Differential Argument Marking with the Latvian debitive

A multifactorial analysis

Ilja A. Seržant[i, ii] and Jana Taperte[ii, iii]
[i]Johannes-Gutenberg-University Mainz / [ii]University of Vilnius /
[iii]Latvian Language Institute of the University of Latvia

The present paper investigates the NP-triggered Differential Argument Marking phenomena (DAM) found in the debitive construction of Latvian. There are two perspectives: the synchronic perspective aims at providing a coherent description of the DAM in Contemporary Latvian, while comparison with data from Old and Early Modern Latvian allows for a diachronic perspective. The arguments of the debitive, A/S and P, are marked non-canonically by the dative and nominative case, respectively. The emergence of new, canonical case-marking strategies, namely, ACC on the P argument and occasionally NOM on the S argument, has created differential marking for the S (DAT/NOM) and for the P argument (NOM/ACC). We claim that the appearance of these new case-marking strategies is the result of the increasing degree of grammaticalization of the debitive as well as the pressure of the canonical case-assignment patterns. Thus, the debitive incipiently and gradually loses its original lexical properties such as the own case frame (stemming from the possessive predicate) in favor of the case frame of the embedded lexical verb, thereby becoming similar to an auxiliary. In turn, the appearance and spread of the new case-marking strategies is conditioned by various factors and constraints that are established through a multifactorial analysis. Thus, ACC marking proceeds along the accessibility scale starting from the most-accessible NP types, and is additionally conditioned by linear position, animacy and the semantic class of the lexical verb embedded under the debitive. In contrast, the sporadic appearance of the NOM marking on S, i.e. the failure of the underlying NOM to be turned into the DAT otherwise required by the debitive, is found with NPs with low-individuated referents and may appear with existential verbs only. The postverbal position is an additional attracting factor here.

DOI 10.1075/vargreb.3.05ser

1. Introduction[1]

In this pilot study, our aim is to scrutinize the changes in case-assignment that are taking place in the debitive construction of Contemporary Latvian in contrast to the reconstructed state, Old Latvian and Early Modern Latvian. In our multifactorial analysis we establish various factors determining the choice of the new case marking, such as animacy, accessibility degree, the linear position of the NP as well as the semantic verb class of the embedded lexical verb and some others factors.

The debitive construction primarily encodes different kinds of necessity, as well as related epistemic meanings (see, *inter alia*, Bergmane (ed.) 1959: 615; Fennel 1973; Holvoet 2007; Kalnača & Lokmane 2014). It is formed by means of the auxiliary *būt* 'to be' (optional in the present affirmative, cf. Stolz 1987: 225) and the non-finite debitive form of the lexical verb marked by the prefix *jā-* (henceforth: *the debitive gerundive*). The debitive construction imposes its own case frame, overriding the regular case assignments of the lexical verb in the following way: A/S arguments (also referred to as (non-canonical) subjects, cf. Fennell 1973; Stolz 1987) are marked by the dative case (DAT), while P is marked by the nominative case (NOM) with all NP types except for personal pronouns denoting speech act participants (SAP) and the reflexive anaphora. The latter are obligatorily marked by accusative case (ACC). By way of example, consider the transitive NOM-ACC verb *redzēt* 'to see' in (1) and the corresponding debitive construction in (2) in which the arguments are marked by DAT and NOM, respectively, while (3) illustrates the SAP pronoun triggering ACC on the object-like argument:[2]

(1) (Standard Latvian)
Kāpēc es šo filmu redzu?!
why 1SG.**NOM** DEM.**ACC**.SG film.**ACC**.SG see.PRS.1SG
'Why do I watch this film?!' [Constructed example]

(2) (Standard Latvian)
Kāpēc man šī filma ir
why 1SG.**DAT** DEM.**NOM**.SG.F film.**NOM**.SG be.PRS.3
jā-redz?!
DEB-see
'Why do I have to watch this film?!' [Constructed example]

---

1. Seržant designed the study, performed the statistical analyses and wrote the paper; Taperte was responsible for the data collection.

2. Cf. Timberlake 1974: 140–154, Seržant, this volume, on the object analysis of the second NOM/ACC argument of the debitive (differently Holvoet & Grzybowska 2014).

(3) (Standard Latvian)
Kāpēc man tevi ir jā-redz?!
why 1SG.DAT 2SG.ACC be.PRS.3 DEB-see
'Why do I have to see you?!' [Constructed example]

While this is the situation found in the normalized, conservative Standard Latvian, Contemporary Latvian[3] is undergoing the expansion of the ACC marking onto the other NP types. Moreover, under very special conditions to be laid out in detail below, the DAT marking of the subject-like argument is rarely replaced by the NOM. Thus, both arguments of the debitive alternate their case-marking.

The phenomenon of Differential Argument Marking (DAM) is widely discussed in the literature under different headings (Differential Object/Subject/Agent Marking, Optional Ergative Marking, etc., cf., *inter alia*, Bossong 1982, 1998; De Hoop & de Swart 2008; Fauconnier 2011; McGregor 1992, 1998). DAM entails that different marking strategies are available for one and the same argument bearing one and the same macrorole (Seržant & Witzlack-Makarevich, to appear). In particular, we will be dealing in this paper with *restricted, NP-triggered Differential Argument Marking* (according to the classification in Seržant & Witzlack-Makarevich, to appear), that is to say, both the NOM/ACC alternation and the rare DAT/NOM one are conditioned by *NP-internal* properties and both are *restricted* to the debitive predicate only; moreover, the rare DAT/NOM alternation is additionally *restricted* to the debitive predication formed out of existential and similar verbs (see details in 5.3).

In our investigation, we scrutinize these two DAM systems separately in four idealized diachronic stages: (i) the reconstructed state, namely, the construction that was the source of the later debitive construction, (ii) Old Latvian, (iii) Early Modern Latvian of time span 1850–1900 and (iv) Contemporary Latvian. Our study is both synchronic and diachronic: we provide synchronic descriptions of both DAM systems at each of these stages (albeit with the main focus on Contemporary Latvian) and reveal the changes that have occurred from the reconstructed state to Contemporary Latvian in a multifactorial approach.

In order to do this we proceed as follows. We first present our four samples and the way they have been compiled (Section 2). In Section 3, we discuss the reconstruction of the debitive construction. The main discussion of the paper follows in the next two sections: we scrutinize both the differential marking of the object-like argument (Section 4) and the rare differential marking of the subject-like referent

3. We take Contemporary Latvian to be here the average linguistic system resulting from some sort of conglomeration where the conservative standardized language is intermingled with colloquial and often innovative features.

(Section 5), taking into account the four idealized diachronic stages. Section 6 summarizes the motivations and factors that constrain the expansion of the new case-markings and provides conclusions.

2. Our samples

Before we turn to the discussion Sections 3–5, we somewhat extensively describe our samples of Old Latvian (2.1), Early Modern Latvian (2.2) and two samples of Contemporary Latvian, the first covering differential object marking (2.3) while the second exemplifies differential subject marking (2.4). Notably, none of the samples includes occurrences of a control or raising verb in the debitive construction; we can thus make sure that the relevant case is assigned by the debitive construction and not by a lower verb. Finally, in 2.5, we discuss some philological preconditions that allow us comparing these samples with each other and taking them as representatives of the three diachronic stages of *mutatis mutandis* one and the same language.

2.1 The Old Latvian sample

Old Latvian texts are mostly translations made by German clerics whose skills in the Latvian of that time were varying. Nevertheless, the data gathered from different texts thanks to the digitalized collection of Old Latvian texts SENIE provides a coherent and uniform picture of the case-marking patterns, attesting to the rigid case frame DAT-NOM. Hence, in our view, it can be taken to be representative for the earliest attested language layer despite the general non-native character of the texts.

Our sample consists of 200 entries that were found by searching the corpus SENIE by the combinations *ja-* (sic!)[4] with 5 (111 relevant hits), 4 (79 relevant hits) and 3 (10 relevant hits) subsequent letters. The following verbs are contained in our Old Latvian sample: *art* 'to plow' (3), *berzēt* 'to oil' (1), *braukt* 'to drive, to ride' (1), *cerēt* 'to hope' (1), *darīt* 'to do' (26), *domāt* 'to think' (5), *dot* 'to give' (16), *ēst* 'to eat' (19), *gādāt* 'to take care of' (7), *gaidīt* 'to wait' (7), *gulēt* 'to sleep, to lie' (5), *iet* 'to go' (23), *kāpt* 'to ascend' (3), *karot* 'to be at war' (2), *kopt* 'to take care of' (1), *krāt* 'to save' (1), *likt* 'to put' (4), *lūgt* 'to ask' (3), *mācīt* 'to teach' (1), *maksāt*

4. While the debitive marker *jā-* is actually a long syllable, the Old Latvian writings often do not indicate lengths. Thus, *ja-* may be read as [jaː] or [ja]. There are also spellings such as *jaa-* which we have not taken into account.

'to pay' (6), *meklēt* 'to look for' (2), *mest* 'to throw' (2), *mirt* 'to die' (11), *mist* 'to inhabit, to live' (1), *nākt* 'to come' (6), *ņemt* 'to take' (4), *nest* 'to carry' (5), *pirkt* 'to buy' (1), *pūst* 'to blow' (1), *raudāt* 'to cry' (1), *raudzīties* 'to look' (1), *redzēt* 'to see' (10), *runāt* 'to speak' (7), *sacīt* 'to say' (5), *sildīt* 'to warm' (1), *slavēt* 'to praise' (1), *stāvēt* 'to stay' (2), *turēt* 'to hold' (2), *vest* 'to carry' (1).

2.2 The Early Modern Latvian sample

For a statistically coherent picture of the Early Modern Latvian period, we have collected examples for exactly the same verbs as in the Contemporary Latvian Sample (see below) in newspapers from 1850 to 1900, using the newspaper corpus PERIODIKA. The intention was for the Early Modern Latvian Sample to be analogous to, and hence comparable with, the Contemporary Latvian Sample. However, in contrast to 3193 examples in the Contemporary Latvian Sample, we found only 399 examples for the same verbs in the newspapers of the aforementioned period. Among those listed in 2.3 below, only the following verbs are attested in the debitive form in the newspapers of this period: *apvainot* (1) 'to offend', *kaitināt* (2) 'to annoy', *klausīties* (37) 'to listen', *lasīt* (110) 'to read', *ņemt* (53) 'to take', *nest* (52) 'to carry', *pirkt* (26) 'to buy', *rakstīt* (65) 'to write', *sacīt* (52) 'to say', *trenkāt* (1) 'to push in, chase'. The design of annotations and the makeup of this sample is the same as in 2.3.

2.3 The Contemporary Latvian Sample

In order to statistically assess the distribution of ACC vs. NOM with different NP types we have created a sample of total 3193 entries, manually collecting authentic examples by Google search. We included all examples found by Google, excluding repletion and dubious examples (non-authentic, spelling errors, etc.). In order to have a less biased sample we always included all examples found by Google but no more than the first 200 relevant hits for each verb. Thus, the examples stem from different sources and registers, though the colloquial register prevails; cf. Table 3 below.

Our Google search string contained the following verbs in their debitive forms: *apsaukāt* 'to call so. offensive names' (number of hits: 29), *apvainot* 'to offend' (68), *besīt* 'to annoy' (colloquial only) (7), *dzirdēt* 'to hear' (68), *iznīcināt* 'to destroy' (180), *kaitināt* 'to annoy' (159), *klausīties* 'to listen' (129), *laist* 'to admit' (192), *lasīt* 'to read' (111), *mīlēt* 'to love' (160), *ņemt* 'to take' (203), *nest* 'to carry' (113), *piekaut* 'to pummel' (43), *pirkt* 'to buy' (134), *provēt* 'to test, to try (in its lexical meaning)' (colloquial only) (32), *pucēt* 'to dress, clean' (colloquial only) (64), *rakstīt* 'to write' (238), *redzēt* 'to see' (112), *sacīt* 'to tell, to say' (182), *satikt*

'to meet' (205), *sazvanīt* 'to call' (95), *slēgt* 'to close' (202), *stūķēt* 'to push, to slide' (114), *tēlot* 'to imitate, to feign' (92), *trenkāt* 'to chase' (colloquial only) (61).[5] We aimed at a balanced collection of verbs allowing animate, inanimate and both kinds as the object input: (i) verbs with animate objects *jā-apsaukā* 'have to call someone offensive names', *jā-piekauj* 'have to pummel so.'; (ii) verbs with inanimate objects *jā-pērk* 'have to buy', *jā-provē* 'have to test'; (iii) verbs used with both animate and inanimate objects, e.g., *jā-pucē* 'have to dress up smb., have to clean smth.', *jā-redz* 'have to see', *jā-dzird* 'have to hear'.

We have purposely aimed at such a high number of examples in order to warrant the solidity of the linguistic results to be discussed below. Specifically, the high number of examples from different source texts, we think, makes the sample into a kind of a sociolinguistic average glossing over various sociolinguistic factors such as variation across different speakers, generations or styles and thus provides a sample representative for an average speaker of Contemporary Latvian. Our contention is that the aforementioned factors might have the strong impact on the variation in case selection but only if a considerably smaller number of examples had been selected which could have allowed accidental biases for or against a particular social group or speaker in the sample.

Furthermore, we purposely selected for the study those verbs that we expected to occur more frequently in free-style writings (such as blogs, forums and social networks) than in official, highly standardized texts. Thus, our aim was to create a sample more representative of the colloquial than of the standard register. Finally, we selected verbs from different semantic classes in order to have an evenly balanced sample in this respect as well:

Table 1. Distribution of verbal semantic classes in the sample

Semantic class	Total hits
Agentive verbs	48% (1518)
Experiencer verbs	23% (731)
Verbs of communication	12% (369)
Verbs of transfer	11% (337)
Verbs of meeting	6% (205)
Conative verb	1% (32)

5. Although there is an annotated corpus of Latvian, it did not provide reasonable search facilities for our study. Its automatic annotation of Latvian had no tools to resolve homonymy (which would have been important for our study); moreover, given the free word order of Latvian and lack of syntactic annotation there was no way to look for all ACC and for all NOM objects with the debitive of a particular lexical verb.

The examples were annotated according to the following factors:

Table 2. Annotated factors and their values

Case	**ACC, NOM**
Verb class	Agentive, Experiencer, Meeting, Conative, Transfer, Communication
Aspectuality	Prefixed, non-prefixed (verbs)
Position relative to the debitive predicative	Before, after
Definiteness of the object NP	Definite, indefinite
Animacy of the object NP (animals count as animate)	Animate, inanimate
Animacy of the subject NP	Animate, inanimate
Part of speech of the object NP	Full NPs,[1] third person pronoun,[2] demonstrative pronoun,[3] reciprocal pronoun,[4] indefinite pronoun,[5] proper name, relative and question (*wh*) pronoun,[6] universal quantifier *viss*

1: nouns modified (e.g. by a demonstrative, quantifier, etc.) and bare nouns, bare quantifiers such as *daži* 'some', *viens* 'one' (in the relevant sense);

2: third-person pronouns: *viņ-*;

3: demonstrative pronouns (independent, not modifying a noun; also if used anaphorically): *tāds* 'such (as this one)', *tas* 'that one', *šis* 'this one', *šitais* 'this one' and their combinations;

4: reciprocal pronouns: *cits citu*, *viens otru*;

5: indefinite pronouns (independent, not modifying a noun): *kāds* 'some, someone', *jebkurš* 'anyone', *kaut kas* 'something', *cits* 'other', *kas cits* 'some other', *kas* 'some, any, something, someone', *katrs* 'each', *nekas* 'none, nothing', *neviens* 'no one', *viens* 'some';

6: *wh*-pronouns – relative and question (*wh*) pronouns: *kas* 'what/who', *kurš* 'what/who/which', *kāds* 'which'.

Furthermore, we additionally classified the examples into *colloquial* and *unclear*. The former is meant to be clearly colloquial examples with spontaneous speech from blogs, comments, etc., containing slang words, smiles, etc., while *unclear* stands for contexts whose classification as colloquial either was not entirely clear or which were unequivocally non-colloquial text. The following table shows the proportion in the sample:

Table 3. The degree of Contemporariness of the sample

Unequivocally colloquial	60% (1756)
Non-colloquial or examples with unclear classification	40% (1188)

Our hypothesis here was that unequivocally colloquial contexts would show higher percentages of ACC than the other contexts did. This hypothesis was not corroborated by the data. Therefore, in what follows we ignore this annotation.

2.4 The Contemporary Latvian Sample 2: The marking of the subject-like argument

Finally, we collected another sample with examples of the dative marking being replaced by the nominative one or, to be more precise, with the original nominative marking unexpectedly not overridden by the dative of the debitive. Here we thank Andra Kalnača, who first drew our attention to these examples. While being rare, the nominative use has already attracted the attention of purists and been condemned as not original Latvian, along with the accusative replacing the older nominative (as discussed above): "The Latvian case government is currently endangered: other languages' cases slip into Latvian necessity sentences – first, the accusative" … "and then also the nominative (***tas*** *tā jā-būt*). This still sounds repulsive to the non-distracted language feeling." (Grīsle 2005: 8).[6] Grīsle (2005: 8) herself quotes an example that she has come across:

(4) (Contemporary Latvian)
tas *tā* *jā-būt*
DEM.NOM.SG.M SO DEB-be
'This has to be this way.'

