), *Валентностные классы двухместных предикатов в разноструктурных языках* [*Valentnostnye klassy dvухmestnyх predikatoв v raznostrukturnyх jazyках*], 198–210. Saint Petersburg: ILS RAN.