JS Journal of HistoricalSyntax
Volume 6, Article 8: 1–40, 2022

A LONG-HAUL CHANGE: DIFFERENTIAL OBJECT MARKING IN EARLY SLAVONIC *

Hanne Eckhoff
University of Oxford

Abstract This article uses extensive treebank data from the PROIEL and TOROT treebanks to track the much-debated rise of the animacy category in Russian, which in this article will be analysed as a change from at least partly definiteness-driven differential object marking in Old Church Slavonic via constructionally-conditioned variation in Old East Slavonic to fully fledged animacy subgender marking in late Middle Russian. The change is interesting from a methodological point of view as well, since it requires us to annotate data through an ongoing change, and also since conventional treebank annotation is not enough to capture the conditions of the observed variation and change: annotation for semantics and information structure is necessary too. The article describes and defends a conservative approach to annotation in the face of change: the analysis that fits the first attested stage of a change is retained as long as possible.

1 Introduction

Diachronic treebanks large enough to use in serious studies of syntactic change are a relatively new thing, simply because creating them is time-consuming and demanding. Given the nature of historical texts, using purely automatic methods does not seem feasible – the texts, particularly the earliest ones, are often linguistically complex and highly unstandardised, and the source material for some periods is so limited that we cannot afford to work with analyses with success rates of 80–90%, especially if the phenomenon under scrutiny is of low or medium frequency.

This article takes advantage of the recently developed diachronic treebank resources for Russian and Old Church Slavonic (OCS) to revisit a much-

* This study would not be possible without the extensive work on information structure annotation done on the project Pragmatic Resources in Old Indo-European Languages (PROIEL), where Dag Haug was PI and I was lucky enough to be a postdoc in 2008–2012. My warmest thanks go to all the project members who were involved.

debated topic in Russian and Slavonic linguistics: the question of the origins and development of the animacy category. While much has been written on the topic, explicit corpus studies are relatively scarce. In particular, although studies of the rise of animacy in the history of Russian regularly refer to OCS data, there is an obvious opening for a rigorous corpus study analysing data from both the earliest attested situation in OCS *and* the development from Old East Slavonic (OES) towards modern Russian. The topic is also methodologically interesting in several ways. It raises the question of how we can annotate a phenomenon undergoing change in a way that ensures data extraction with minimal amounts of noise. This is especially important since the earliest stages of the development are very scarcely attested, so that it is essential to be able to retrieve all the relevant data with as little noise as possible. The topic is also one that demonstrates the limits of conventional treebank annotation: in the early stages of the development the alternation between nominative-accusative (NA) and genitive-accusative (GA) is clearly conditioned by semantic and pragmatic factors, raising a need for detailed and principled annotation for semantic features, information status and anaphoric relations.

I shall argue that the modern Russian state of affairs came about through two separate change processes that were very different in pace and nature. First, we see a long, slow change from at least partially definiteness-based differential object marking to a situation where the GA, which can be hypothesised to have been the original definiteness marker, increasingly becomes the default marker for animate direct objects, and the NA, which can be hypothesised to have been the original indefiniteness marker, becomes increasingly specialised: first functionally in OCS, then constructionally in OES. Only the tail end of this development is attested, when the GA was already well on the way to being the default choice, and it is not possible to make proper sense of the OES data without taking the OCS data into consideration. Second, when the GA is fully generalised in the singular, comes a quick, sharp S-curved change in the plural, with no clear functional specialisation.

2 The animacy category

All modern Slavonic case languages have animacy-based subgenders, "an additional gender distinction within a minimal subset of the paradigm" (Baerman (1998: 187) following Corbett (1991: 163)). In Russian, for example, the masculine gender or agreement class is thus affected: if the noun in question is animate,[1] it will differ from the rest of the masculine agreement class in its

1 The terms 'animate' and 'animacy' are used in a strictly semantic sense throughout the article, for items that refer to humans or animals.

accusative singular and plural agreement. The feminine and neuter agreement classes are affected too, but only in the plural, where animate nouns again differ in their accusative agreement. The deviant forms are identical with the genitive (GA), but otherwise follow the usual pattern of their agreement class in the same way as inanimate nouns do, as seen in (1). Inanimate nouns have nominative-accusative syncretism instead (NA).