The occurrence of the NOM marking is quite rare and not acceptable for many speakers; crucially and to anticipate the skeptical reader, we have collected only those examples that we could assume were produced by native speakers. For this we checked the context of each of the example in the sample to see whether there were mistakes typical of a non-native speaker. Having said this, we cannot entirely rule out the possibility that in particular instances the nominative marking is in fact due to typos and/or spelling errors, since we obtained our examples from written sources. In total, given the quantitative and qualitative evidence, this phenomenon cannot be reduced to just performance errors or sloppy writing. Moreover, as we argue below, the NOM is found in semantically clearly definable contexts (see 5.1 below).

This sample consists in total of 126 examples, most of which contain the verb *būt* 'to be' (121) including a number of copular uses of the verb *būt* with the

6. "Latviskā rekcija patlaban ir apdraudēta: latviešu vajadzes teikumos iespiežas citu valodu locījumi – vispirms akuzātīvs … pēc tam arī nominātīvs (***tas*** *tā jābūt* …). Nenotrulinātai valodas izjūtai tas vēl skan pretīgi." (Grīsle 2005: 8).

resultative/passive participle (glossed as PPP below) or with predicatively used nouns and adjectives alongside *kļūt* 'to become' (1), *palikt* 'to remain' (1) and *piemist* 'to be characteristic of, be inherent to' (3).

Since the sample was collected by Google search we could not look for just any nominative marked subject-like argument in the combination with a debitive predicate. Instead, we have searched for, and entered into our sample all occurrences of a particular debitive form with a nominative (both ways preceding and following the debitive gerundive) of *tas/tā* 'that one (M/F.SG)', *šis/šī* 'this one (M/F.SG)', *viņš/viņa* 'he/she', *cilvēks* 'person', *iespēja* 'possibility (M.SG)' and some other nominative NPs with the following debitive gerundives: *jābūt* 'DEB-be', *jā-paliek* 'DEB-stay', *jā-kļūst* 'DEB-become', *jā-piemīt* 'DEB-be inherent to'. For example, the subset of the copular uses is selective and determined by the search string as we looked for particular lexical items only. At the same time, we have preliminary and selectively checked whether or not other types of verbs – such as transitive and agentive intransitives – attest this phenomenon in the debitive. To conclude, this sample is not fully exhaustive or representative of the lexical range that occasionally admits the nominative marking of the subject-like argument.

2.5 On the comparability of the samples

Before we turn to the discussion part, we provide argumentation for why the data collected for the three idealized stages, namely Old Latvian, Early Modern Latvian and Contemporary Latvian, can be compared with each other.

We argue that, first, there is dialectal continuity among them: all three stages represent exactly the same dialect, namely, the Middle Latvian dialect. This can be demonstrated independently, for example, on the basis of phonetics. Both High Latvian and Livonian dialects exhibit quite prominent phonetic features that make their identification unproblematic. For example, High Latvian (= Latgalian) has demonstrably very early undergone a sort of "great vowel shift" with basically no Proto-Latvian vowel remaining in the same place (cf., *inter alia*, Seržant 2005) while Middle Latvian phonetics remained very conservative. Livonian subdialects, on the other hand, exhibit such features as drastic shortening of word final syllables or such grammatically remarkable phenomena as the total loss of the gender distinction. Nothing of this kind could be observed in the Old Latvian or Early Latvian texts we have used. Rather, the dialect found there was unequivocally Middle Latvian. Having said this, for obvious reasons, we cannot make claims about a particular Middle Latvian subdialect from which the changes to be discussed have started to appear, and whether and/or when similar processes might have occurred in the other Latvian dialects.

Furthermore, another problem that a skeptical reader might raise is that there is a difference between the Old Latvian data, on the one hand, and the Early Modern and Contemporary Latvian data, on the other. Old Latvian was mainly written by German translators, while Early Modern and Contemporary Latvian data stem from native speakers. This is, we concede, an important difference. Nevertheless, we argue that the Old Latvian sample is a valid data source to the extent needed for this study and, hence, may be compared with the other two: the case marking at issue is remarkably uniform across the Old Latvian texts of the sample, as our counts below (see 4.2 and 5.2) show. The alternative hypothesis that the German translators consistently perceived and reproduced the wrong case is not plausible. Still another alternative hypothesis, namely that the uniformity of the case-marking is due to their native language is even less plausible, for the simple reason that a German DAT-NOM pattern expressing necessity simply does not exist. Even if there were some influence from their native language, it would result in a different pattern, forcing the ACC and not the attested NOM marking of the object-like argument.

Furthermore, for historical reasons, it is highly unlikely that Old and Early Modern Latvian were subject to such specific prescriptivist rules that would prescribe the patterns we are interested in, namely, the word order and case assignment of the debitive construction. This is corroborated by philological evidence. It has been discussed in the philological literature that there is hardly any direct mention of the case-marking of the object-like argument in the early Old Latvian grammars (Fennell 1984: 252; Andronovs 1998: 156, 163, 165, 172). The important grammar *Manvdvctio Ad Lingvam Lettonicam*… (published in 1644) and most of the subsequent grammars up to Stender's of 1761 do not provide any direct information on this (Fennell 1984: 252; Andronovs 1998: 156, 163, 165, 172).

Although questions about language standardization planning have been addressed ever since Georg Mancelius (1593–1654) (cf. Vanags 2012), the ideas of language cleansing – first of all as regards the lexicon and a dedicated Latvian orthography but not case patterns – developed only in the second half of the 19th c. (around 1850–1860) in the circle of the so-called Young Latvians (*jaunlatvieši*). In turn, massive language standardization of the grammar came later, in the beginning of the twentieth century in the circle of Riga's Latvian society (*Rīgas Latviešu biedrība*). Thus, Rūķe-Draviņa (1977) argues that the most important role in the standardization process of Latvian was played by the grammar of K. Mühlenbach and J. Endzelīns from 1907 (*Latweeschu gramatika*) containing a number of syntactic treatments. Our Early Modern Latvian sample is purposely restricted to periodicals from 1850 to 1900 – a time when the grammar must have been least affected by this process.

Finally, the three samples are somewhat different as to the diversity of their lectal registers. While Old and Early Modern Latvian samples are based on the language of religious texts, Contemporary Latvian is a collection of examples found by Google search, representing a wider range of registers with a strong inclination towards colloquial language. However, notably, even our Old Latvian sample contains different text genres including not only religious but also secular texts, for example, the Latvian translation of the Swedish army statutes *Krigz-Artiklar*… from 1696. The same holds for Early Modern Latvian, which is based on various periodicals of that time. Moreover, it is highly improbable that the difference between the language of religious texts and the colloquial language could have been significant, given that the language of religious texts is not older than 16th c.

3. The origin of the debitive construction: The reconstructed state

3.1 Reconstruction of the general pattern

Before we eventually turn to the discussion of the diachronic processes found in the argument marking of the debitive construction, in this section we provide a brief overview of the grammaticalization path of this construction. The comparative evidence suggests that the development of the debitive is recent, most probably shortly before the literary tradition, i.e. around the 14th–15th centuries. No traces of the debitive can be found in the closely related Lithuanian (Wälchli 2000: 202), and it is only barely found in High Latvian, which regularly employs the synonymous modal predicate *vajadzēt* 'have to' (Nau 2012: 482). Most probably the debitive originated in the Middle Latvian (serving as the basis for Contemporary Latvian) and/or Livonian dialect, and then only later and only partially affected High Latvian.

It is commonly assumed that the debitive construction emerged from a biclausal construction out of the *mihi-est* type possessive construction in the main clause and a relative clause formed by means of the generalized relativizer *jā-* (dialectally also *ju-*) and the respective lexical verb (Endzelīns 1951: 971–3; Holvoet 1998),[7] similarly to the present-day construction in (5):

7. This analysis may be in fact as old as the first grammars of Old Latvian. Thus, Adolphi in his grammar *Erster Versuch Einer kurz-verfasseten Anleitung Zur Lettischen Sprache* (published in 1685) translates the debitive into Latin with a relative pronoun *quod* 'what': *ja=ehd* 'DEB-eat' *quod edat* (Andronovs 1998: 159).

(5) *Man ir maize, ko ēst*
1SG.DAT be.PRS.3 bread.NOM.SG REL.ACC eat.INF
'I have bread (which) to eat.'

Semantically, it parallels the English debitive modal *have to,* which developed from *I have something to eat* into *I have to eat something* (cf. Wälchli 2000). The original possessive meaning is typical of 16th-century Old Latvian, while already the 17th century more frequently attests obligation meanings than the possessive ones (Vanags 2000: 152). The Latvian counterpart of the English infinitive *to eat* is coded by the debitive gerundive (details in 3.1.1 below); both represent non-inflected, non-finite verb forms syntactically dependent on an auxiliary. In the process of grammaticalization, this gerundive and the possessive clause merged into one predication.

The following examples from Old Latvian attest the more original possessive meaning and the bi-clausal structure consisting of: the dative (predicative) possessor *töw* '2SG.DAT' in (7) and *jums* '2PL.DAT' in (6), ellipsis of the nominative possessee (implying indefinite 'any') in (7) and overt possessee in *kas* 'INDEF.NOM' in (6), the existential *Irra-g* in (6)/*gir* in (7) 'is' in its lexical meaning (to be later the auxiliary in the debitive) and the prefix *ja-* (etym. *jā-*) (sometimes written separately in Old Latvian) with the lexical verb *-maxa* 'to pay' (formally in the third person present):

(6) (Old Latvian, SENIE, *Lett. Handbuch* 1685, p.41, 15)
Irra=g jums śche kas jaehd?
be.PRS.3=Q 1PL.DAT here INDEF.NOM DEB.eat
'Do you have here something to eat?'

(7) (Old Latvian, SENIE, Manc1637, p.73, 18)
Aißto ja töw nhe gir jamaxa
therefore when 2SG.DAT NEG be.PRS.3 DEB.pay
tad töw tawu Ghulltu appackßchan töw
then 2SG.DAT your.ACC.SG bed.ACC.SG below 2SG.DAT
attjembs.
take.away.FUT.3
'Therefore, if you don't have [what] to pay, then you will be disseized of your bed under you.'

In addition to the possessively used matrix verb *būt* 'to be' the source construction could be used on the existential reading of this verb, cf. the following example in which the possessive reading of the verb is excluded:

(8) (Old Latvian, Ps 95a, Vanags 2000: 147)
Winge *slawe* *gir* *youke* *vnde*
3.GEN.SG.M praise.NOM.SG be.PRS.3 nice.NOM.SG.F and
mylige *ya* *klouse*
lovely.NOM.SG.F DEB listen
'His praise is beautiful and lovely to listen to.'[8]

The word-order distribution found in Old Latvian supports the analysis suggested. Thus, despite the fact that the regular word order in Latvian is SVO, the debitive construction predominantly attests NOM-DEB – i.e. the NOM argument preceding the debitive gerundive – in our Old Latvian sample. This is accounted for by the above explanation, namely, that the nominative argument is the subject (possessee) of the higher possessive predicate in Old Latvian:

Table 4. Position relative to the debitive gerundive

NOM-DEB	DEB-NOM	n/a
92% (105)	8% (9)	85

Furthermore, the subject-like argument is consistently marked by the dative case, or is just left unexpressed for generic or discursively anchored referents and on the existential reading of the verb:

Table 5. Case marking of the subject-like NP

Dative	Ablative-like PP	Impersonal
130	1	68

In one instance, the ablative-like preposition *no* 'from' with the genitive case is found, most probably in analogy to passives which – in contrast to present-day Latvian – could employ this PP to code the agent phrase in Old Latvian, in accordance with the German passives with *von* 'from'.

3.1.1 *The origin of the debitive gerundive*

As to the morphological shape of the debitive gerundive, it has been argued in the literature that it was originally formed on the basis of the infinitive, which is also the source of the modal meaning (Endzelīns 1901, 1905; Holvoet 1998). Indeed, the verb *būt* 'to be' forms the debitive gerundive from the infinitive form, which might be an indication of this being the original morphological pattern, cf. *jā-būt*

8. Cf. the German original: *Sein Lob ist schën vnd lieblich an zu hören*

(DEB-be.INF) 'have/has to be', because this form is morphologically isolated. All other debitive forms are morphologically derived from the third person present of the lexical verb, cf., e.g. *jā-brauc* (DEB-drive.PRS.3) 'have/has to drive' except for *jā-iet* 'have/has to go' which is ambiguous in not morphologically distinguishing the third person from the infinitive, both being *iet* '(s)he/it/they go(es), to go'. Problematic for this account is that there are only two attestations of the debitive gerundive in the unpublished manuscript by Elger derived from the respective infinitives in Old Latvian: *jā-runāt* and *jā-tiesāt* (originally *iarunnat, iatêsôt* EvE 37,24) (Vanags 2000: 153 fn. 23), while the third-person present forms are used elsewhere, also with the same verbs. These and other facts lead Wälchli (2000) to suggest that the third person form underlying the debitive gerundive is original, while the infinitive in *jā-būt* (DEB-be.INF) is secondary. Indeed, the data attested support this view. Thus, Vanags (2000) and Andronovs (1998: 156–157) argue that the debitive gerundive of *būt* 'to be' is late, which is also expected given that the source construction was originally used only with transitive verbs and the appearance of intransitive verbs is only possible under the secondary, obligation meaning (Vanags 2000). Hence, *jā-būt* (DEB-be.INF) must be young. Furthermore, since the debitive construction has functionally replaced the old modal-infinitive construction of Baltic extensively discussed in the literature (cf., *inter alia*, Kiparsky 1969; Timberlake 1974), the infinitive-like form of *jā-būt* might be the result of some kind of morphological adaptation to this former, synonymous construction which was crucially based on infinitives.

Having said this, we are nevertheless inclined to adhere to the traditional account assuming that the infinitive is original and the source construction must have been structurally similar to (5) above. The main reason for this is that it is not clear how finite present verb forms would have yielded the modal meaning inherent to the debitive construction, as has been pointed out by Holvoet (1998). In languages with necessity expressions from originally possessive constructions, the obligation meaning stems from the infinitival verb forms (Bybee et al. 1994: 184). Bybee et al. (1994: 184) argue that the metaphorical possession of an activity yields the prospective, "the projected sense" precisely because the lexical verb is not in a form denoting the present (or past) tense. Indeed, even in Baltic, the modal meaning is typically connected to the non-finiteness of the lexical verb, cf. Wälchli (1996, 2000), who also compares the Latvian debitive with the synonymous constructions in the neighboring Livonian and Estonian dialects which are based on an infinitive.[9]

9. One solution to this (suggested in Wälchli 2000: 206–207) is the assumption that the debitive gerundive (in our terms) developed prior to the rise of the debitive construction and became very early a non-analyzable infinite verb form. While this is a possible option, it is not very likely

It is however possible that the replacement of the infinitive by the third-person verb form might have taken place before the subordinate clause merged into one verb form. Thus, Fraenkel (1928: 49, fn. 1) following Mühlenbach (1907) argues for precisely this scenario on the basis of the following examples from Lithuanian:

(9) (Dialectal Lithuanian, Fraenkel 1928: 47)
ar buvo kas valgo?
Q be.PST.3 INDEF.NOM eat.PRS.3
'Was there something to eat?' [lit. 'was there something (that one) eats?']

Notably, also for Fraenkel (1928) examples such as (9) are historical modifications of the original construction with an infinitive as in (10):

(10) (Lithuanian, constructed)
nebuvo kas valgyti
NEG.be.PST.3 what.NOM eat.INF
'There was nothing to eat.' [lit. 'there is nothing one could eat']

The question about the original form of the verb in the subordinate clause is immediately related to the origin of the debitive marker.

3.1.2 *The origin of the debitive marker* jā-*: An old relativizer or conjunction?*
The reconstructed source construction of the debitive has at least since Endzelin (1901; 1905) been compared with another construction that is productive in the present-day Latvian and Lithuanian, and certain conclusions about the origin of the debitive construction have been made on the basis of this modern construction, cf. (5) repeated here as (11) for convenience:

(11) *Man ir maize, ko ēst*
1SG.DAT be.PRS.3 bread.NOM.SG REL.ACC eat.INF
'I have bread (which) to eat.'

Indeed, this construction is structurally similar to the debitive one. The relative pronoun *ko,* inflecting only for case but not for gender and number in Baltic, is compared to the debitive prefix *jā-* under the assumption that *jā-* was a former relative pronoun of the same sort (Endzelīns 1901; Holvoet 1998). This point is somewhat problematic, not only methodologically. Let us look at this in more detail.

---

in our view: Baltic, closely related Slavic and areally related Finnic languages, to our knowledge, do not attest such a development from a finite verb form into a verbal noun/gerundive (except, perhaps, Bulgarian/Macedonian which, however, generally lack infinitives); instead a number of developments in the opposite direction may be found here. Moreover, there is no comparable construction with *jā-* in the closely related Lithuanian and Slavic in whatever function, suggesting that the gerundive itself is recent and an internal Latvian development.

All previous accounts assumed the historical derivation of *jā-* from an alleged relative pronoun **jo-* inflected according to the *o*-stem paradigm (*inter alia*, Endzelīns 1901; Stolz 1987; Holvoet 1998; Wälchli 2000). While Proto-Indo-European indeed had a phonetically similar relative pronoun **$h_2$jo-* (cf. Dunkel 2013: 312ff) which phonetically would have yielded Proto-Baltic **jo-*, the pronoun at issue is a different one, namely, the Proto-Indo-European demonstrative/third person pronoun **i/ei-* originally inflected in the *i*-stem paradigm (cf. Dunkel 2013: 363ff). Only this pronoun is found in Baltic (including Old Prussian), namely, in the cliticized forms of the definite adjectives as well as the third-person pronouns in *j-* in Lithuanian and Latgalian. It has secondarily acquired a number of case-inflection forms from the *o*-stem paradigm (an independent process that many *i*-stem nouns and pronouns underwent in Baltic) and thereby became morphologically similar to the paradigm of the alleged relative pronoun **jo-* (while the latter probably never existed in Baltic).