(1) a. *Ja videla malen'kij stol*
I saw small.NA table.NA
'I saw the small table'

b. *Ja videla malen'kogo mal'čika*
I saw small.GA boy.GA
'I saw the small boy'

c. *Ja videla malen'kie knigi*
I saw small.NA books.NA
'I saw the small books'

d. *Ja videla malen'kix devoček*
I saw small.GA girls.GA
'I saw the small girls'

While we can only determine a noun's (sub)gender on the basis of its agreement, gender and declension class are closely intertwined in Russian and the other Slavonic languages (cf. Corbett 1982), and the noun's case-number endings are normally affected too. (1 a) and (1 b) both have masculine first-declension[2] objects, but since *mal'čik* 'boy' is an animate noun, it gets the GA agreement and GA ending instead of the NA. Likewise, (1 c) and (1 d) both have feminine second-declension objects, but since *devuška* 'girl' is an animate noun, it gets the GA agreement and GA ending instead of the NA in the plural. In the singular the second declension has a dedicated accusative form (no nominative-accusative syncretism) that both nouns would get regardless of their semantic class, and likewise there is a dedicated, non-syncretic feminine accusative singular form in the adjective paradigm which is triggered instead of subgender agreement in such examples. However, masculine nouns control GA agreement in the accusative singular, regardless of their declension class, as seen in (2). Here *deduška* 'grandfather' is a second-declension animate masculine noun. Since it belongs to the second declension, it gets the

2 I follow Corbett (1982) in positing four noun declensions in modern Russian, where the first declension is the historical masculine ŏ-stem declension, the second declension is the historical ā-stem declension, the third declension the historical ĭ-stem declension and the fourth declension the historical neuter ŏ-declension.

dedicated accusative form, but nonetheless controls GA agreement in the adjective.

(2) *Petja vstretil starogo dedušku*
Petja met old.GA grandfather.ACC
'Petja met his old grandfather'

In modern Russian this is an entirely stable morphological phenomenon, there is practically no variation with individual lexemes or within semantic classes. It is driven by animacy alone.

The situation is the outcome of a development that started well before historical times, triggered by nominative-accusative merger in a number of Common Slavonic noun declensions. Old Church Slavonic provides the earliest attestation of the Slavonic animacy category. Here, we find the genitive-accusative (GA) in *variation* with the nominative-accusative (NA). Masculine singular ŏ-stem common nouns denoting male persons *may* get GA marking, as may masculine singular pronouns and nominalised adjectives and participles, but it is not obligatory, as illustrated in (3) and (4). At this point, therefore, the use of the GA is clearly not driven by animacy alone, and the subgenders cannot be deemed to be fully established until it is.

(3) *ašte žena puštьi mǫža posagnetъ za inъ*
if woman released man.GA marries after other.NA
'If a woman divorces her husband and marries another …'
(Mar. Mk. 10:12)

(4) *idi prizovi mǫžь tvoi*
go summon.2SG man.NA your.NA
'Go, summon your husband.' (Mar. Jh. 4:16)

It is clear that the development is relatively advanced already in OCS. The GA is more frequent than the NA, it is found both with direct objects and after accusative-governing prepositions, and GA agreement can be triggered by a controller that is itself not GA-marked (5), as it can in modern Russian (2).

(5) *prizъvavъ slugǫ svojego glagola jemu*
having-summoned servant.ACC REFL.POSS.GA said.3SG he.DAT
'having summoned his servant, he said to him'
(*Codex Suprasliensis* 8, L-61v 4)

Nonetheless, over time a number of studies (Meillet 1897, Comrie 1978, Huntley 1993, Eckhoff 2015) have pointed out that the distribution of NA and GA in OCS is not random, merely conditioned by morphological factors or due to differences in social prominence, but is conditioned by definiteness to some extent. A reasonable hypothesis therefore seems to be that the origins of the animacy category are to be found in differential object marking, i.e. a situation where "one set of direct objects is case marked in one way and another set in a different way depending on features of the object" (Malchukov & de Swart 2009: 345), and where this alternation does not depend on the verb, but on semantic and pragmatic features of the object (Aissen 2003: 435). The semantic and pragmatic properties in question are usually animacy, definiteness, and "other properties which are subject to discourse-related requirements, such as specificity and topicality" (Ledgeway, Schifano & Silvestri 2019: 1).

In early Slavonic it is clear that we are dealing with *discriminatory* DOM, triggered by sound changes that caused nominative-accusative syncretism in several noun declensions. At a very early point NA was thus the only option. The DOM is only found in declensions with such syncretism (see also Seržant 2019: 164–167). It was at first limited to animate singular and, as I shall argue, definite objects, since they shared so many of the typical properties of subjects. This view is not uncontroversial, most notably, Krys'ko (1994: 166) completely rejects the idea that the NA-GA variation was ever conditioned by definiteness or any other semantic considerations beyond animacy. I believe that Krysko's view cannot be maintained if the OCS[3] data is considered alongside the Old East Slavonic and Middle Russian data.[4]

3 Animate objects in the PROIEL and TOROT treebanks

The data for this article are all taken from the PROIEL[5] and TOROT[6] treebanks, both part of a larger family of treebanks for early attestations of Indo-European languages (Haug, Jøhndal, Eckhoff, Welo, Hertzenberg & Müth

3 OCS is, of course, not a direct ancestor of Old East Slavonic, as it has clear South Slavonic features and can be classified as a dialect of Old Bulgarian/Macedonian with some elements from other early Slavonic dialects. However, in this as in many other matters, it can reasonably be expected to be close to Late Common Slavonic and may be taken to represent an earlier state of affairs than that observed in early East Slavonic texts.

4 This is not to say that Krys'ko does not consider OCS data, but when he does perform counts (e.g Krys'ko 1994: 63–64), he tends to be dismissive of the statistical tendencies he finds, for further discussion see Živov (2017: 768–769).

5 Pragmatic Resources in Old Indo-European Languages, https://proiel.github.io, Haug & Jøhndal (2008)

6 Tromsø Old Russian and OCS Treebank, https://torottreebank.github.io, Eckhoff & Berdicevskis (2015)

2009). The PROIEL treebank is at its core a parallel treebank of early Indo-European languages, containing the New Testament in Greek, Latin, Gothic, Classical Armenian and OCS. It contains the *Codex Marianus* fully lemmatised, annotated for morphology and syntax, and aligned at token level with the Greek text.[7] The TOROT treebank is an expansion of the OCS part of the PROIEL treebank (containing about 150,000 words of further canonical and non-canonical texts), and also contains a new Old and Middle Russian treebank (about 250,000 words of texts of various types: chronicles, religious texts, legal texts, charters, birchbark letters, tales, personal correspondence etc.). The texts are all fully lemmatised and annotated for morphology and syntax (the lemmatisation and morphological annotation is automatic with manual correction, the syntax is fully manual).

The annotation scheme is a form of enriched dependency grammar, which deviates from classical dependency grammar (such as the Prague Dependency Treebank scheme) by using null verbs and conjunctions, secondary dependencies and a richer set of relation label tags. This makes the formalism more expressive (for example, it handles e.g. ellipsis and control very well), but also more complicated to parse and query. The treebank infrastructure also allows extra layers of annotation, notably for information status, anaphoric relations and semantic features such as animacy. Analyses generally exploit the interplay between part of speech, morphology and syntax. This is also the case for our annotation policy for animate direct objects.

There can be little doubt that the GA is an accusative marker morphologically even in the earliest texts, as GA agreement can be triggered by masculine singular human objects with a regular accusative marker even in the very earliest texts, as seen in (5). The most "correct" analysis might then perhaps have been to tag GAs as plain accusatives. This would, however, not have been good for retrievability, as it would have been impossible to distinguish NA-marked objects from GA-marked ones. Another option would have been to introduce a separate case tag for the GA, but this would have lead to a number of difficult annotation decisions, as there are several contexts where the GA is ambiguous: under negation, which triggers near-obligatory genitive of negation, after supines, which predominantly take genitive objects, and in contexts that trigger partitive genitives. We therefore opted for a simple, but obviously 'wrong' solution, namely to tag all genitive-looking items as genitives, even in cases where they were undoubtedly in the GA. The rationale was that this was the simplest solution both for automatic taggers and for human annotators, and that in any case there was no way of retrieving the relevant set

7 The Greek text is von Tischendorf (1889). It is clear that this is not the Greek source text used for the OCS translation, and this will necessarily cause some noise in the data.