Crucially, neither Lithuanian nor Latvian (let alone Old Prussian) ever attested the pronoun **i/ei-* functioning as a relative pronoun. Instead, all three Baltic languages preserve the original demonstrative (grammaticalized as definite adjectival clitics) and the third-person pronoun functions (Lithuanian and Latgalian). The next possible connection is the Old Church Slavonic relative pronoun *i-že* which is (i) far too distant given that the Latvian debitive emerged very late and (ii) the historical phonetics of Slavic is such that it is virtually impossible to determine whether *iže* represents the old relative pronoun **$h_2$jo-* or has secondarily extended the use of the demonstrative **i/ei-* (cf. German *der* which is both demonstrative, third person and relative); in the second case, Slavic is not a good parallel for assuming a Baltic relative pronoun **jo-*.

Moreover, there is another, old relative pronoun well attested in Baltic with stem in *k-* that is the descendant of the Proto-Indo-European **k^uó-* which also had relative-pronoun functions (cf. Dunkel 2013: 452ff; see Hettrich 1988: 503–505 on its relative function in Hittite and Latin). This would require assuming two different relative pronouns at some point during the Proto-Baltic period – a situation not impossible and not unattested but, given the lack of independent evidence, such an assumption is a violation of the Occam's razor principle, in our view.

Moreover, the very form *jā-* is not straightforwardly analyzable as a case form that would fit a construction like (11). Holvoet (1998), following Endzelīns (1901; 1951), assumes that it stems from the genitive (of negation) in negated contexts. He shows indeed that the negative context has played an important role in the emergence of the construction. Dialectally there is another form of the debitive marker alongside the standard *jā-* namely *ju-*, which may be taken as an accusative singular and/or genitive plural of the alleged relative pronoun **jo-* (Holvoet 1998; Vanags 2000; Wälchli 2000).

However, given the lack of evidence for the very existence of this relative pronoun in Baltic, one is tempted to provide an alternative explanation. We suggest that the prefix *jā-* may rather stem from a conjunction **jā* that is well attested in Lithuanian *jo* (< historically stemming from **jā*) or, more frequently, extended by the particle *gi* as in *jog*[10] with the meaning '(so) that' (cf. LKŽ, *vide sub verbo*):

(12) (Lithuanian, K. Lenartaitė, p.c.)
*pasakė tai tyliai, **jog** visai nieko ne-girdėti*
say.PST.3 so softly **that** at.all nothing.GEN.SG NEG-hear.INF
'He said (this) so softly **that** no one could hear it.'

The prefix *ju-* (shortened form *juo/jū*) might represent just a dialectal variant of *jā-*: The pair *jā-* and *ju-* precisely correspond to the pair Middle and Standard Latvian *ja* 'if' vs. East Latvian *ju* (or *jū*) 'if' (ME II, 115 and 124–125). Analogously, the marker *ju-* may be considered to be the High Latvian phonetic variant of the original *jā-*. Indeed, the *ju-* forms are found almost exclusively in Selonian subdialects of the High Latvian dialect (Vanags 2000: 144–145).[11]

The debitive gerundive would then simply be a *reinforced infinitive* discussed at length in Haspelmath (1989: 304ff): infinitives tend to acquire additional markers in order to reinforce their purposive meaning but then, subsequently, generalize them to other readings, cf. the "double-marking" of infinitives in German with the clitic *zu* and the suffix *-(e)n*. The comparison with the German originally possessive obligation construction is telling:

(13) (German, constructed)
Ich habe zu esse-n
1SG.NOM have.PRS.1SG INF eat-INF
'I have to eat / I have something to eat.'

Structurally parallel to the Latvian debitive gerundive alias reinforced infinitive are, for example, the infinitives in Koine Greek which are often extended by *hṓste* 'so that, in order to' or *hína* 'in order to' (the latter made a notable career in Modern Greek) (Haspelmath 1989). In our view, the account assuming the reinforced infinitive is more coherent with what is actually attested in Baltic than the one assuming a relative pronoun in **jo-*.

Wälchli (2000: 206–207) suggests that the original subordinate clause may have grown together into one word unit (gerundive) independently of the grammaticalization of the debitive, yielding a "verbal noun" (i.e. gerundive in our terms). This

10. *gi* is a particle widely used in Lithuanian.

11. Alternatively, the prefix *ju-* (shortened form *juo/jū*) may also stem from a conjunction *juo*. The latter has the meaning 'because' in modern Latvian and 'as, than' in modern Lithuanian.

is supported by the rare dialectal examples (listed in Endzelīns 1951: 973; Vanags 2000: 144–145) with the matrix verbs other than the possessive/existential *be* taking the subordinate debitive gerundive, cf.:

(14) (Latvian, dialectal; Endzelīns 1951: 973)
Ne-devu *nevienam* *jā-brauc* *savas*
NEG-give.PST.1SG none.DAT.SG DEB-ride RPO.ACC.PL.F
kamaniņas
sledge.ACC.PL
'I didn't give my sledge to anyone to ride/for a ride.'

These examples all attest the purposive meaning of the debitive gerundive. This might be taken as another piece of evidence for the origin of the debitive gerundive from a reinforced infinitive, because infinitives tend to be first reinforced when precisely used in their purposive meaning (cf. Haspelmath 1989).

Another advantage of the reinforced-infinitive explanation is that it creates the link to the old, modal-infinitive construction as in (15), which except for the lack of the debitive marker would be structurally very similar to the source of the debitive construction: it employed exactly the same auxiliary *būt* 'to be', it has the same case frame DAT-NOM and it also employs the infinitive of the lexical verb (cf. Endzelīns 1951: 992, 994), cf. the second clause in the following Lithuanian example:

(15) (Lithuanian, CCLL, courtesy of K. Lenartaitė)
Ji *nutilo,*
3.NOM.SG.F fall.silent.PST.3
ir ***man*** ***buvo*** ***matyti*** ***pusė***
and 1SG.DAT be.PST.3 see.INF half.NOM.SG
jos *veido*
3.GEN.SG.F face.GEN.SG
'She fell silent and I could see half of her face.'

Given that the meaning of this old construction was very similar to the grammaticalized debitive, it must have had some interrelation with it. Thus, Stender in his *Lettische Grammatik* from 1783 treats both constructions as two subtypes of one and the same category *Modus Necessitatis* (*apud* Andronovs 1998: 171–172). It is assumedly the synonymy of both constructions (from the second half of the 17 c. on) that made the old construction redundant, which in turn led to its full replacement by the debitive construction in Latvian (cf. Vanags 2000: 149, 152). This did not happen in Lithuanian precisely because there was no new synonymous construction in Lithuanian.

In contrast to the traditional explanation, our account places the debitive construction among a number of analogical developments in other languages, while the account employing a relative clause has, to our knowledge, no parallels among the grammaticalized obligation construction cross-linguistically. Thus, the emergence of the debitive would be analogical to, for example, English constructions as below, cf. the existential (16) and the possessive (17):

(16) *She is to see the dean tomorrow at 4.* (Bybee et al. 1994: 184)

(17) *What a student has to pay for housing these days is outrageous.* (Bybee et al. 1994: 184)

Finally, our account is not fully incompatible with the traditional account assuming the relative-pronoun function of *jā-*/*ju-*. Thus, if the traditional account assumes the existence of the relative pronoun *$*(h_2)jo$- in Proto-Baltic, then, the conjunction *$jā$ (Lithuanian *jo*) or *juo* served to reinforce the infinitive may be considered to be originally a case form of this relative pronoun sustained in the conjunction function, as one of the reviewers suggested.

4. ACC replacing NOM

Above we have discussed the reconstructed source construction of the debitive. Now we can trace back the changes that have occurred in the case-marking of the logical object.

4.1 Reconstructed state

The original case-marking of the object-like argument was the nominative, because syntactically it was the subject of the first (possessive/existential) clause of what was originally a bi-clausal structure (see Section 3 above). The nominative case must be reconstructed for all persons (*pace* Holvoet 1998) including the first and second person pronouns, as it is attested in isolated and clearly conservative examples from Old Latvian and folklore texts, cf. the following example from Old Latvian with the second person *pro*-dropped nominative pronoun:

(18) (Old Latvian, SENIE, VLH1685_Syr, 55A. lpp., 28.)
kà ja-kaŗŗo eśśi
that DEB-hassle be.PRS.2SG
'…that you are to hassle'

This example is not entirely unequivocal, as the author does commit mistakes here, cf. *tawu Prezzu* 'your.ACC.SG good.ACC.SG' instead of the correct *tawu Prezzi* just a few lines below. However, the nominative case of SAP pronouns is attested also elsewhere and therefore must be assumed to have existed for a period of time. Moreover, the nominative case of SAP pronouns was also attested in dialects:

(19) (Middle Latvian dialect, Valmiera subdialect, Endzelīns 1951: 972)
tâ pat tu pats ar bij jà-glabā
so even 2SG.NOM self.NOM.SG.M also be.PST.2SG DEB-save
'You, yourself, also had to be rescued.'

From a fairy tale:

(20) (Folklore Latvian, Šmits 1925/1937 *apud* Endzelīns 1951: 972)
tad tu man arī būsi jā-kuopj
then 2SG.NOM 1SG.DAT also be.FUT.2SG DEB-care
'Then, I will also have to take care of you.'

Holvoet (1998) is skeptical because the examples with the nominative first and second person pronouns are rare. However, if we assumed that the accusative marking with these pronouns was original, then the alleged secondary introduction of the nominative marking would run against the strong general tendency to introduce accusative for all other NP types (see Sections 4.3 and 4.4 below). Given that this NP was originally the subject of the possessive/existential verb 'to be', there is no way to explain the nominative forms as secondary and analogical. However, it is possible, as Holvoet (1998) points out, that both options co-existed from the very beginning, depending on whether the pronoun was regarded as subject of the originally possessive clause or as the object of the dependent debitive gerundive. While we cannot rule out the latter option completely, it seems less probable for the following reasons. First, the second clause contained an amalgamated structure, namely, the complementizer/conjunction (or relativizer) **jā* fused with the infinitive of the lexical verb, and, hence, was unlikely to accommodate constituents inside it. Secondly, the strong preference of 92% for the word order with the object-like NP preceding the debitive predicate in Old Latvian (cf. Table 3 above) equally suggests that the "object-like" NP was rather part of the first clause and not of the second one and was the subject of the matrix clause of that time.

4.2 Old Latvian

Before we turn to the state of affairs in Early Modern Latvian and Contemporary Latvian, we briefly examine the case assignment patterns found in Old Latvian (16th–19th centuries). The distribution between the NOM vs. ACC forms in the respective sample is as follows:

Table 6. The case-marking of the object-like argument in Old Latvian

	(i)		(ii)	(iii)
NOM	ACC=NOM	**ACC**	GEN	no object found
81	21	**5**	8	84

The proportion important for our context excludes equivocal counts such as (i) feminine plural forms which do not distinguish NOM vs. ACC, (ii) GEN forms (genitive under negation and partitive genitive) which can be equally based on "underlying" nominatives or accusatives, as well as (iii) occurrences of intransitive verbs and object ellipses. Hence, the relevant proportion is as follows:

Table 7. The proportion of NOM vs. ACC marking in Old Latvian

NOM	**ACC**	**Total**
95% (81)	5% (4+1)	100% (85+1)

Moreover, the five examples with accusative may be reduced to even fewer accusatives. First of all, the following example appears in the SENIE corpus twice with the same wording and stems from the translations of the Swedish army statutes *Krigz-Artiklar…* from 1696 (the original appeared in Stockholm in 1683), representing thus just one example:

(21) (Old Latvian, SENIE, SKL1696_RA, 0., SKL1696_KB, 0.)
… pahr tahm Leetahm/
about DEM.DAT.PL.F matter.DAT.PL
ko *teem ślikteem Saldahteem jadarr/*
REL.ACC DEM.DAT.PL.M common.DAT.PL.M soldier.DAT.PL DEB.do
un paklaußigi japanahk …
and obediently DEB.achieve
'… about those matters **which** the common soldiers have to do and obediently have to carry out…'

After having excluded this repetition, we get four examples of ACC. The following example is questionable:

(22) (Old Latvian, SENIE, Manc1654_LP1, 196. lpp., 27.)
Kad tawa Stunndinja daggajußśi gir
when your.NOM.SG.F hour.NOM.SG arrive.PPA.NOM.SG.F AUX.PRS.3
*/ und töw jaeet **tawu Zeļļu***
and 2SG.DAT DEB.go **your.ACC.SG.M way.ACC.SG**
'When your hour has come, and you have to go **your way**.'

The verb 'to go' is a genuinely intransitive verb in Latvian. The unexpected accusative marking of 'your way' may be explained if one assumes that this is an adjunct that has not been fully integrated into the case frame of the verb, and is therefore not "promoted" as a regular object into the nominative. We have no direct proof for this assumption. However, the extremely low frequency of the ACC in the debitive as well as the predominantly intransitive use of this verb in the debitive construction (and elsewhere) make this assumption plausible: out of 23 instances in the SENIE corpus there are only 2 instances of a transitive-like use including the example above.

The following example is more reliable, although the effect of priming cannot be entirely excluded here due to the accusative form of the 'vinegar' in the protasis:

(23) (Old Latvian, SENIE, Hag1790_IM, p.5, 29.)
Kad Wihna= Ettiki ne warr dabbuht,
when wine.GEN.SG vinegar.ACC.SG NEG can.PRS.3 get.INF
tad ja-ņemm ßtipru Allus= Ettiki.
then DEB-take **strong.ACC.SG** beer.GEN.SG **vinegar.ACC.SG**
'When wine vinegar cannot be got, then one should take **strong** beer **vinegar**.'

Finally, the following example contains the accusative form of the quantifier *wiśśu*:

(24) (Old Latvian)
Skohlasbehrneem wiśśu jadarra, ko
school.child.DAT.PL **all.ACC.SG** DEB.do what.ACC
Skohlmeißters pawehl, wiņņa Mahziśchanu
schoolmaster.NOM.SG order.PRS.3 3.GEN.SG.M teaching.ACC.SG
labbi usklaußidami.
good.ADV hear.CNV
'The schoolchildren have to do **everything** that the teacher requests and attend to his teaching properly.' (SENIE, SL1789, p.12, 6.)

Here, the effect of priming on the relative pronoun *ko* in the second clause cannot be entirely excluded. Moreover, the converb *usklaußidami* is a *same-subject-converb,* which means that it should be coreferential with a nominative subject of the preceding clause. Yet, it is coreferential with the dative subject-like argument of the debitive construction *Skohlasbehrneem*. It is unclear whether it represents a

reinterpretation of the latter into a full-fledged syntactic subject despite its dative case-marking or whether it is a mistake caused by the non-native writer. In the latter case, the accusative marking may also be cast into doubt. To conclude, while the accusative marking of the logical object in the debitive construction cannot entirely be ruled out already for the Old Latvian period, its frequency is significantly low (4–5%). Given that there were hardly any prescriptive rules as to the avoidance of the accusative marking in Old Latvian (as they are found today), it is safe to assume that the accusative marking – if present at all – was generally extremely rare in Old Latvian.

4.3 Early Modern Latvian (1850–1900)

While Old Latvian attests around 4% ACC (excluding the repetitions), the Early Modern Latvian Sample attests 7 ACC forms vs. 388 NOM forms (in addition to 4 examples with morphologically ambiguous NOM=ACC forms), which is around 1.8%. The figure is unexpectedly lower than the 4% found in the Old Latvian Sample. However, this deviation seems to be rather due to the very low percentage of animates in the Early Latvian Sample: 13 animate vs. 386 inanimate referents, i.e. 3% in total. This animacy ratio might be responsible for such a low rate of ACC; as will be shown below, animacy is an important predictor of ACC forms in Contemporary Latvian (cf. 4.5.3 below). Notably, all the seven ACC hits are inanimates.

4.4 Contemporary Latvian

There are two major shifts at this stage: (i) the ACC marking becomes obligatory with the first and second person pronouns and (ii) the ACC marking gains more ground with all other NP types becoming more frequent in average (see Table 9 below).

As to (i), Contemporary Latvian requires the SAP pronouns to be marked with ACC in the debitive construction while the NOM marking is ungrammatical here. In turn, all other NP types must bear NOM in this position in the standard language.[12] Unfortunately, our samples for Old Latvian and Early Modern Latvian do not allow us to make claims as to when exactly the ACC of SAP pronouns appeared and became – differently from all other NP types – mandatory, NOM

12. One exception is the reflexive pronoun *sev-*. However, here the explanation seems rather to be purely morphological: this pronoun simply does not have a nominative form at all, hence, the accusative form has to be used here.

becoming ungrammatical. Examples with the NOM marking such as (18) and (19) above suggest that, at least, the generalization of the ACC marking with SAP pronouns must be quite recent.