Context	Specific tag	Non-specific tag	
Discourse	OLD	NONSPEC-OLD	previously mentioned
Scenario	ACC-INF	NONSPEC-INF	accessible by inference
Encyclopaedic	ACC-GEN		accessible from world knowledge
Situation	ACC-SIT		accessible by deixis
—	NEW	NONSPEC	not previously mentioned
	KIND		kind-referring
		QUANT	quantified

Table 1 Contexts and tags in the PROIEL corpus, adapted from Haug, Eckhoff & Welo (2014)

of NAs and GAs a) without animacy annotation of some sort, and b) without using the syntactic annotation to ascertain that the item in question was not negated, not dependent on a verb in the supine (which regularly take the genitive), and not on a verb that could occasionally govern the genitive.[8] Since the NA-GA variation persists, first in the singular, then in the plural, throughout the period covered by the texts represented in PROIEL and TOROT, this policy was maintained throughout. This conservative policy is in line with a number of decisions made in the PROIEL/TOROT annotation: the earliest stage of a development is generally privileged, and when a phenomenon undergoes change, the analysis is not changed unless strictly necessary. It also fits in with a general ethos of breaking down the analysis into components that are easier to agree on objectively, and then combine these components to retrieve the relevant data.

The PROIEL treebank makes rich use of the extra annotation layers. The OCS texts have full animacy annotation at lemma level,[9] and the Greek Gospels are fully annotated for information status and anaphoric relations. The annotation assigns information status tags to discourse referents (in the sense of Karttunen 1969). In addition, anaphoric links (from anaphors to antecedents) are assigned, in order to provide data on the properties of immediate antecedents and also on the length and density of anaphoric chains. An addressee-based notion of givenness is employed, based on the idea that the addressee will consult various contexts in order to establish the reference.

8 This is for example the case for several perception verbs. The solution was to tag all genitive dependents of such verbs with the supertag ARG, which was used when it was uncertain whether an item was a direct object (OBJ) or an oblique argument (OBL).

9 The tags are HUMAN, ORGANISATION, ANIMAL, CONCRETE, VEHICLE, NONCONCRETE, PLACE, TIME (Haug et al. 2009: 42–43).

The annotation scheme is built on a basic three-way distinction between OLD-ACCESSIBLE-NEW in the tradition after Prince (1981).[10] If something has been explicitly mentioned in the previous discourse, it is OLD.[11] The "accessible" category is divided into three by knowledge context: a discourse referent may be accessible by inference, world knowledge or deixis.[12] Finally, if a discourse referent is not available from any knowledge context, it is tagged NEW. In addition there are parallel tags for non-specific contexts, where direct and indirect anaphora are also relevant. These tags are used for discourse referents which only exist inside certain embeddings, such as negation, modality etc., what we may call short-term referents in Karttunen's (1969) terms. There is some limited scope for picking up referents within such embeddings, in examples such as *John wants to catch a* **fish** *and eat* **it** *for supper* (Karttunen 1969: example 25). Finally, there are separate tags for kind-referring and quantified[13] nominals.[14]

The scheme can be applied with sufficiently high interannotator agreement (Haug et al. 2014: 47–49) and allows quite sophisticated analysis of information structure on its own and in combination with other annotation layers. For this article it crucially allows a workable operationalisation of the notion of definiteness. For the purposes of this study I will take definiteness to be about identifiability, namely that the speaker directs the hearer to the referent of the noun phrase by signalling that s/he is in a position to identify it (Lyons 1999: 5–6).[15] Semantically, then, a noun phrase is definite if its referent is previously mentioned (OLD, NONSPEC-OLD, OLD-INACT), if it is accessible from inference, world knowledge or deixis (ACC-INF, ACC-GEN, ACC-SIT, NONSPEC-INF) or if it is kind-referring (KIND). If its referent is not previously mentioned (NEW, NONSPEC) it is semantically indefinite.[16]

10 For a fuller description and problematisation of the annotation scheme and its theoretical background, see Haug et al. (2014).

11 The tag was only used if the antecedent occurred within a 13-sentence window. If the antecedent was outside that window, the tag OLD-INACT (old inactive) was used. The two tags are collapsed in all statistics in this article.

12 The term is thus not used as it is in the framework of Ariel, e.g. Ariel (1988), who is primarily interested in the *degree* of accessibility, i.e. how easy it is for a hearer to access a particular referent.

13 The decision to consider QUANT a tag for non-specific discourse referents was a matter of convention. The organisation of tags in to specific/non-specific pairs forced a decision even though the choice is not obvious: Quantified NPs are usually treated as non-referential in semantic analyses, but it is clear that many of them just refer to plural discourse referents (Haug et al. 2014: 34).

14 In the following, the three accessible tags are collapsed into one, and so are OLD and OLD-INACT.

15 I acknowledge that there are some uses of definiteness markers that are better accounted for by the notion of uniqueness or inclusivity (Lyons 1999: 7–15), but I will not pursue that analysis in this article.

16 Quantified noun phrases are more difficult, but can often be seen as definite if a uniqueness

Only limited use has been made of the extra annotation layers in the TOROT, so the Old East Slavonic and Middle Russian treebanks have no animacy annotation and no full-scale information status annotation, and had to be handled differently, see Sections 5.1 and 6.1.

4 Snapshot 1: DOM in OCS

Ever since Meillet (1897) it has been suggested repeatedly in the literature that the NA expresses indefiniteness and the GA definiteness to some extent. In this section I will briefly recapitulate my analysis in Eckhoff (2015), where I found that the choice of NA and GA in OCS is indeed driven by information status and potentially also by other discourse factors.

For the study all human masculine singular objects and complements of accusative-governing prepositions were extracted from the *Codex Marianus* and the *Codex Zographensis*, two archaic gospel manuscripts that are part of the Old Church Slavonic text canon,[17] on the condition that they

- had a lemma that did not end in -a, which would in most cases mean that they belonged to the ŏ-stem declension (for nouns)
- were not in an obligatory or nearly obligatory genitive position (in the scope of a negation, dependent on a supine[18])
- were not governed by a verb that could possibly take genitive objects[19]

perspective is applied. However, there was only a single such example in the datasets used for this article, see Table 8.

17 At the time the TOROT also contained a large share of the *Codex Suprasliensis* (now complete), but this data was not included in the study since the share of NA was considerably lower than in the *Marianus* and the *Zographensis*, and since it was also clearly limited to only a few of the texts in the codex (Eckhoff 2015: 237). It would thus seem that the language of the *Suprasliensis* was innovative in this respect as in so many other respects. Similarly, the canonical OCS *Psalterium Sinaiticum* does not contain a single NA-marked human masculine singular ŏ-stem object. There may thus not be much scope for expanding the OCS dataset.

18 In the *Codex Marianus* and *Codex Zographensis* there is only a handful of examples of supines with accusative rather than genitive objects (in *Marianus*, for example, there are four examples of accusatives, but 23 examples of genitives). In comparison *Codex Suprasliensis* has 12 accusatives and 25 genitives, so again it seems that this later manuscript is linguistically innovative.

19 This was operationalised by extracting only items with the OBJ relation label, which was reserved for a) NA-marked objects, b) objects marked by non-syncretic accusative forms, c) objects of reliably transitive verbs with face-value genitive objects (whether GA, genitive of negation, partitive genitive or genitive due to a supine). If the verb was not deemed strictly transitive, but able to take either accusative and genitive objects, the genitive-marked argument was tagged ARG. Note that this means that the query did not rule out partitive genitive objects in principle, but in practice human singular objects are never partitive objects. To extract

- had a Greek token alignment which was tagged for information status

NA and GA were found to be unevenly distributed across lemmas – only 13 out of 48 object lemmas were attested in actual variation in the dataset (Table 2), while 9 were attested in the NA only and 26 in the GA only (Eckhoff 2015: 241–243). In most cases there were very few occurrences of each lexeme, which makes it difficult to determine whether this distribution was down to pure chance, or of there is some lexical differentiation in the material, for example along the lines of social prominence, which is the usual textbook explanation (see e.g. Lunt (2001: 56)). To avoid any possible bias from such lexical differentiation, all statistics are given for the full dataset as well as for a dataset consisting of the lemmas that were attested in actual variation.[20]

	NA	GA
bogъ 'god'	2	40
gluxъ 'deaf'	1	1
gospodinъ 'master'	1	3
gǫgnъnivъ 'having a speech impediment'	1	1
drugъ 'other'	4	4
mǫžь 'man'	1	3
oslabiti 'weaken'	3	1
pastyrь 'shepherd'	1	3
rabъ 'servant, slave'	12	14
razboinikъ 'robber'	1	4
synъ 'son'	26	32
cěsarь 'emperor, king'	2	6
člověkъ 'man'	3	28

Table 2 Object lemmas attested in variation in OCS dataset (data from Eckhoff (2015))

The OCS data (Table 3) show that there is indeed a correlation between information status and object case choice.[21]

complements of prepositions, I queried for dependents of prepositions from a list of known accusative-taking prepositions, tagged with the relation label COMP.

20 The lemma *bogъ* 'God' was excluded from the variation dataset since it had only two somewhat dubious NA occurrences and had such high frequency that it was likely to skew the results, see Eckhoff (2015: 241).