(ii) While SAP pronouns are obligatorily marked ACC, in what follows we focus on how the ACC marking expands onto other NP types in Contemporary Latvian. Note first that Standard Latvian does not allow ACC marking here, adhering to the more conservative pattern:

(25) (Standard Latvian)
Viņam ir jā-redz sav-s draug-s
3.DAT.SG.M AUX.PRS.3 DEB-see RPO-NOM.SG.M friend-NOM.SG /
/ **sav-u* *draug-u*
*RPO-ACC.SG *friend-ACC.SG
'He has to see his friend.' [Elicited]

(26) (Standard Latvian)
*Viņ-am ir jā-redz *es* / *mani*
3-DAT.SG.M AUX.PRS.3 DEB-see *1SG.NOM / 1SG.ACC
'He has to see me.' [Elicited]

However, even in the conservative (and normalized) Standard Latvian, there is a strong preference to generalize the ACC marking of the object that is semantically licensed by the lower verb in the constructions with some matrix verb in the debitive (see already Fraenkel 1928: 49; Endzelin 1901: 72):

(27) (Standard Latvian)
Jā-cenšas atjaunot vēsturnieku komisiju.
DEB-try renew.INF historian.GEN.PL committee.**ACC**.SG
'Attempts should be made to renew the historians' committee.'[13]

The conative verb *censties* 'to try' (in the debitive here) takes the infinitival-clause complement, and it is the lower verb that assigns ACC to its object. What this phenomenon shows for our study is that the debitive's case frame ceases to be operative here in overriding the ACC case-marking of the embedded complex predicate 'to try to renew'. One may argue that this does not tell us anything about the incipient loss of the debitive case frame but is just the effect of the syntactic structure that is opaque for the debitive to superimpose its case frame on the lower verb's arguments. The following examples show, however, that this is not so, and occasionally exceptions may be found in which the debitive predicate does assign the NOM marking to the lower object (note that the NOM is less regular in these examples and can always be replaced by ACC, cf. Fennell & Gelson 1980: 430):

13. http://nra.lv/viedokli/maris-krautmanis-3/141427-jacensas-atjaunot-vesturnieku-komisiju.htm

(28) (Contemporary Latvian)
Vienmēr jā-cenšas uztvert teikuma jēga
always DEB-try perceive.INF sentence.GEN meaning.NOM.SG
'One always has to try to understand the meaning of the sentence'[14]

We consider the nominative case here as a residual of the DAT-NOM government of the debitive that is gradually ceasing to apply on objects of dependent infinitives.

The following example is analogous to (28): although the underlying matrix verb *sākt* 'to start' does not provide for semantic roles which, in turn, are assigned by the lower verb, the debitive form of it does provide a case frame for the NPs semantically belonging to the lower verb:

(29) (Contemporary Latvian)
Tad jau laikam jā-sāk atbalstīt arī
then PTC perhaps DEB-begin support.INF also
šķība dziedāšana, greiza dejošana utt.
wry.NOM.SG.F singing.NOM.SG skew.NOM.SG.F dancing.NOM.SG etc.
'Then one should start supporting also wry singing, skew dancing, etc.'

The difference between (28) and (29) is that, in (28), the verb *censties* 'to try' provides a semantic role only for the first NP, in (29), the verb *sākt* does not provide semantic roles for both NPs. Both arguments of the lower verb are assigned cases according to the case frame of the higher, debitive predicate. Notably, the linear position of the nominative NP is not in any way restricted with this case assignment, and all possible orderings obtain, cf. Table 8 and Example (30) with NOM in the sentence-initial position:

(30) (Standard/Contemporary Latvian)
Gads jā-cenšas pabeigt pozitīvi
year.NOM.SG DEB-try complete.INF positive.ADV
'The year should be ended up positively.'

We have collected 33 examples with the NOM marking of the object argument stemming from the lower verb on the Internet in order to see which word orders are possible here. While Table 8 is not meant to be statistically representative, the counts show that all possible word orders may obtain:

Table 8. Attested linear positions of the NOM NP with control and raising verbs

NOM-DEB-INF	DEB-NOM–INF	DEB-INF-NOM
5 (15%)	10 (30%)	18 (55%)

14. http://www.rigastulki.lv/lv/rigas-tulki-tehnisku-un-sarezgitu-tekstu-tulkosana/rekomendacijas-tulkotajiem-1-dala/

We analyze examples like (28) and (29) as residuals on the grammaticalization cline which retain the (original) debitive case frame, while the standard ACC marking as in (27) is one of the first steps towards the loss of the debitive's own case frame.

The background for this is the following. Typically for Baltic languages, and for Latvian, in particular, the modal predicates and the respective constructions have their own case frames that are often superimposed upon the case frame of the lexical verb, which is historically the residual of their former lexical nature. In the same way, the debitive operator coded by the auxiliary and the marker *jā-* has the DAT-NOM case frame of its own. In the course of grammaticalization modals lose their lexical properties, and the loss of the case frame is symptomatic of this (see also 6.1).

While the ACC marking is standard with lower objects of complex predicates and obligatory with SAP pronouns, let us now consider how ACC marking expands onto the third person NP types with simple verbs. As our analysis will reveal, not all NP types are affected to the same extent. However, first we compare the samples as to the average frequency of the ACC vs. NOM ratio:

Table 9. ACC vs. NOM ratio across all samples

	NOM	ACC	*p*-value compared to Contemporary Latvian
Old Latvian	95% (81)	5% (4)	significant ($p < 0.05$)
Early Modern Latvian	98.2% (388)	1.8% (7)	significant ($p < 0.05$)
Contemporary Latvian	87% (2778)	13% (412)	

There is a significant increase of ACC in Contemporary Latvian as compared to Old and Early Modern Latvian. As has been argued above in 4.3, the extremely low ratio of 1.8% of ACC forms in Early Modern Latvian is rather accidental and due to the fact that among the total of 395 examples there were only 3% of animate NPs. Our contention is that if the animates vs. inanimates proportion had been more balanced, the Early Modern Latvian ratio would correspond to the one in Old Latvian. Crucially, we observe here a considerable change in frequency: from 5% in Old Latvian to 13% in Contemporary Latvian. Note that the figure for Contemporary Latvian would have been much higher if there had not been such strong prescriptive rules in Latvia while this cannot be said about Old and Early Modern Latvian, which were not subject to prescriptivism to the same extent.

In what follows we address the conditioning factors for ACC selection in a multifactorial analysis of the Contemporary Latvian Sample in order to find out how exactly the ACC marking spreads across different categories. As has been mentioned above, our model glosses over sociolinguistic type factors on the assumption that they have weaker impact than the factors to be presented below. Our

resulting multifactorial model (schematized in (61), Section 6.3) includes a number of factors/variables responsible for ACC/NOM selection organized hierarchically in accordance to their effect strength.

4.4.1 *Primary-importance factors*

The ACC has spread onto all NP types, and there is no NP that would disallow ACC marking in the contemporary language. However, we argue that this spread does not affect all possible types and occurrences of the object NP to the same extent in terms of statistically significant tendencies. In subsections 4.4.2 to 4.4.7, we analyze the statistical dispositions of the predictors for the ACC marking from the strongest to the weakest ones. Yet, while 13% is the average preference for selecting the ACC marking,[15] in what follows, we aim at establishing factors that clearly favor or disfavor ACC marking, being statistically significantly above or below this value. Our investigation comprises the NP-related factors (4.4.2–4.4.6) and predicate-related factors (4.4.7).

The 13% of ACC vs. NOM distribution shows a significant overall preference for NOM, independently of these factors. Responsible for this ratio is, in our view, the combination of two factors (unfortunately we cannot disentangle them): (i) the very expansion of ACC is quite incipient in the language and (ii) strong influence of prescriptive rules. It is obvious that such a high ratio of NOM (87%) cannot be accounted for simply in terms of purification rules and standardization norms albeit, confessedly, the latter are very strong in Latvia. The reason is that this assumption would imply that the speakers are completely re-learning this part of grammar in the schools and universities. More likely we should assume that both factors are at work here: (i) there is a gradual expansion of ACC which is (ii) impeded by the prescriptive rules. In what follows, we will not dwell on these two factors but will imply that they are the strongest ones constraining the ACC/NOM variation. Below we refer to these factors as *the primary-importance factors*. In what follows we discuss the *secondary-importance factors*.

4.4.2 *Linear position of the object NP relative to the debitive predicative*

The most important secondary predictor for case selection we found is the linear position of the object NP relative to the debitive predicative, with two values: after vs. before the debitive gerundive (DEB):

---

15. Note that this average is highly biased for full NPs since the prevailing majority of all NP types in our sample is full NPs which, in turn, are less prone to ACC than, for example, different kinds of pronouns, see the detailed exposition in 4.4.3 below. The average excluding the overweighting of the too frequent full NPs is 23% (which is the average percentage of different NP types in Table 12 below).

Table 10. The interrelation of position relative to the debitive predicative and case (p-value < 0.001, highly significant)

	DEB-NP		NP-DEB	
ACC	150	ACC in the After-position	262	ACC in the Before-position
NOM	1633	**8%**	1145	**19%**

As can be observed from the table, there is a considerable effect size; there are both less NOM and more ACC (than the overall average) in the preverbal position, while there are more NOM and less ACC in the postverbal position. As will be shown below, *wh*-pronouns (relative pronouns and question/focus pronouns) score second highest with respect to ACC selection as compared to other NP types. Even though the word order in Latvian is largely syntactically unconstrained, *wh*-pronouns precede the predicate. In order to establish the *before*-position as the crucial factor for ACC among the secondary-importance factors, *wh*-pronouns should therefore be excluded from the counts in Table 10, since they themselves represent a factor independently favoring ACC but, at the same time, occur in the *before*-position only. The improved figures are shown in Table 11 below:

Table 11. The interrelation of position relative to the debitive predicative and case *without wh-pronouns* (p-value < 0.001, highly significant)

	DEB-NP		NP-DEB	
ACC	151	ACC/NOM proportion	207	ACC/NOM proportion
NOM	1633	in the After-position	983	in the Before-position
		8%		**17.3%**

Indeed, even after *wh*-pronouns are removed from the figures the *NP-DEB*-position remains to be a strong predictor for ACC.

The reason for selecting the ACC marking in the position preceding the predicate is the following. The original NOM marking is ambiguous as regards the grammatical roles and, even more importantly, as regards the interpretation of the macroroles A vs. P: in the vast majority of Latvian sentences outside the debitive, the NOM marking is tendentially associated with the A role (exceptions being passives and some modal constructions). Similarly, the preverbal position is also tendentially associated with A (and S) rather than with P role in Latvian. Yet, having a NOM-marked NP in the preverbal position yields the wrong expectation of this NP's coding the A-participant. Therefore, for an NP used preverbally – in order to avoid this (frequency-based) expectation – the ACC marking is selected in order to signal the unexpected P-macrorole of the NP.

4.4.3 *Different input types*

In what follows we explore whether different input types create significant preferences in case selection. The following table orders different NP types into a scale according to their preferences for ACC:

Table 12. The interrelation of the input type and case

	ACC	NOM	*p*-value if compared to the preceding NP type	*p*-value if compared to the total average
Reciprocal pronouns	**70%** (7)	3		***significant***[2]
wh-pronouns	**25%** (54)	164	***significant***[2]	***significant***
Demonstrative pronouns	**21%** (64)	246	*not significant / chance*	***significant***
Viss 'all'	**19%** (12)	51		*not significant / chance*
3rd person pronoun	**19%** (14)	59		*not significant / chance*
Indefinite pronouns	**11%** (16)	125		*not significant / chance*
Full NP (total)	**10%** (232)	1998		***significant***
Proper names	**9%** (13)	134		*not significant / chance*
Total	**13%** (412)	2780		

Notably, *proper names* typically score high on various scales (such as the definiteness or animacy scale) and acquire new marking strategies in terms of differential object marking (DOM) earlier than many other NP types. Thus, in Old Spanish, proper names are consistently marked (with the direct-object DOM preposition *a*) while other nouns (even the definite animate ones) are only optionally marked (Laca 2006: 242; von Heusinger, to appear). The Latvian debitive – while otherwise *mutatis mutandis* following the scales – creates this "gap": not only do proper names follow full NPs, but what is more, they fall below the average of 13%, which must be interpreted as indicating that proper names is a negative predictor for ACC.

It is the reciprocal pronouns that score highest by preferring ACC over NOM in absolute values: 7 vs. 3. Note, however, that the total figures for the reciprocal object are low and, therefore, not fully reliable. However, the DAT-NOM case frame of the debitive seems indeed to pose a problem for speakers. Reciprocals (*cits citu* 'other.NOM.SG.M other.ACC.SG.M' and *viens otru* 'one.NOM.SG.M second.ACC.SG.M/F') mirror both arguments of the verb and the respective case-markings. Thus, for example for the direct object of a transitive verb, the NOM-ACC case frame *viens otru* 'one.**NOM**.SG.M second.**ACC**.SG.M/F' is used. Accordingly, one would expect to find DAT-NOM with the debitive. Indeed, there are two examples in our sample with the DAT-NOM case marking of the reciprocal compound:

(31) (Contemporary Latvian)
Komunikācijas procesā pusēm jābūt godīgām, objektīvām …
'Opponents have to be fair and objective in the process of communication'
nevis ***vienai*** ***otra*** *jā-nosoda*
and.not **RECIPR.DAT.SG.F** **RECIPR.NOM.SG.F** DEB-condemn
un *jā-apvaino.*
and DEB-offend
'and they should not condemn and offend each (DAT) other (NOM).'

Differently from other NP types, reciprocals (like reflexives) can never occur in the subject position in Latvian. This renders the NOM marking of the second part of the reciprocals – which is generally associated and agrees with the object (cf. the discussion in Seržant, this volume) – somewhat unusual. The fact that the speakers do have difficulties in rendering the Standard Latvian DAT-NOM can be observed from the following example where the reciprocal pronouns render NOM-NOM:

(32) (Contemporary Latvian)
meitenes, *vai* *tiešām* ***viena*** ***otra***
girl.NOM.PL Q PTC **RECIPR.NOM.SG.F** **RECIPR.NOM.SG.F**
tagad *ja-apvaino!!*
now DEB-offend
'girls, is it now really necessary to offend each other!!'[16]

On the other hand, the DAT-ACC pattern is semantically straightforward in that the part that is semantically associated with the direct object of the verb elsewhere receives the expected case, that is, ACC. Indeed, this case-marking of the reciprocals is the most frequent here:

(33) (Colloquial Latvian)
Tāpēc *nav* *jā-kaujas*
therefore NEG.be.PRS.3 DEB-struggle
un ***vienam*** ***otru*** *jā-apsaukā*
and **RECIPR.DAT.SG.M** **RECIPR.ACC.SG** DEB-call_offensively
'Therefore [you] should not fight and should not call each other names.'[17]

We summarize the results in the following scale; the NP types on the left are stronger affected by the ACC expansion than those on the right to the arrow:

(34) *Effect scale of expansion of the ACC marking*
Reciprocal pronouns < *wh*-pronouns < demonstratives < *viss* 'all' & third person pronoun < indefinite pronouns < full NPs < proper names

16. http://twilight.town.lv/2010/04/16/turpinam-lasit-fanu-stastus/

17. http://staburags.diena.lv/citas-zinas/viedokli/mierigs-16-marts-11737

Since proper names fall considerably below animate NPs and all the other NP types to their left, the scale in (34) runs counter to a number of hierarchies, such as the Extended Animacy Hierarchy in Croft (2003: 130). Other hierarchies, such as the definiteness hierarchy, do not explain the differences in (34) appropriately, since, for example, proper names and third-person pronouns are not distinct with regard to their degree of definiteness. At the same time, the Effect scale in (34) largely correlates with the *Accessibility Marking Scale* in (35) as suggested in Ariel (1990: 73; 2001: 31), compare (34) and (35):

(35) *Accessibility Marking Scale* (Ariel 1990: 73)
Full name + modifier > Full ('namy') name > Long definite description > Short definite description > Last name > First name > Distal demonstrative + modifier > Proximal demonstrative + modifier > Distal demonstrative (+ NP) > Proximal demonstrative (+ NP) > Stressed pronoun + gesture > Stressed pronoun > Unstressed pronoun > Cliticized pronoun > Extremely High Accessibility Markers (gaps, including pro, PRO and *wh* traces, reflexives, and Agreement)

Proper names provide for the most exact reference (on the assumption that proper names have unique reference) and allow the interlocutors to retrieve the least accessible referents. Ariel (1990: 73) distinguishes between full names and last/first names in (35); the latter are used with more accessible referents and are therefore less informative. In our sample we did not draw this distinction, and we encompass both types under *proper names* because there does not seem to be any significant difference between these two subtypes as regards ACC vs. NOM marking. This is the only respect in which the Effect scale deviates from the Accessibility Marking Scale. Thus, full NPs are less strict reference retrievals than full names and therefore can designate only somewhat more accessible referents. Finally, the pronouns are highly dependent on some antecedent or some additional knowledge and can refer to even more accessible referents only. What is more, even the four different types of pronouns in (34) correlate with different degrees of accessibility: the reciprocal pronouns refer to the most accessible referents, i.e. those referents that just have been or will be (re-)activated in the same clause, while *wh*-pronouns in their relative use refer to entities that just have been or will be activated in an immediately neighboring clause and are, hence, also highly accessible, but not as highly as the referents of reciprocals. Finally, demonstratives may provide reference to somewhat less accessible and more distal referents than these pronouns but, in contrast to full NPs and proper names, they nevertheless need some additional knowledge for their correct reference retrieval. Correlation between (34) and (35) suggests that the more accessible the referent the stronger it attracts ACC marking, while the NOM marking is associated with low accessibility.

Once we accept that the Effect scale in (34) is constrained by the differences in degree of accessibility, the obligatory ACC marking of the first and second person pronouns (and potentially of the reflexive pronoun) discussed above (4.4) reveals itself to be fully expected: these pronouns are inherently among the most activated and, hence, accessible referents. Therefore, they attract the ACC most strongly and were, assumedly, historically the very first ones among all NP types to allow ACC marking. Consequently, by the time of Contemporary Latvian, the ACC marking is already generalized here as the only marking option. Now, the Effect scale in (34) can be established more precisely taking the SAP pronouns into account:

(36) *Effect scale of expansion of the* ACC *marking* (refined)
SAP pronouns (& reflexive pronoun) < Reciprocal pronouns < *wh*-pronouns < demonstratives < *viss* 'all' & third person pronoun < indefinite pronouns < full NPs < proper names

The Effect scale in (36) now reads as follows: from the obligatory ACC marking on the very left, to the more preferable ACC marking with reciprocals, further to the right the other NP types with decreasing frequency of, and association with, the ACC marking, up to the proper names which have the least preference for ACC. The Effect scale (36) correlates with the accessibility scale of Ariel (1990) and is not an instance of the animacy scale as traditionally assumed (cf. Timberlake 1974), for the simple reason that the opposition between, say, SAP pronouns and third-person pronouns is not primarily driven by animacy. Animacy, in turn, does play a role as another factor, which is, however, orthogonal to accessibility.

4.4.4 *Animacy of the object referent*

Animacy is another factor that constrains the preferences for ACC/NOM marking. Consider Table 13:

Table 13. Animacy effect with full NPs

	ACC	NOM	*p*-value for the animate vs. inanimate reference	*p*-value if compared to the total
full NP (animate)	**13%** (89)	573	**significant**	**significant**
full NP (inanimate)	**9%** (143)	1425		not significant / chance
Total	**10%** (232)	1998		

As can be observed from Table 13, the distance between the effect size of animates and the effect size of inanimate NPs is not large – inanimate NPs come slightly below the overall average of 13%. Interestingly, while it is the value [+animate] that additionally increases the inclination towards ACC with pronouns (except for the third-person pronoun, see immediately below), with full NPs, animacy achieves

just the average of 13% which is constrained by the primary-importance factors only (as in 4.4.1). In turn, it is the lack of [+animate] feature that additionally stimulates full NPs to be marked with NOM.

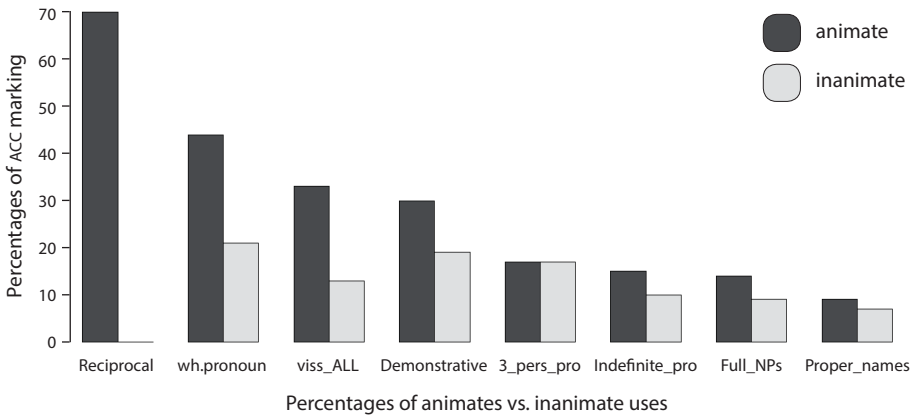
The following table integrates both factors: the degree of accessibility and (in)animacy and illustrates the respective distances by means of odds ratio. As can be observed, animacy is an important predictor (among the secondary-importance factors) because animate subsets of the respective NP types always score clearly higher with ACC than their inanimate counterparts:

Table 14. Differences in the size effect depending on animacy and accessibility degree

	ACC	NOM	**Odds ratio measuring the difference from the overall average (the latter is 0.15)**
Reciprocal pronouns (animates only)	**70%** (7)	(3)	16
wh-pronouns (animates)	**44%** (17)	(22)	5
Viss 'all' (animates)	**33%** (6)	(12)	3
Demonstrative pronouns (animates)	**30%** (14)	(32)	3
wh-pronouns (inanimates)	**21%** (37)	(142)	1.8
Demonstrative pronouns (inanimates)	**19%** (50)	(213)	1.6
3rd person pronoun (animates)	**17%** (13)	(54)	1.6
3rd person pronoun (inanimates)	**16%** (1)	(5)	1.3
Indefinite pronouns (animates)	**15%** (7)	(41)	1.1
Viss 'all' (inanimates)	**13%** (6)	(39)	1
Full NP (animate)	**13%** (89)	(573)	1
Indefinite pronouns (inanimates)	**10%** (9)	(84)	0.7
Full NP (inanimate)	**9%** (143)	1425	0.7
Proper names (animates)	**9%** (11)	(109)	0.7
Proper names (inanimates)	**7%** (2)	(25)	0.5

There are two conclusions to draw from these data. First, the *prima facie* average of 13% of accusatives discussed above (4.4.1) is not particularly telling because the percentage range varies remarkably depending on just the two factors at issue: from 7% with inanimate proper names up to 70% with reciprocals. In turn, the 13% results from the fact that precisely those NPs that are most frequent (full NPs) are – expectedly (cf. Bybee & Hopper 2001: 17–18; Bybee 2001: 12) – conservative as to the new, ACC marking, but by virtue of their high frequency (also in our sample) have too large a weight in the overall average. Then, the largest distance is found among the most accessible and, at the same time animate, definite pronouns, as can be observed from Table 14. In turn, the third person pronouns (both animate and inanimate), indefinite pronouns, all kinds of full NPs and proper names do not differ from each other to the same extent. Here, animacy is a less important factor.

Secondly, Table 14 allows us to draw the conclusion that the accessibility degree of a particular NP is a stronger attractor of ACC than animacy because the rows are ordered according to accessibility in the first place. The plot below visualizes this by illustrating the percentages distribution of ACC vs. NOM with different accessibility types split according to the factor [+/– animate]:



Plot 1. Animacy distinctions in ACC/NOM selection along the Accessibility scale

Plot 1 also shows that, indeed, animacy plays an important role which can be observed with every particular NP type. Furthermore, the plot illustrates that certain NP types are indifferent to animacy, exhibiting similar behavior with animate and inanimate referents. Most obviously, the third person pronoun is one of them.

In order to corroborate the claim that animacy is among the secondary-importance factors, we have run the *t-test*. However, before doing so, we excluded categories that are not sufficiently well represented to give reliable results: *third-person pronouns* which only have 6 inanimate occurrences in total[18] and *reciprocal pronouns* with zero inanimate occurrences. The *t-test* was fed with percentage values (in bold in Table 14) for each category and not with the absolute number of occurrences, so as to exclude absolute-frequency effects. It yields the following results:

Table 15. Significance of the animacy value on the ACC selection with different NP types

With *proper names*	almost significant, p-value = 0.05593
Without *proper names*	significant, p-value < 0.05

18. In Latvian, generally the demonstrative pronoun *tas* 'this' is regularly used in the anaphoric function when referring to inanimates, while the regular third person pronoun *viņ-* is used mostly with animate referents.

4.4.5 *Definiteness of the NP*

Since Latvian does not have a grammaticalized means to code (in)definiteness (except for the (in)definite adjectives), it is not an easy task to coherently annotate the examples according to this parameter. Note that we assume that definiteness – apart from its semantic core – cross-linguistically is not a uniform semantic category, and languages do exhibit variation as to the non-core instances such as singular generics, plural generics, abstract nouns, mass nouns, etc. We proceeded here as follows: NPs containing markers that are unequivocal indicators of definiteness (definite adjectives, demonstratives, various genitive modifiers) have been annotated as definite. On the other hand, generics, mass nouns, abstract nouns, etc. have all been annotated as indefinite unless one of the markers has been used. In total we get the following proportion:

Table 16. Definite vs. indefinite NPs; (full NPs only, all NP types of the sample in brackets)

	Definite	**Indefinite**	**Percentage of definite NPs**
ACC	126 (264)	106 (148)	54% (64%)
NOM	1052 (1634)	945 (1144)	53% (59%)

As can be observed in Table 16, definiteness does not seem to be a relevant parameter.

4.4.6 *Interim: NP factors*

While definiteness is not important, the preverbal position, high accessibility degree and animacy are the most important factor values among the secondary-importance factors that favor ACC marking (4.4.2–4.4.4). In our view, these three values can be subsumed under the notion of *saliency.* Now our interim results can be formulated in the following way: as regards NP-related factors, the more salient the NP is the more it is affected by the expansion of ACC and, consequently, the earlier it started appearing with ACC.

4.4.7 *Verbal factors: Distribution of ACC among different verbs*

In what follows we split our analysis into two subsections: preferences of particular semantic classes (4.4.7.1) and preferences of particular lexical verbs (4.4.7.2).

4.4.7.1 *Semantic verb classes.* To provide a lexeme-independent picture of ACC/NOM distribution across different verbs, we have annotated verbs according to their semantic classes (only the Try-verb class consists of just one member):

Table 17. Frequencies of acc across different verb classes

Verb	ACC	NOM	%	Odds ratio, relative to the overall average
Try verb	20	12	*63%	*11
Experiencer verbs	133	598	18%	1.5
Communication verbs	44	325	12%	0.9
Agentive verbs	167	1351	11%	0.8
Meeting verb	21	184	10%	0.8
Transfer verb	27	310	8%	0.6

The verb *provēt* 'to try' scores highest which, however, is not the result of its semantic class. Instead, it is motivated by the syntactic structure *provēt* typically occurs in. In subsection 4.4.7.2 below, we discuss this in more detail.

The direct object of the Transfer verbs is typically inanimate which is – as we assume – underlyingly the reason why Transfer verbs are below the overall average of 13%. Moreover, as can be observed from Table 17, the experiencer verbs slightly set themselves apart from the other verb classes, which range from 12% to 8%. Indeed, the experiencer verbs are more prone to ACC as opposed to all other verb classes, with the difference being 7%:

Table 18. Frequencies of acc across different verb classes

	ACC	NOM	%	*p*-value
Experiencer verbs	74 (133)	271 (598)	**21%** (18%)	significant (< 0.01)
All other verbs except *provēt*	91 (259)	575 (2170)	**14%** (11%)	

To determine whether semantic verb class is a significant predictor for ACC itself or is just epiphenomenal, we have excluded the lexically determined distinctions in object NP selection: in Table 18, the figures represent animate objects for both semantic classes only (whereas the respective total is given in brackets). The outcome is that (i) the experiencer verbs represent another secondary factor for the ACC selection and that (ii) other verbs do not contribute to the case selection, being close to the overall average of 13%. Having said this, in what follows we discuss preferences of particular lexical verbs.

4.4.7.2 *Lexical verbs.* Table 19 provides an overview of how frequently the ACC marking is found among different lexical verbs.

As can be observed from the table, there is considerable variation among different lexical verbs as to how often they have ACC objects in the debitive construction. There are some additional factors that have to be excluded before interpreting

Table 19. Frequencies of acc across different lexical verbs

Verb	ACC	NOM	%	Verb	ACC	NOM	%
ņemt 'take'	15	188	**7%**	*lasīt* 'read'	15	96	**14%**
nest 'carry'	8	105	**7%**	*tēlot* 'feign'	13	79	**14%**
rakstīt 'write'	16	221	**7%**	*dzirdēt* 'hear'	10	58	**15%**
sazvanīt 'call'	7	88	**7%**	*klausīties* 'listen'	20	109	**16%**
pirkt 'buy'	12	122	**9%**	*laist* 'let'	30	162	**16%**
pucēt 'dress, clean'	6	58	**9%**	*iznīcināt* 'destroy'	30	149	**17%**
slēgt 'close'	18	184	**9%**	*mīlēt* 'love'	32	128	**20%**
stūķēt 'push'	10	104	**9%**	*redzēt* 'see'	22	90	**24%**
satikt 'meet'	21	184	**10%**	*apvainot* 'offend'	18	50	**26%**
taisīt 'make'	22	180	**11%**	*apsaukāt* 'call so. offensive names'	14	15	**48%**
piekaut 'pummel'	5	38	**12%**				
trenkāt 'push in, chase'	7	53	**12%**	*provēt* 'try'	20	12	**63%**
sacīt 'say'	24	158	**13%**	*besīt* 'annoy'	0	7	#

these data. First, the verb *besīt* 'to annoy' occurs too rarely to be able to contribute to the understanding of ACC vs. NOM distribution. Secondly, the verb *provēt* 'to try, to taste' can be and is mainly used as a conative, subject-to-object raising verb, in most of its occurrences taking an infinitival complement clause. As has been stated in 4.4 above, the debitive NOM object marking has been largely abandoned with complex predicates in favor of the ACC marking even in the conservative Standard Latvian, let alone Contemporary Latvian; hence, the strong preference of *provēt* 'to try' for ACC is due to its high frequency of being a raising verb:

(37) (Contemporary Latvian)
ka ir jā-provē vismaz atrast k(aut)ko
that AUX.PRS.3 DEB-*provēt* at_least find.INF something.ACC
'that one has at least to try to find something'[19]

This verb occurred mostly with a clausal complement, and although we generally excluded examples with complement clauses in our sample, at times it was difficult to distinguish between the lexical use of this verb with a regular object NP and its being used as a raising verb with an ellipsis of the subordinate infinitive, as in the following example:

19. http://klab.lv/community/gardumi/104540.html

(38) (Contemporary Latvian)
Pusdienlaikā Sandra mūs aizveda paēst, labi, ka viņa mums varēja ieteikt,
kas *ir* *garšīgs* *un* ***ko***
what.NOM be.PRS.3 tasty.NOM.SG.M and what.ACC
mums *obligāti* *jā-noprovē.*
1PL.DAT obligatorily DEB-try
'During lunch time, Sandra took us to a restaurant; good that she could advise us on what is tasty and what we absolutely have to try [?to eat].'[20]

Moreover, we assume that these two facets are not strictly distinguished by the speakers, which is why the more frequent pattern, namely *provēt* with the infinitival complement clause, exerts a strong influence on the rarer, lexical use with an object NP.

As regards other verbs scoring considerably above the overall average of 13%, we assume that their high scores (e.g. 48% with *apsaukāt* 'to call so. offensive names') are due to frequency-based associations. The reason why particular verbs or verb classes may favor one option in DOM more strongly than the other is what has been called in Haspelmath (2008: 191) *complementary expected association*. In particular, the overall preference of particular verbs for animate vs. inanimate direct objects creates particular complementary associations, as has been shown by von Heusinger (2008) on the basis of the Spanish DOM. Here, verbs that are more frequent with animate objects in general have a stronger association with animacy and, hence, facilitate the new marking; in turn, verbs that are less frequent with animates are accordingly less strongly associated with animacy and are, hence, less prone to favor the new marking (von Heusinger 2008). The complementary expected association accounts for divergent ACC preferences of different verbs, very much in parallel to the Spanish situation in von Heusinger (2008), since animacy is an important predictor for ACC with the Latvian debitive as well. Here, the experiencer verbs generally more frequently take animate objects, with 47% in our sample (animate 345 hits vs. inanimate 386 hits), than other verbs do, with 27% of all occurrences in our sample (666 animates vs. 1763 inanimates). This difference (amounting to an odds ratio of 2.35) accordingly results in different preferences for ACC marking, as illustrated above in Table 18. We assume – drawing on Haspelmath (2008) and von Heusinger (2008) – that each verb's differing association degrees largely constrain its preferences for ACC or NOM:

20. http://vinetap.wordpress.lv/2012/10/13/katra-diena-ka-piedzivojums/

(39) Associations relevant for ACC vs. NOM selection:
   i. *Verbs having a higher association ratio with official texts will generally favor NOM (even in colloquial contexts), while verbs having a higher association ratio with colloquial texts will favor ACC (even in official contexts).*
   Underlying assumptions: (a) associations are determined by frequency effects: the more frequently the verb occurs in official contexts, the higher the association rate; (b) official contexts generally favor NOM due to prescriptivism;
   ii. *Verbs having a higher association ratio with animate objects will generally favor ACC (even with inanimate objects), while verbs having a higher association ratio with inanimate objects will favor NOM (even with animates).*
   Underlying assumptions: as has been shown in 4.4.4 animacy is a strong predictor which may impose implicationally coding preferences via frequency effects.

Consider the verbs *ap-vainot* 'to offend' and *ap-saukāt* 'to call so. offensive names' in Table 19. These two verbs score second highest after *provēt* 'to try' with respect to ACC selection. With respect to the associations in (ii), these verbs are strongly associated with animacy because they mostly take animate objects: *ap-vainot* 'to offend' has 91% (62/6) and *ap-saukāt* 'to call so. offensive names' 83% (24/5) of animate objects, in contrast to, for example, *dzirdēt* 'to hear' with 3% (2/66) of animate objects. Yet, *ap-saukāt* and *ap-vainot* have 48% and 26% of ACC, respectively, while *dzirdēt* attests only 15% of ACCs.