21 Meillet (1897) suggests that the GA first appeared in participle constructions. If we look at occurrences with participle heads (42 occurrences in the dataset reflected in Table 3), we find that the GA is possibly overrepresented, as there are only four examples with NA. Interestingly, those four occurrences have an information status that fits the overall trend: one is new, one

	NA	GA	NA	GA
new	27	21	14	13
anchored	24	32	14	4
accessible	3	65	3	11
old	19	132	13	57
non-specific new	15	17	7	6
non-specific inferred	5	15	5	5
non-specific old	0	8	0	2

Table 3 Information status by OCS case, full set vs. variation set, data from Eckhoff (2015)

Old and accessible objects have a strong tendency to be GA-marked. New and anchored objects, on the other hand, show variation: they are about as likely to occur in the NA as in the GA. A new referent is one that has not been previously mentioned (6, 7).[22]

(6) a. *privěsę* *člvkъ* *němъ* *běsenъ*
brought.3PL man.NA.NEW mute possessed-by-demon

b. *prosēnegkan* *autōi* *anthrōpon* *kōphon* *daimonizomenon*
brought.3PL they man.ACC.NEW mute possessed-by-demon
'they brought a man, mute and possessed by a demon'
(Mar. Mt 9:32)

(7) a. *i* *se* *mǫži* *na* *odrě* *nosęšte* *člvka*
and lo men on stretcher carrying man.GA.NEW

b. *kai* *idou* *prosepheron* *autōi* *paralutikon* *epi* *klinēs*
and lo carried.3PL they paralytic.ACC.NEW on stretcher
'And behold, some men were carrying a man on a stretcher'
(Mar. Lk 5.18)

An anchored referent, on the other hand, is a referent that has not itself been previously mentioned, but has a syntactic dependent which is not new, and

is non-specific new and two are anchored.

22 All OCS examples (a) are given with the Greek parallel text (b).

which 'anchors' it to the context (8, 9). In languages with definite articles, such 'anchored' referents are seen to group either with new referents (no article, as in English *my son*) or with old and accessible referents (with article, as in Norwegian *sønn-en min* 'son.DEF my'). In the OCS dataset we see that they group with new referents more often than not. In Greek anchored referents strongly tend to have definite articles, and anchored referents are in fact responsible for a large share of the 'mismatch' examples where Greek has a definite article, but the OCS translation has NA (Eckhoff 2015: 246).

(8) (A man was having a big banquet and invited many)

a. *i* *posla* *rabъ* *svoi* *vъ* *godъ* *večerě* *rešti*
and sent.3SG servant.NA his on time banquet.GEN say
zъvanymъ
invited.DAT

b. *kai* *apesteilen* *ton* *doulon* *autou* *tēi* *hōra* *tou*
and sent.3SG the servant.ACC his the time the
deipnou *eipein* *tois* *keklēmenois*
banquet.GEN say the invited.DAT
'and he sent his servant on the day of the party to tell the invited'
(Mar. Lk 14:17)

(9) a. *edinъ* *že* *otъ* *stojęštixъ* [...] *udari* *raba* *arxiereova*
one PTC of standing struck.3SG servant.GA archpriest's

b. *heis* *de* *tis* *tōn* *parestēkotōn* [...] *epaisen* *ton*
one PTC someone the standing struck.3SG the
doulon *tou* *arkhiereōs*
servant.ACC the archpriest.GEN
'one of those standing there [...] struck the archpriest's servant'
(Mar. Mk 14:47)

The observation that the NA is more common with objects with the reflexive possessive pronoun *svoi* is an old one, hailing back to Tomson (1908/2006: 40). Tomson argued that since the reflexive possessive pronoun effectively disambiguates the NA form (it cannot be the subject because the reflexive demonstrates that there is a different subject), the GA is not needed. But as he also notes, the situation in attested OCS has developed well beyond such a (hypothetical) state, so that the GA was found as well. However, it is important to note that the most common anchor is *not* the reflexive possessive pronoun (5 occurrences), but the first-person possessive pronoun *moi*

(13 occurrences), which does not disambiguate the NA in the way indicated by Tomson (1908/2006). It therefore seems clear that the figures we see for anchored objects are not down to a lack of need to discriminate.

There is no discernible difference between specific and non-specific objects, as far as we can tell by the very small number of non-specific examples: new non-specific objects display NA-GA variation (10,11), while old and accessible non-specific objects are mostly GA.