Regarding associations in (i), these verbs are less apt to be used in official texts for purely pragmatic reasons. So here ACC is also favored. Although we do not have counts for the average usage ratio of these verbs in colloquial-style vs. official texts, we have the following evidence that it indeed has an important impact. First, the very lexical meaning of these verbs is such that it would hardly fit official contexts. Second, other verbs with a similar animacy association (ii) fall considerably below these two verbs, for example, *satikt* 'to meet' (10% ACC) and *sazvanīt* 'to call, to phone' (7% ACC), cf. Table 19 above (with animacy associations of 99% (203/2) and 100% (95/0), respectively). Yet, these verbs are frequently found in various official texts: e.g. *to meet the prime minister* or *to call an institution in a particular case of emergency*. One may, of course, alternatively interpret this as saying that animacy is just not an important predictor and something else is at work here, but this option is simply ruled out for independent reasons: we have shown above that animacy (4.4.4) and, consequently, animacy association (immediately above) *is* an important predictor at the overall level of semantic verb classes where particular lexical peculiarities are leveled out. Therefore we assume that another, lexeme-specific factor imposing stronger effects than the animacy association is at

work here. We claim that this is mainly factor (ii) and, perhaps, some other minor lexeme-specific factors.

In addition to the associations laid out in (39), there are other, minor and lexeme-specific factors, one example being the association of *provēt* 'to try' with complex predication discussed above. Furthermore, the verbs *ap-vainot* 'to offend' and *ap-saukāt* 'to call so. offensive names', scoring second highest with respect to ACC marking after *provēt,* both have the prefix *ap-,* literally 'around', which on its non-lexical use often has the applicative function (of "accusativizing a syntactically peripheral NP", cf. Kozhanov, this volume), cf. *strādāt ar kaut ko* (lit.) 'to work with smth.', itr., vs. *ap-strādāt* (lit.) 'to process smth.', tr.[21] This morphological marking may additionally favor ACC marking specifically with these verbs.

4.5 Changes in the word-order preferences: Old ~ Early ~ Modern

We mentioned above (4.4.2) that word order is an important secondary predictor for the ACC selection. In this respect, it is not irrelevant that there are considerable changes in word-order preferences from Old Latvian into Contemporary Latvian. While Latvian is a regular "free-word-order" SVO language in all its stages, the word order in the debitive construction exhibits a considerable affinity towards SOV constellations. Thus we have 56% of VO and 44% OV, which makes an assumption of one of these being the basic word order of the debitive construction empirically unjustified. However, and importantly for our context, these preferences have nevertheless considerably changed since Old Latvian, as the data in Table 20 unequivocally testify:

Table 20. Position relative to the predicate

	OV	**VO**	**n/a**
Old Latvian	**92%** (105)	**8%** (9)	85
Early Modern Latvian	**63%** (250)	**37%** (149)	–
Modern (Contemporary) Latvian	**44%** (1407)	**56%** (1783)	–

We observe the shift in preferences towards the VO word order that is elsewhere basic in Latvian. Note that the label O above is meant semantically, indicating the second-rank argument of the lexical verb that is its direct object outside the

21. The non-lexical uses are, of course, not restricted to the applicative meaning only but may also have the "perfectivizing" function, or rather the function of a bounder, cf. *ēst* 'to eat' vs. *ap-ēst* 'to eat up'. Notably, both non-lexical functions are highly lexically restricted.

debitive construction. Yet, this argument has been the subject of the debitive construction to begin with, certainly in the Old Latvian period, hence, the basic word order OV (historically SV) in Old Latvian. However, this syntactic role – given the semantic proximity to the direct object – undergoes a gradual change towards objecthood in the course of grammaticalization of the debitive construction. The changes in word order highlighted above represent the direct consequence of this subject-to-object development. In Seržant (this volume), it is argued that the nominative-marked constituent behaves very much like an object from the syntactic point of view in Contemporary Latvian (*pace* Holvoet & Grzybowska 2014). Needless to say, the rise of ACC marking is another consequence of this development, which is particularly preferred in the preverbal position in order to cancel the incorrect expectations of this NP encoding the A participant (see 4.4.2 above).

5. The subject-like argument of the debitive

While Section 4 was devoted to the object-like argument of the debitive, in this section, we turn to the subject-like argument coded by DAT. The DAT argument normally precedes the debitive predicate in Contemporary Latvian (cf. already Stolz 1987):

Table 21. The preferred word order of the DAT argument

DAT-DEB	DEB-DAT
80% (237)	20% (58)

In what follows, we discuss here two phenomena: the increase of the impersonal usage of the debitive construction with the dropped DAT argument (5.1) and the occasional failure of the underlying NOM argument to be demoted into DAT with existential and existential-like verbs (5.2).

5.1 Changes in the preferences for the overt vs. *pro*-dropped usage

The first 1083 examples of our first Contemporary Latvian sample (Section 2.3) and all examples of our Old Latvian and Early Modern Latvian samples have been annotated for the word order of the DAT argument. Consider Table 22 below, which illustrates the changes that have been found with respect to personal (overt subject-like argument) vs. impersonal (*pro*-dropped) usage of the debitive from Old into Contemporary Latvian:

Table 22. Diachrony of the encoding of the subject-like argument

	Overt (DAT-marked)	**Impersonal (*subject* drop)**	**Other**
Old Latvian	66% (130)	34% (68)	0.5% (1)
Early Modern Latvian	48% (191)	52% (208)	–
Modern Contemporary Latvian	27% (295)	73% (788)	–

In contrast to the 34% of impersonal usage in Old Latvian and 52% in Early Modern Latvian, we found 73% (788 out of 1083) of impersonal use in Contemporary Latvian, which supports earlier claims (Stolz 1987: 225) of the predominantly impersonal nature of the debitive in the contemporary language. These 73% include also referential use of the dropped DAT argument referring to accessible referents. However, the latter are clearly in the minority, and most of the 73% contain utterances that have only generic human reference. Note also that the subject-like argument – dropped or overt – almost exclusively refers to humans (12 inanimates as opposed to 3178 animates, i.e. 0.4%).

The data in Table 22 shows that the overt DAT argument gradually becomes less and less frequent, while the impersonal usage of the debitive prevails with time. This development is somewhat reminiscent of the Polish modal *trzeba* 'have to, should' that equally superimposed the DAT case-marking on the (canonical) subject of the embedded lexical verb in Old Polish. However, this modal was more and more used impersonally and finally developed into an impersonal modal with no option to explicitly code the subject of the underlying verb in Modern Polish (Weiss 1993; Hansen 2001).

Furthermore, Table 22 indirectly supports the reconstruction suggested above that the debitive construction has been grammaticalized on the basis of the possessive use of the verb *būt* 'to be'. The possessive construction does not make much sense, if the possessor is not overtly coded. In turn, the gradual increase in grammaticalization and, hence, deviation from the original possessive meaning may indeed be responsible for the losing the semantic constraints requiring an overt possessor. This is another indication for the increasing semantic bleaching and hence grammaticalization degree of the debitive.

5.2 Occasional failure to demote the underlying NOM into DAT with existential verbs in the debitive

5.2.1 *Non-demoted NOM NPs*

The reconstruction of the debitive construction (cf. Section 3 above) suggests that DAT must have been the only possible case-marking strategy for the first argument (corresponding to the subject argument of the lexical verb elsewhere), since the case-marking of this argument directly stems from the predicative possessor dative in the former *mihi-est*-type possessive construction. Indeed, the Old Latvian and Early Modern Latvian samples support this by attesting the DAT strategy as the only possible strategy to mark the underlying subject. As has been mentioned in 3.1, there is one exception to this in Old Latvian in which the subject-like argument is marked by the ablative-like preposition *no* 'from' with the genitive case, which is an analogy to passives and most probably simply a mistake. Furthermore, the DAT option is also the only option found in the more conservative folklore texts of *Dainas* (Grīsle 2005: 8).

We find occasional failure to demote the underlying NOM subject argument of existential verbs into the DAT that presumably just has started in Contemporary Latvian and is not acceptable for most of the speakers (the relevant examples below all carry one question mark to highlight this). The earliest example known to us so far is from the newspaper *Latvijas Kareivis* from 1935. Crucially, the NOM marking is attested neither in our Old Latvian Sample nor in the Early Modern Latvian Sample.

In subsection 2.4 above, we have extensively argued that the rare NOM marking cannot be considered to be just performance errors, let alone typos, for the following reasons: (i) we have collected 126 authentic examples (see 2.4 above for why we consider our examples authentic) which is quite numerous given that our Google search was restricted to specific lexemes; (ii) the NOM marking is discussed by purists as a non-eligible option (cf. the instructions for native speakers in Grīsle 2005: 8); finally, (iii) as will be argued below, these examples form a clearly definable group both semantically and grammatically, which is not typical for performance errors. The following examples illustrate the phenomenon:

(40) (Contemporary Latvian)
?*Vai* ***viņš*** *jā-būt* ***līdzīpašnieks***
Q 3.NOM.SG.M DEB-be co-owner.NOM.SG
vai vienkārši ***algots*** ***darbinieks?***
or simply paid.NOM.SG.M employee.NOM.SG
'Does he (NOM) have to be the co-owner (NOM) or just an employee (NOM)?'[22]

22. http://www.lvportals.lv/visi/e-konsultacijas/6121-/?show=coment

(41) (Contemporary Latvian)
?*Kad lietotājs saņem vēstuli no bankas,*
viņš *jā-būt pārliecinātam,*
3.NOM.SG.M DEB-be convince.PPP.DAT.SG.M
ka tā patiešām saņemta no bankas
'When the user receives a message from the bank, he (NOM) has to be confident that it indeed comes from the bank.'[23]

(42) (Contemporary Latvian)
?*Neapšaubāmi* ***tas*** *jā-būt saistītam*
undoubtedly DEM.NOM.SG.M DEB-be relate.PPP.DAT.SG.M
ar faktu […]
with fact.ACC.SG
'Undoubtedly, this (NOM) must be related to the fact […]'[24]

(43) (Contemporary Latvian)
?*Viņam jā-piemīt* ***laba humorizjūta***
3.DAT.SG.M DEB-be.inherent.to good.NOM.SG.F sense.of.humor.NOM.SG
'He must have a good sense of humor (NOM).'[25]

As can be observed in the examples, the phenomenon is found only with intransitive S arguments of existential verbs or verbs containing the existential component in their lexical meaning. No agentive intransitive, let alone transitive, verbs have been found. Our preliminary interpretation is that the set of verbs allowing for the NOM marking is *mutatis mutandis* restricted by the verbs containing existence as part of their lexical meaning (alongside auxiliary and copular usage of *to be*). The other three verbs found with the NOM in the debitive also were all implying existence: *kļūt* 'to become' (= 'to begin to be'), *palikt* 'to stay, remain' (= 'to continue to be (at a certain place)'), *piemist* 'to be characteristic of, inherent':

(44) (Contemporary Latvian)
?***tas*** *jā-paliek noslepumā*
DEM.NOM.SG.M DEB-remain secret.LOC.SG
'This has to remain a secret.'

Regarding the verb *būt* 'to be' (124 examples out of 126), there are the following subtypes (in what follows, we base our findings on our second Contemporary Latvian Sample discussed in subsection 2.3 above):

---

23. http://www.inarchive.com/lv/2/2v.lv/5092343/2011-03-12-description/132/Katrs_Ir%C4%81kas_parlamenta_deput%C4%81ts_divdesmit_min%C5%AB%C5%A1u_laik%C4%81_nopeln%C4%ABjis_pa_90000_ASV_dol%C4%81riem/

24. www.carelinks.net/languages/latvian/Latvian_Bible_Basics.docx

25. http://ask.fm/streetlife/answer/14298237149

Table 23. The distribution of the uses of the verb *būt* with no change in case marking

Existential	Possessive	Copular			
		Nouns	Adjectives	P-oriented resultative participles	Adverbs
50	30	2	9	27	6

The position of the non-demoted NOM argument is preferably postverbal relative to the debitive gerundive:

Table 24. The position of the NOM argument relative to the debitive gerundive

NOM-DEB	DEB-NOM	No overt subject-like argument, only the predicative is in the NOM
25% (27)	75% (83)	16

Notably, most of the nouns occurring postverbally are abstract nouns. The reason for their preference for the postverbal position is that they are used generically and enter a somewhat closer semantic relationship with the verb *būt* 'to be', yielding one semantic unit (somewhat reminiscent of light-verb constructions), cf. *iespēja* 'possibility' with *jā-būt* [lit. 'there should be a possibility'] meaning 'it should be possible / one should be able to' or *vēlme* 'wish' with *jā-būt* [lit. 'there should be a wish'] meaning 'one should wish/want to'. Thus, *iespēja* with *jābūt* is found in 97% of all instances postverbally (55 out of 57 examples) and *vēlme* in 100% of all instances (5 out of 5 examples). Thus, the figures in Table 24 may be accidental due to the overweight of *iespēja* in the sample – with 57 examples it covers almost half of the sample. However, we will argue below that this is not coincidental and the non-referentiality is an important factor responsible for the occasional failure to demote the lexical NOM marking because the NOM is almost exclusively found in the context of weak or no referentiality.

Once we split all NOM NPs into nouns and pronouns, the following picture emerges. While nouns clearly prefer the postverbal position (with 87%), the pronouns are uniform in selecting the preverbal position only (100%):

Table 25. The position of the non-demoted NOM NP relative to the debitive gerundive

Nouns		Pronouns	
NOM-DEB	DEB-NOM	NOM-DEB	DEB-NOM
13% (12)	87% (83)	100% (15)	0% (0)

In order to test whether this positional distribution is specific for the non-demoted NOM NPs only, we have compiled an auxiliary sample. We included here the first

40 examples with nouns and all examples with definite pronouns (personal and demonstrative pronouns) found by searching Google for *jā-būt* 'DEB-be' (repetitions and dubious examples excluded). Only assertive examples have been taken into account. The following table illustrates the word-order preferences of the DAT-marked argument of *jābūt*:

Table 26. The position of the demoted DAT NP relative to the debitive gerundive

Nouns		Pronouns	
DAT-DEB	DEB-DAT	DAT-DEB	DEB-DAT
80% (32)	20% (8)	100% (6)	0% (0)

Contrast the positional preferences of the non-demoted NOM NP in Table 25 with the regularly demoted DAT NP in Table 26. While there is no difference with pronouns, the nouns show entirely reverse preferences: the DAT-marked nouns clearly prefer the preverbal position while the NOM marked nouns the postverbal position. From this we conclude that one of the factors favoring the NOM marking of the S argument with nouns is their postverbal position. The latter is, of course, not self-explanatory and results from various factors such as information-structure properties of the NOM NP and its individualization/referentiality degree.

Furthermore, out of 112 (93%) total instances of NOM, we have 106 inanimates and only 8 (7%) animates. Notably, even the animate NPs are non-individuated. Thus, all five instances of *cilvēks* 'person, human' attest generic reference, and none of them a specific, let alone definite, reference, cf. (45)–(47):

(45) (Contemporary Latvian)
?*Vai obligāti mājas tehniskā apskatē*
jā-būt ***cilvēks*** *ar* *būvinženiera*
DEB-be person.NOM.SG with construction_engineer.GEN.SG
diplomu?
diploma.ACC.SG
'Is it necessary that during the technical inspection at home there should be **a person** with a construction-engineer diploma?'[26]

(46) (Contemporary Latvian)
?***Cilvēks*** *jā-būt* *reģistrētam*
person.NOM.SG DEB-be register.PPP.DAT.SG.M
pie *ģimenes* *ārsta.*
at family.GEN.SG doctor.GEN.SG
'[Any] **person** must be registered at [her/his] family doctor.'[27]

26. http://www.ljb.lv/?c=diskusijaslasit&id=1724

27. www.lkndz.lv/vadlinijas/lkndz3.swf

(47) (Contemporary Latvian)
?*Bet tak ārstam pirmajā vietā*
but PTC doctor.DAT.SG first.LOC.SG.DEF place.LOC.SG
ir jā-būt ***CILVĒKS*** *nevis viņa samaksa.*
AUX.PRS.3 DEB-be human.NOM.SG NEG 3.GEN.SG.M pay.NOM.SG
'but then the doctor's priority should be the **PERSON** (=the patient) in the first place and not his pay.'[28]

Moreover, even the third person pronoun is found to have a generic antecedent, cf. the following example:

(48) (Contemporary Latvian)
?*Kad lietotājs saņem vēstuli no bankas,*
viņš *jā-būt pārliecinātam,*
3.NOM.SG.M DEB-be convince.PPP.DAT.SG.M
'When the user receives a letter from the bank **he** has to be convinced (that this letter is indeed from his bank.)'[29]

Another pronoun often occurring coded by the non-demoted NOM is the demonstrative/inanimate-third-person pronoun *tas*. In all its occurrences in our sample it refers to some situation and not to a particular entity:

(49) (Contemporary Latvian)
?*bet kad valdība ir novedusi valsti un tautu līdz bankrotam,*
un aizņemās miljardiem, un atkal uz tautas rēķina,
tad ***tas*** *tā jā-būt!!!!*
then DEM.NOM.SG.M SO DEB-be
'when the government brought the country and all the people to bankruptcy and ran into billions of debt, again at people's charge, then **this** has to be that way!!!'[30]

Thus, we conclude that the choice of the speakers not to change the underlying NOM into DAT with existential verbs (*sensu lato*) is facilitated by a low degree of individuation of the respective NP of which the low percentage of animates (7%) is also indicative. Furthermore, in the absolute majority of the examples in our sample the NOM NP is an abstract noun such as *vēlme* 'wish, desire' (5), *laba*

28. Note that it happened to us several times that an example has been removed from the web. This is especially frequent with various official blogs where there seems to be a proofreader correcting the "mistakes" and making our examples disappear from the web.