(10) a. *jako* *na* *razboinikъ* *li* *pridete* *sъ* *orǫžьemь*
as against robber.NA PTC arrived.2PL with weapon.INSTR

b. *hōs* *epi* *lēistēn* *exelēluthate* *meta* *maxairōn*
as against robber.ACC have-come-out.2PL with swords.GEN
'Have you come as if against a robber, with weapons?'
(Mar. Lk 22:52)

(11) a. *jako* *na* *razboinika* *li* *izidete* *sъ* *orǫžьemь*
as against robber.GA PTC came-out.2PL with weapon.INSTR

b. *hōs* *epi* *lēistēn* *exēlthate* *meta* *maxairōn*
as against robber.ACC came-out.2PL with swords.GEN
'Have you come out as if against a robber, with weapons?'
(Mar. Mk 14:48)

This two-way separation between new and anchored on the one hand and old and accessible on the other hand is statistically significant, both in the full set and in the variation set (Table 3). If we take 'definite' to mean old or accessible, and 'indefinite' to mean new (regardless of specificity or anchoring), then OCS GA marking looks strongly like a limited kind of definiteness marking.

The fact remains that only about half of the indefinite human singular objects (after this definition) are NA-marked. However, there are signs of specialisation in the OCS data: in a dataset of new, non-specific new and anchored objects, NA objects were found to be much more likely to be picked up (Table 4, Eckhoff (2015: 249)) than GA objects.

It is possible, then, that the NA had specialised as a marker of referential persistence: it could be used to indicate the introduction of an important participant who would be mentioned again multiple times, as in (12) and (13).

	full set	variation set
NA	6.30	7.29
GA	2.43	3.04

Table 4 Mean number of pickups by object case, new, non-specific new and anchored objects, p-value=0.00025 in variation set

(12) Parable of the wicked tenant (8 further mentions)

a. *posledь že posъla kъ nimъ synъ svoi*
after PTC sent.3SG to them son.NA his

b. *husteron de apesteilen pros autous ton huion autou*
after PTC sent.3SG to them the son.ACC his
'Finally, he sent his son to them' (Mar. Mt 21.37)

(13) Healing of the paralytic (11 further mentions)

a. *i se priněsę emu oslablenъ žilami na odrě ležęštъ*
and lo brought.3PL him weakened.NA sinews on stretcher lying

b. *kai idou prosepheron autōi paralutikon epi klinēs beblēmenon*
and lo brought.3PL him paralytic.ACC on stretcher lying
'and, behold, they brought him a paralytic lying on a stretcher' (Mar. Mt. 9.2)

If this was not important, you could introduce your referent in the GA, as in (14), where Simon of Cyrene is introduced, but has no storyline beyond this verse: we are only told that he is forced to carry the cross.[23]

23 This constitutes one further mention in the OCS text. In the Greek text (on which the statistics are based) he technically has two further mentions because the verb *arēi* is taken to have a null subject coreferent with Simon.

(14) a. *isxodęše že obrětǫ čka kuriněiska. imenemь*
going-out PTC found.3PL man.GA Cyrenean.GA name.INST
simona. semu zaděšę ponesti krstъ ego
simon.GA. this.DAT forced.3PL carry cross his

b. *Exerkhomenoi de heuron anthrōpon Kurēnaion,*
going-out PTC found.3PL man.ACC Cyrenean.ACC
onomati Simona. touton ēggareusan hina arēi ton
name.DAT Simon.ACC that.ACC forced.3PL so-that took.3SG the
stauron autou.
cross his

'As they went out, they found a man of Cyrene, Simon by name. They compelled this man to carry his cross.' (Mar. Matt. 27:32)

However, although this is a statistically significant tendency, there are counterexamples too. For example, in the Mar. Lk. 5:17–26 version of the healing of the paralytic, the paralytic is introduced in the GA (7) even though he has 20 further mentions, unlike in the Matt. 9:1–8 version cited in (13). We are thus not able to account for all the variation in the material.

To sum up, it seems indisputable that the distribution of NA and GA in our OCS dataset is sensitive to information status and can be understood as a type of definiteness marking. It is also clear that the GA is increasingly the default marker, and that it can to some extent be used for referents that have not previously been mentioned. The NA may have developed into a specialised marker for referentially persistent new participants, but this observation is not enough to account for all the variation in the dataset – perhaps due to the default status of the GA, perhaps due to the nature of the texts, which are handed down to us in copies that are about two centuries later than the date of composition.[24]

5 Snapshot 2: DOM in Old East Slavonic

In Old East Slavonic, we still find NA-GA variation, but it is much less obvious that the variation is conditioned by information status.