29. http://www.inarchive.com/lv/2/2v.lv/5092343/2011-03-12description/132/Katrs_Ir%C4%81kas_parlamenta_deput%C4%81ts_divdesmit_min%C5%AB%C5%A1u_laik%C4%81_nopeln%C4%ABjis_pa_90000_ASV_dol%C4%81riem/

30. http://www.tvnet.lv/zinas/latvija/323740-vai_viegli_iegut_maznodrosinata_statusu/comments

humorizjūta 'a good sense of humor' (1), *iespēja* 'possibility' (62 instances) or a mass noun such as *informācija* 'information' (17). These two types of nouns have notoriously vague reference.

In (62) below, we summarize the factors constraining the appearance of the NOM marking in form of a hierarchy (the factors are in italics). The preceding discussion suggests that an important condition for the NOM marking is that it enters a somewhat closer relationship with the predicate. This is evidenced by the word-order preference towards *postverbal position relative to the debitive predicate* and by the fact that the NOM marking – if it occurs at all – tends to be found with *low-individuation referents (generic reference, abstract/situation referents, non-referential predicatives, predominantly inanimates*). In turn, the verb must be existential or existential-like (*existential verbs only* in (62)).

Moreover, additionally, albeit less strongly in our view, the *presence of another, heterogeneous ("non-structural") dative*, either a free dative, experiencer dative (50), or possessor dative (if the *mihi-est*-possession construction was embedded into the debitive), cf. (52), may favor the NOM marking in order to avoid two DAT NPs bearing different semantic and syntactic roles in the same clause:

(50) (Contemporary Latvian)
?***Tas*** *jā-būt* ***katram*** *skaidrs*
DEM.NOM.SG.M DEB-be each.DAT.SG.M clear.NOM.SG.M
'This has to be clear to everyone.'[31]

The utterance in (50) has the complement dative *katram* 'to everyone' (corresponding to the dative in the indicative) which has a different semantic (experiencer) and syntactic role (complement/subject-like oblique) from the demonstrative *tas* 'it, this' (semantically the stimulus and syntactically the full-fledged subject). The latter also should have had the dative marking in the standard language:

(51) (Contemporary Latvian)
Tam *jā-būt* ***katram*** *skaidram*
DEM.DAT.SG.M DEB-be each.DAT.SG.M clear.DAT.SG.M
'This has to be clear to everyone.' [Constructed]

(52) (Contemporary Latvian)
?*Kādēļ nestrādā par velti*,
'Why don't they work for free'
ja jau ***ārstam*** *jā-būt* ***vēlme*** *tikai palīdzēt cilvēkiem?*
if PTC doctor.DAT.SG DEB-be wish.NOM.SG only help.INF people.DAT.PL
'if the doctor has to be willing to only help people'[32]

---

31. https://www.e-klase.lv/lv/zinas-komentari/zinas/aktualitates/eksperte-jauniesiem-vairak-jamaca-izprast-inovaciju-rasanas-nevis-japiespiez-macities-fiziku-un-kimiju/3/

32. http://cosmo.lv/forums/topic/85528-/?sort=desc

Finally, the non-demoted NOM occurs often in questions introduced by *kas* 'what/who':

(53) (Contemporary Latvian)
?Bet ***kas*** *jā-būt pārreģistrācijas dokumentos?*
but what.NOM DEB-be re-registration.GEN.SG document.LOC.PL
'But what should the re-registration documents contain?'[33]

Problematic here is that the expected DAT marking of the *wh*-pronoun *kas* is more likely to be interpreted as animate 'who', which would not fit the context.

Furthermore, we observe that *Distinguishability* plays an important role constraining the appearance of the NOM marking. Distinguishability is one of the general functions of case: to distinguish between the two arguments of a monotransitive clause in order for it to be correctly interpretable (cf., *inter alia*, Comrie 1978, 1989; Dixon 1979; Silverstein 1976; Kibrik 1997; formalized in Aissen's *Markedness Relation* 2003). While Distinguishability is typically not among the most important functions of case cross-linguistically (cf., *inter alia*, Malchukov 2008; DeHoop & DeSwart 2009, eds.), it does play some role here. Distinguishability manifests itself in the fact that the NOM marking has replaced the original DAT marking precisely with intransitive and not with transitive verbs. Thus, while both NOM and DAT marking do not create misinterpretations and can be correctly processed with the S argument, cf. (54), this is not the case with a transitive verb ('to indicate') in the same context, cf. (55):

(54) (Contemporary Latvian)
Bet ***?kas/kam*** ***jā-būt*** *pārreģistrācijas*
but **WHAT.NOM/DAT** DEB-be re-registeration.GEN.SG
dokumentos?
document.LOC.PL
[lit.] 'But **what (?NOM/DAT)** should there be in the re-registration documents?'
[NOM – attested in (53) / DAT – constructed on the basis of (53)]

(55) (Contemporary Latvian)
a. *Bet* ***kas*** ***jā-norāda*** *pārreģistrācijas*
but **WHAT.NOM** DEB-indicate re-registeration.GEN.SG
dokumentos?
document.LOC.PL
'But **what (NOM)** should be indicated in the re-registration documents?'
[Constructed on the basis of (53)]

33. http://www.vid.lv/lv/forums/21/71/1__2_2_1/2072

(55) b. *Bet* ***kam*** ***jā-norāda*** *tas* *pārreģistrācijas*
but **WHAT.DAT** DEB-indicate DEM.NOM.SG re-registeration.GEN.SG
dokumentos?
document.LOC.PL
'But **who (DAT)**[34] should indicate this in the re-registration documents?'
[Constructed on the basis of (53)]

Thus, in addition to the conditions for the non-demotion of the NOM identified above, *Distinguishability* plays an important role.

5.2.2 *Case agreement on the predicative (adjective, participle, etc.)*
Concomitantly with the incipient penetration of the NOM into the DAT domain with existential verbs,[35] there is also an incipient penetration of the NOM case-marking on the predicatively used participles and adjectives:

(56) (Contemporary Latvian)
?*Manuāļos* *jā-būt* ***norādīts*** *kādam*
manual.LOC.PL DEB-be **indicate.PPP.NOM.SG.M** which.DAT.SG.M
spiedienam *ir* *jābūt.*
pressure.DAT.SG AUX.PRS.3 DEB-be
'It has to be indicated in the manuals how [strong] the pressure should be.'[36]

In this example, the S argument of the second clause *kādam spiedienam* 'which pressure' is regularly marked with the dative case. However, the predicative resultative participle *norādīts* 'indicated' of the first clause is in the nominative instead of the expected dative case. The first clause is an impersonal resultative embedded under debitive. We have argued above that non-referentiality and close ties with the verb are the conditions in which the failure to demote the NOM may be found. The participle *norādīts* indeed is not linked to any referent (via agreement), while the (standard) dative marking on the participle might have yielded the wrong expectation that there is some specific *pro*-dropped dative subject with which it agrees.

While (56) does not represent a case of agreement with non-demoted NOM, in what follows we focus on agreement with a failure to demote NOM into DAT. Thus, the following example illustrates agreement in NOM case:

---

34. Additionally, there is a third reading which is not relevant here.

35. Penetration is meant here diachronically. Synchronically, the NOM must be analyzed as just not undergoing the transformation into DAT.

36. http://audi-style.lv/forum/topic/32511-audi-a4-1-8t-1998g-motora-problema/

(57) (Contemporary Latvian)
?*informācija* *jā-būt* *redzama*
information.**NOM**.SG DEB-be visible.**NOM**.SG.F
'the information has to be visible.'

The next examples exhibit agreement only in number and gender but not in case:

(58) (Contemporary Latvian)
?*Latvijā* *informācija* *jā-būt* *pieejamai*
Latvia.LOC information.**NOM**.SG.F DEB-be accessible.**DAT**.SG.F
latviešu *valodā*
Latvian language.LOC.SG
'In Latvia, the information must be available in the Latvian language.'[37]

(59) (Contemporary Latvian)
?*Cilvēks* *jā-būt* *reģistrētam* *pie*
person.**NOM**.SG.M DEB-be register.PPP.**DAT**.SG.M at
ģimenes *ārsta*
family.GEN.SG doctor.GEN.SG
'[Any] person must be registered at [her/his] family doctor.'[38]

Given that other agreement features such as gender and number are always copied by the predicative participle, we assume that this is not an instance of agreement suspension, but simply a failure to demote one of the NOMs.[39]

The evidence from the predicative participles and adjectives also supports our claim above that low degree of individuality and close relation with the predicate favor NOM marking. This is, of course, true of any predicatively used expressions, including the predicative adjective in (58) and participle in (59) which only ascribe particular properties to the referent but are not referring themselves.

37. http://www.pvd.gov.lv/lat/_gv/module_faq/cat-id_64/page_37?text_ver=1

38. www.lkndz.lv/vadlinijas/lkndz3.swf

39. Note that the very fact that there is agreement in number and gender but not in case is not necessarily an indication for performance errors, contrary to what a reviewer suggested. Thus, in a similar construction of Russian, the modal-infinitive construction, we likewise find only agreement in number and gender, not in case, cf. dative-instrumental *Emu ne byt' sil'nym* 3.SG.M.**DAT** NEG.be.INF strong.SG.M.**INS** 'He will never be(come) strong.'

6. Conclusions

We have discussed the Differential Argument Marking systems of the Latvian debitive construction from both diachronic and synchronic perspectives and with respect to both the marking of the subject-like argument (the occasional non-demotion of NOM) and the object-like argument (NOM > ACC) in present-day Contemporary Latvian. In what follows, our general explanation for these two alternations will consist of two parts: (6.1) motivations for the rise of new case-marking strategies and (6.2) factors and conditions that constrain the gradual and incipient expansion of these new case-marking strategies but are themselves not motivations for their appearance. While we assume that the motivations are largely the same for both NOM > ACC and DAT > NOM, the factors constraining their occurrence are quite different, cf. 6.2.1 and 6.2.2. We begin first with some general motivations for these DAMs and then proceed to more specific factors.

6.1 Motivations

In general terms, the development towards canonical case assignment and, at the same time, towards the case assignment required by the embedded lexical verb represents gradual abandonment of the debitive's own case frame. That is, the competition between the overridden case frame of the lexical verb and the case frame of the debitive leads to the enforcement of the lexical verb's case frame and syntax (see Seržant 2013: 349–351 on the typological discussion).[40] This, in turn, represents, in our view, another step towards grammaticalization or, rather, syntacticization of the debitive construction, for the following reason. A fully grammaticalized modal should not have a case frame and mapping rules for syntactic roles of its own, because it does not license arguments or assign semantic roles that need to be coded (by case); instead, the arguments, their semantic roles and, consequently, the case assignment come from the respective lexical verb. Generally, in the gradual development from a lexical verb into a new modal verb the original lexical government may be retained in the transitional period as a lexical residual. This is exactly the situation with the debitive here. Note that the case frame of the debitive is the original one of the matrix lexical verb *to be* in its possessive use: *to me is X* 'I (DAT) have X (NOM)'. This possessive verb turned into an auxiliary, and hence synchronically the case frame of the embedded lexical verb must be analyzed as overridden by the case frame of the whole debitive construction or, in formal

40. Note that this process is not unique to the debitive construction only. Other modal constructions of Baltic underwent the same process. Thus, the modal-infinitive construction before it was lost in Old Latvian was almost consequently used with DAT-ACC (cf. examples in Andronovs 1998: 174, *passim*; Vanags 2000, *passim*) instead of the original DAT-NOM.

terms, by some head higher than the VP.[41] Notably, concomitant or perhaps, even causally primary (cf. the *Behavior-before-coding principle* in Haspelmath 2010) to the introduction of new case is the gradual change in the assignment of syntactic roles in the debitive (Seržant, this volume), again in accordance to the assignment of the syntactic roles by the lexical verb:

(60) *Gradual re-assignment of grammatical relations in the debitive construction*
From original $\text{DAT}_{\text{COMPLEMENT}}$-$\text{NOM}_{\text{subject}}$ into $\text{DAT}_{\text{subject}}$-$\text{NOM}_{\text{object}}$.

The incipient penetration of the syntactic roles and case marking of the embedded lexical verb discussed here is another indicator of the increase of internal dependencies and semantic bleaching, respectively, that the former matrix possessive verb (and now auxiliary) underwent. Furthermore, the fact that the competition between the case and syntactic role assignments of the lexical verb and the ones of the debitive construction is being incipiently abandoned is yet another indication for an increasing syntacticization of what was originally two clauses into one complex clause, hence, into a more bound grammatical unit with more internal dependencies. Both the loss of lexical properties and the increase of internal dependencies might be interpreted as a further increase in the degree of grammaticalization of the debitive (cf. Heine, Claudi & Hünemeyer 1991: 2; Haspelmath 2004; Givón 1979: 208 on grammaticalization).

Note that the increasing degree of grammaticalization of the debitive in Latvian may have been facilitated by external factors, such as the influence of the neighboring languages as originally suggested by purists (cf., *inter alia*, Grīsle 2008: 5). Thus, we cannot exclude the influence of the two major regional languages that historically have had a strong impact on Latvian, namely German and Russian. Both languages indeed employ obligation modals which do not alter the case-assignment patterns of the lexical verb, cf. the synonymous verbs in German *sollen*, *müssen*, *haben zu* and the Russian predicative adjective *dolžen*.[42] It is the task of future research to find out to what extent, and whether at all, there was some external influence.

The increasing grammaticalization degree of the debitive manifests itself in adjustments of its own cases and syntactic roles to those of the embedded lexical verb, producing in effect a more canonical structure. We claim that this process

41. It is our conviction that the case frame of the debitive is not ascribable to a particular part of the construction, although one is tempted to host it at the debitive gerundive. The latter is doubtful because precisely the same case frame is found in the analogous modal-infinitive construction of Lithuanian (cf., *inter alia*, Seržant, this volume), and for Lithuanian one would have to ascribe the case frame to the infinitive, which is not tenable.

42. Note that marginally, even in Estonian, there is the modal verb *vajama* 'have to' (lexically 'to need') that equally does not alter the case frame of the lexical verb.

is the main motivation for the occasional NOM marking instead of the old DAT on the subject-like argument, and for the ACC marking instead of the NOM on the object-like argument. Additionally, there are specific constraints on how exactly the expansion of these new case-marking strategies proceeds.

We have mentioned above (5.2) that Distinguishability plays an important role in creating constraints on where these new case-marking strategies may appear, which is why the occasional non-demotion of the NOM into DAT is found with S arguments only and never with A arguments (the latter would have yielded *NOM-NOM alongside *NOM-ACC structures). Furthermore, regarding the properties of the referents, we observe the tendency for the NOM marking to be associated with low-individuated referents, while the oblique marking, namely the ACC, is associated with highly individuated referents; the DAT marking is the default for A/S and is found with both low-individuated and highly-individuated referents. Thus, the difference between the two processes discussed in the paper, namely, NOM > ACC and DAT > NOM, is that the new marking (NOM) in the subject domain may only be found with low-individuated NP types whereas, in contrast, the expansion of the new marking (ACC) in the object domain first targets highly-individuated/salient referents:

Table 27. Tendencies in case selection

Subject-like argument		Object-like argument	
Highly-individuated referents	**Low-individuated referents**	**Low-individuated referents**	**Highly-individuated referents**
DAT	DAT / (**NOM**)	**NOM** / (ACC)	ACC / (NOM)

Notably, this picture is vaguely reminiscent of an active/inactive alignment system (or split-S alignment) in which there is a split between $A/S_{individuated}$ (DAT) vs. $S_{non\text{-}individuated}/P$ (NOM) with an additional differentiation into $P_{individuated}$ (ACC) vs. $P_{less\ individuated}$ (NOM). In very general terms, the processes discussed run against the tendency to distinguish only those arguments that are atypical for the grammatical relation they bear, for example, to mark only atypical A's and P's (cf. Comrie 1978, 1989): highly-individuated A's would have been expected to have NOM marking since they are typical A's anyway, while highly-individuated P's do indeed confirm this tendency by tending to the ACC marking. At the same time, the NOM marking shows some inclination towards low-individuated NPs.[43]

43. In Baltic there is a general problem in disentangling morphologically more and morphologically less marked options, since all three cases at issue are morphologically (un)marked to the same degree. To account for this from a different perspective, one may assume that the nominative case is the functional default (and, hence, the unmarked) option, while DAT and ACC are functionally more specific, but this is not mirrored in morphology (cf. Seržant, this volume).

6.2 Conditions and constraints

6.2.1 *The object domain*

We have shown that the NOM marking is being gradually replaced by ACC. While only a few examples of this are found in Old and Early Modern Latvian, the change has gained ground particularly in Contemporary Latvian, which requires the ACC marking with SAP pronouns and the reflexive pronoun, strongly tends to have ACC with reciprocals, and frequently admits ACC marking with other definite pronouns, cf. the Effect scale in (36). We emphasize that the change from NOM into ACC for all third-person NP types is incipient, affecting on average only 13% of all third-person object arguments in the sample (the factor "*Incipiency of the spread of ACC*" in (61) below) and it is impeded by prescriptive rules which are strong in Latvia ("*Prescriptive rules imposing the retrograde effect*" in (61)). In addition to these primary-importance factors there are five secondary-importance factors.

ACC is not randomly distributed across all third-person NP types. We have argued that there are four factors constraining the preference for ACC or NOM within the limits in which the change has taken place. The most important constraint is the linear position relative to the debitive predicate ("*the positional factor*"): OV order favors ACC marking while VO order favors the NOM marking (4.4.2). The second important constraint is the degree of accessibility of the respective NP ("*the accessibility factor*") represented in (36) above (4.4.3): the higher the accessibility degree, the stronger preference towards ACC. Then, animacy ("*the animacy factor*") dependent on, and additional to, accessibility is the third factor favoring ACC (4.4.4).

We have argued that the preverbal position, high accessibility degree and being animate may be cumulatively referred to as saliency. Thus, we claimed that the appearance and expansion of the ACC marking in the debitive construction gradually proceeds from the most salient to the less salient NPs. The reason why salient NPs are affected first is the conflict of contradictory expectations associated with saliency, case marking and syntactic roles: a salient NP has a frequency-based expectation to be the subject, and even more so if it is marked NOM case (cf. Seržant, this volume, on the varying correlation between subjecthood and different NP types marked NOM); yet, at the same time, the NOM NP is not a true subject but has rather acquired a number of object-like properties according to (60). To solve this conflict, the ACC case is introduced gradually from the most unexpected towards less unexpected constellations. The very appearance of the ACC here is facilitated by those NPs that are absolutely not expected to occur in the subject position, such as reciprocals. The latter morphologically may have dedicated nominative forms in Latvian but for obvious reasons simply do not occur in the subject position. In order for them not to be unexpectedly associated with subjects, they acquire accusative marking. Similarly, the SAP pronouns – differently from other NP types in Baltic – *must* be subjects if they bear the nominative case

(cf. Seržant, this volume); the emerging object-like properties must have forced them therefore to acquire ACC marking.

Finally, different verb classes have different preferences for ACC vs. NOM case selection ("*the verbal factor*"), experiencer verbs scoring highest with ACC marking (4.4.7). In turn, among the experiencer verbs, there are particular lexemes that additionally facilitate the choice of ACC such as *apsaukāt* 'to call so. offensive names' due to their frequency-based associations and other, minor factors ("*the lexical factor*"), cf. 4.4.7. We have shown that definiteness of the input NP has revealed itself as not being significant (4.4.5).

6.2.2 *The subject domain*

The occasional failure to demote the underlying NOM into DAT found with existential verbs in the debitive construction is mainly constrained by low-individuation of the referent, as we have argued above. Furthermore, and admittedly a consequence of this, the NOM marking tends to occur in the VO position relative to the debitive predicate, whereas the DAT marking with the same verb *būt* 'to be', in contrast, tends to occur preverbally.

6.3 All motivations and predictors: Summary

The following motivation and predictors scale summarizes the results from the preceding sections, schematically presenting the factors and their impact on NOM/ACC case selection in the object-like domain in (61) and on DAT/NOM case selection in the subject-like domain in (62):

(61) Factors and motivations predicting ACC case selection in the object-like domain
Motivations: *Increasing grammaticalization/ Incipient loss of the debitive's own case frame / Canonicization* > **Primary-importance factors:** > *Incipiency of the spread of* ACC > *Prescriptive rules imposing the retrograde effect* > **Secondary-importance factors:** > *The positional factor* > *the accessibility factor* (36) > *The animacy factor* > *The verbal factor* > *The lexical factor*

(62) Factors and motivations predicting occasional NOM case selection in the subject-like domain
Motivations: *Increasing grammaticalization/ Incipient loss of the debitive's own case frame / Canonicization* > **Primary-importance factors:** > *Distinguishability* > *Occasional nature of the appearance of* NOM *(for dative)* > *Prescriptive rules imposing the retrograde effect* > **Secondary-importance factors:** > *Existential verbs only* > *Low individuation (generic reference, abstract/ situation referents, non-referential predicatives, predominantly inanimates)* // *Postverbal position relative to the debitive predicate* // *Presence of another, heterogeneous ("non-structural") dative*

The factors and motivations in (61) and (62) can be further split at a more fine-grained level. The factors *animacy* and the *lexical* factor in (61) yield animacy associations of particular verbs, cf. 4.4.7. Unfortunately, we cannot exactly determine the ranking of the last three factors in (62), and these factors are separated using // to highlight this.

Acknowledgements

We thank Axel Holvoet, Andra Kalnača, Nicole Nau and the external reviewer for their extensive and highly valuable and helpful discussions that improved the paper considerably.

Abbreviations

ACC	accusative	NEG	negation
ADV	adverb	NOM	nominative
AUX	auxiliary	PL	plural
CNV	converb	PPA	past active participle
DAT	dative	PPP	past passive participle
DEB	debitive	PRS	present
DEF	definite	PST	past
DEM	demonstrative pronoun	PTC	particle
F	feminine	Q	question particle
FUT	future	RECIPR	reciprocal pronoun
GEN	genitive	REFL	reflexive
INF	infinitive	REL	relative pronoun
INDEF	indefinite pronoun	RPO	reflexive possessive
LOC	locative	SAP	speech act participants
M	masculine	SG	singular
N	neuter	WH	*wh*-pronoun

Corpora

SENIE: Digitalized collection of Old Latvian texts hosted at http://www.korpuss.lv/senie
PERIODIKA: Digitalized collection of Latvian periodicals hosted at http://www.periodika.lv
CCLL: Corpus of the Contemporary Lithuanian Language. University of Kaunas. (Available online at http://tekstynas.vdu.lt/tekstynas/)

References

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy, *Natural Language & Linguistic Theory* 21: 435–483. DOI: 10.1023/A:1024109008573

Andronovs, Aleksejs 1998. "Vajadzības izteiksme" latviešu valodas gramatiskajā tradīcijā. 1. daļa (XVII–XVIII gs.), *Baltu filoloģija* 8: 154–177.

Ariel, Mira. 1990. *Accessing Noun-Phrase Antecedents*. London/New York: Routledge.

Ariel, Mira. 2001. Accessibility Theory: An overview. In *Text Representation: Linguistic and Psycholinguistic Aspects*, Ted Sanders, Joost Schilperoord and Wilbert Spooren (eds), 29–87. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/hcp.8.04ari

Bergmane, Anna (ed). 1959. *Mūsdienu latviešu literārās valodas gramatika. Fonētika un morfoloģija. Valodas un Literatūras Institūts*. Rīga: LPSR Zinātņu akadēmijas izdevniecība.

Bossong, Georg. 1982. Der präpositionale Akkusativ im Sardischen. In *Festschrift für Johannes Hubschmid zum 65. Geburtstag*, Otto Winkelmann and Maria Braisch (eds.), 579–599. Bern: Francke.

Bossong, Georg. 1998. Le marquage différentiel de l'objet dans les langues d'Europe. In *Actance et valence dans les langues de l'Europe*, Jack Feuillet (ed), 193–258. Berlin: Mouton de Gruyter.

Bybee, Joan L. 2001. *Phonology and Language Use* [Cambridge Studies in Linguistics, 94]. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511612886

Bybee, Joan L., Perkins, Revere & Pagliuca, William. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago: The University of Chicago Press.

Bybee, Joan L. & Hopper, Paul. 2001. Introduction to frequency and the emergence of linguistic structure. In *Frequency and the Emergence of Linguistic Structure*, Joan L. Bybee and Paul Hopper (eds), 1–26. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/tsl.45.01byb

Comrie, Bernard. 1978. Genitive-accusative in Slavic: The rules and their motivation, *International Review of Slavic Linguistics* 3 (1–2): 27–42.

Comrie, Bernard. 1989. *Language Universals and Linguistic Typology, 2nd edition*. Oxford: Blackwell.

Croft, William. 2003. *Typology and Universals*. 2nd edition. Cambridge: Cambridge University Press.

de Hoop, Helen & de Swart, Peter (eds). 2008. *Differential Subject Marking*. Dordrecht: Springer.

Dixon, Robert M.W. 1979. Ergativity, *Language* 55: 59–138. DOI: 10.2307/412519

Dunkel, George E. 2013. *Lexikon der indogermanischen Partikeln und Pronominalstämme*. Band 2. Heidelberg: Universitätsverlag Winter.

Endzelin, J. 1901. Ursprung und Gebrauch des lettischen Debitivs, *Bezzenbergers Beiträge* 26: 66–74.

Endzelin, J. 1905. Zur Entstehung des lettischen Debitivs, *Bezzenbergers Beiträge* 29: 320–321.

Endzelīns, Jānis. 1951. *Latviešu valodas gramatika*. Rīga: Latvijas valsts izdevniecība.

Fauconnier, Stefanie. 2011. Differential agent marking and animacy, *Lingua* 121: 533–547. DOI: 10.1016/j.lingua.2010.10.014

Fennell, Trevor G. 1984. *Georg Dreszell's Gantz kurtze Anleitung zur lettischen Sprache. Text. Translation. Commentary. Concordance*. Melbourne: Latvian Tertiary Committee.

Fennell, Trevor G. 1973. The subject of Latvian verbs in the debitive mood. In: *Baltic Literature and Linguistics*, Arvīds Ziedonis (ed), 213–221. Columbus, Ohio: Association for the Advancement of Baltic Studies.

Fennell, Trevor G. & Gelsen, Henry. 1980. *A Grammar of Modern Latvian*. 3 vols. Den Haag: Mouton. DOI: 10.1515/9783111398457

Fraenkel, Ernst.1928. Litauische Beiträge. *Indogermanische Forschungen* 46: 44–57.

Givón, Talmy. 1979. *On Understanding Grammar*. New York: Academic Press.
Grīsle, Rasma. 2005. SPĒKILDZE. *Populārvalodniecisku rakstu izlase. 2., elektroniskais izdevums*. (Available online at www.eraksti.lv/fetchbook.php?urlkey=2491909, Accessed on 30 08 2015).
Hansen, Björn. 2001. *Das Modalauxiliar im Slavischen. Grammatikalisierung und Semantik im Russischen, Polnischen, Serbischen/Kroatischen und Altkirchenslavischen*. [Slavolinguistica, 2.] München: Otto Sagner.
Haspelmath, Martin. 1989. From purposive to infinitive – a universal path of grammaticalization, *Folia Linguistica Historica* 10: 287–310.
Haspelmath, Martin. 2004. On directionality in language change with particular reference to grammaticalization. In *Up and Down the Cline: The Nature of Grammaticalization*, [Typological Studies in Language, 59], Olga Fischer, Muriel Norde and Harry Perridon (eds), 17–44. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/tsl.59.03has
Haspelmath, Martin. 2008. Creating economical patterns in language change. In *Linguistic Universals and Language Change*, Jeff Good (ed), 185–214. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199298495.003.0008
Haspelmath, Martin. 2010. The Behaviour-before-Coding Principle in syntactic change. In *Mélanges Denis Creissels*, Franck Floricic (ed), 493–506. Paris: Presses de l'École Normale Supérieure.
Heine, Bernd, Claudi, Ulrike & Hünnemeyer, Friederike. 1991. *Grammaticalization: A Conceptual Framework*. Chicago: University of Chicago Press.
Hettrich, Heinrich. 1988. *Untersuchungen zur Hypotaxe im Vedischen*. Berlin: De Gruyter. DOI: 10.1515/9783110892123
von Heusinger, Klaus. 2008. Verbal semantics and the diachronic development of DOM in Spanish, *Probus* 20 (1): 1–33. DOI: 10.1515/PROBUS.2008.001
von Heusinger, Klaus. To appear. The diachronic development of Differential Object Marking in Spanish ditransitive constructions. In *The Diachronic Typology of Differential Argument Marking*, Ilja A. Seržant, Alena Witzlack-Makarevich and Kelsey Mann (eds). Berlin: Language Science Press.
Holvoet, Axel. 1998. Notes on the rise of grammaticalization of the Latvian debitive, *Linguistica Baltica* 7: 101–118.
Holvoet, Axel. 2007. *Mood and Modality in Baltic*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
Holvoet, Axel & Grzybowska, Marta. 2014. Non-canonical grammatical relations in a modal construction: The Latvian debitive. In *Grammatical Relations and their Non-Canonical Encoding in Baltic* [Valency, Argument Realization and Grammatical Relations in Baltic, 1], Axel Holvoet and Nicole Nau (eds), 97–136. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/vargreb.1.03hol
Kalnača, Andra & Lokmane, Ilze. 2014. Modal semantics and morphosyntax of the Latvian debitive. In *Modes of Modality. Modality, Typology, and Universal Grammar*, Elisabeth Leiss & Werner Abraham (eds), 167–192. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/slcs.149.06lok
Kibrik, Alexander E. 1997. Beyond subject and object: Toward a comprehensive relational typology, *Linguistic Typology* 1: 279–346. DOI: 10.1515/lity.1997.1.3.279
Kiparsky, Valentin. 1969. Das Nominativobjekt des Infinitivs im Slavischen, Baltischen und Ostseefinnischen, *Baltistica* 5 (1969): 141–148.
Laca, Brenda. 2006. El objeto directo. La marcación preposicional. In *Sintaxis histórica del español. Primera parte: La frase verbal*, Concepción Company Company (ed), 423–475. Mexico: Fondo de cultura económica / Universidad Nacional Autónoma de México.

Malchukov, Andrej L. 2008. Animacy and asymmetries in differential case marking, *Lingua* 118: 203–221. DOI: 10.1016/j.lingua.2007.02.005

McGregor, William B. 1992. The semantics of ergative marking in Gooniyandi, *Linguistics* 30: 275–318. DOI: 10.1515/ling.1992.30.2.275

McGregor, William B. 1998. Optional ergative marking in Gooniyandi revisited: Implications to the theory of marking, *Leuven Contributions in Linguistics and Philology* 87 (3–4): 491–571.

ME – J. Endzelins (ed.), *K. Mīlenbacha Latviešu valodas vārdnīca 1*, Chicago: Die Gruppe der lettischen Baltologen in Chikago, 1953. (Quoted after http://www.tezaurs.lv/mev/.)

Mühlenbach, Karl. 1907. O debitivě, *Izvěstija Otděnija russkago jazyka i slovesnosti Imperatorskoj Akademii nauk* 12 (3): 313–333.

Nau, Nicole. 2012. Modality in an areal context: The case of a Latgalian dialect. In *Grammatical Replication and Grammatical Borrowing in Language Contact*, Björn Wiemer, Bernhard Wälchli and Björn Hansen (eds), 471–514. Berlin: Mouton de Gruyter.

Rūķe-Draviņa, Velta. 1977. *The Standardization Process in Latvian: 16th Century to the Present* [Stockholm Slavic studies]. Stockholm: Almqvist & Wiksell.

Seržant, Ilja A. 2005. Otnositeľnaja xronologija osnovnyx fonetičeskix izmenenij v istorii verxnelatyšskogo dialekta, *Acta Linguistica Lithuanica* 53: 39–90.

Seržant, Ilja A. 2013. The diachronic typology of non-canonical subjects. In *The Diachrony of Non-Canonical Subjects* [Studies in Language Companion Series, 140], Ilja A. Seržant and Leonid Kulikov (eds), 313–360. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/slcs.140.14ser

Seržant, Ilja A. & Witzlack-Makarevich, Alena. To appear. Introduction. In *Diachronic Typology of Differential Argument Marking* [Studies in Diversity Linguistics], Ilja A. Seržant, Alena Witzlack-Makarevich and Kelsey Mann (eds). Berlin: Language Science Press.

Silverstein, Michael. 1976. Hierarchy of features and ergativity. In *Grammatical Categories in Australian Languages*, Robert M.W. Dixon (ed), 112–171. New Jersey: Humanities Press.

Stolz, Thomas. 1987. Das Dativsubjekt. Ein Beitrag zur lettischen Kasuslehre. *Indogermanische Forschungen* 92: 220–242.

Timberlake, Alan. 1974. *The Nominative Object in Slavic, Baltic, and West Fennic*. München: Otto Sagner.

Vanags, Pēteris. 2000. Debitīva formu lietojums visvecākajos latviešu rakstos, *Baltu Filoloģija* 9: 143–155.

Vanags, Peteris. 2012. Problems of Standard Latvian in the 19th Century. Changes in the language planning and changes in planners, from Baltic-Germans to Latvian. In *Nation und Sprache in Nordosteuropa im 19. Jahrhundert*, Konrad Maier (ed), 199–207. Wiesbaden: Harrassowitz Verlag.

Wälchli, Bernhard. 1996. Two cases of necessity modality in the NEE Area: External necessity deriving from motion verbs and the Latvian-Livonian debitive as a case of parallel grammaticalization. In *Narmon'Gi. Arbeitspapiere des Berner Projekts zur vergleichenden Darstellung der nordosteuropäischen Sprachen 1*, Simon Christen, Jan Peter Locher and Bernhard Wälchli (eds), 42–51. Bern: Institut für slavische und baltische Sprachen und Literaturen, Universität Bern.

Wälchli, Bernhard. 2000. Infinite predication as a marker of evidentiality and modality in the languages of the Baltic region. *Sprachtypologie & Universalienforschung* 53 (2): 186–210.

Weiss, Daniel. 1993. Infinitif et datif en polonais moderne – un couple malheureux? In *Complétude et incomplétude dans les langues romanes et slaves*, Stanisław Karolak and Teresa Muryn (eds), 443–487. Cracovie: WSP.