While Krys'ko (1994: 166) completely rejects the idea that the NA-GA variation was conditiond by definiteness or any other semantic considerations at any stage, there are several post-Krys'ko voices suggesting that such conditioning still exists in early Old East Slavonic.[25] Timberlake (1997) , ex-

24 We shall see that we have the same problem with the Old East Slavonic texts.

25 For a thorough critique of Krys'ko's views, see Živov (2017: 764–783), who largely supports Timberlake's view that the use of the NA is a matter of low individuation.

amining the *Primary Chronicle*[26] only, suggests that the NA is reserved for a few linguistic templates, which all have in common that the object referent is characterised by his relationship to the subject, and where other properties of this referent are irrelevant, as it does not have participant focus. The core template is the verb *posaditi* 'place' or *posъlati* 'send' + the NA-marked NP *mužь svoi* 'his man' (Timberlake 1997: 51), as in (15). He argues that the template is extendable with a few other nouns (*synъ* 'son', *bratъ* 'brother'). He is thus essentially taking a construction grammar approach to the problem, explaining the remnant NA usage in terms of partially lexically specific constructions, which are still weakly productive.

(15) *i* *prija* *gradъ.* *i* *posadi* *mužь* *svoi*
and took.3SG city and placed.3SG man.NA his
'and he took the city and placed his man there' (*Primary Chronicle* 23.2)

Timberlake convincingly argues that the only relevant property of the object referent in this example is its relationship with the subject: he serves as his representative. This is often underscored by a possessive pronoun, as it is in this example. At the same time, Timberlake says, the GA is clearly the default object case for animate masculine singular direct objects, and can be found in any context.

Another semantic-pragmatic interpretation of OES data can be found in Bratishenko (2003), who suggests that the choice between NA and GA for direct objects and possessive adjective and genitive for possessors are governed by a combined prominence hierarchy where lexical semantics (essentially social prominence), morphological and referential features (definiteness) interact. Intriguingly, she goes as far as to suggest that the GA arose as a way of resolving subject-object ambiguity precisely because the genitive was already being used for untypical, object-like 'possessors' in the adnominal domain. However, in none of these studies has systematic information status annotation been attempted on a larger material, and this is the task I will now turn to.

5.1 Texts, annotation and data extraction

The data was extracted from TOROT, limited to Old East Slavonic texts which displayed robust variation between animate NA and GA objects. This restriction leaves us with only three texts: the legal codex *Russkaja Pravda* (extended edition, compiled in the 12th century, Troitsa manuscript 14th century), and

26 See Section 5.1 for description.

two major Old East Slavonic chronicles: the *Primary Chronicle* (Laurentian manuscript, 1377) and the *Novgorod First Chronicle* (Synodal manuscript, 1260s). Both chronicles were compiled around 1100, but the gap between date of composition and manuscript is much greater for the *Primary Chronicle*.

In the majority of other texts belonging to the OES period, there is no evidence of animate NA objects at all. This is the case for a number of shorter texts, including all of the charters and treaties represented, as well as some longer texts such as the *Tale of Igor's Campaign*. Three longer text sources did have animate NA objects, but were excluded because the distribution seemed very skewed. The excerpts from the *Uspenskij sbornik* (*The Tale of Boris and Gleb*, *The Life of Feodosij Pečerskij*, late 12th/early 13th century manuscript) contained only two animate NA objects in 24,819 words. The texts are written in very high-style Russian Church Slavonic, and it seems tempting to assume that a strong preference for GA was included in that norm.[27] The *Suzdal Chronicle*, also found in the 1377 Laurentian manuscript, but compiled in the early 14th century, displays a distribution that is clearly different from that found in the preceding *Primary Chronicle*: in the 23,345 word excerpt included in TOROT, there were only four NA objects, all of them examples of the lemma *synъ* 'son', limited to only two entries (for the years 6627/1119 and 6648/1140), and it seems reasonable to view this as evidence of a diachronic development. Finally, the collection of birchbark letters (BBLs) included in TOROT poses a different problem: we find 11 animate NA objects in the 1869-word collection, but only three GAs – two of them are from the very early BBL 247, and somewhat dubious, while the third example is from the 14th-century BBL 370. Thus, our small sample (about 10%) of BBLs gives us no clear evidence of variation at all, but this time in favour of the NAs. This is expected, since Old Novgorodian did not have nominative-accusative syncretism in the ŏ-stems – the distinct Novgorod nominative in -e (Zaliznjak 2004: 99–102) prevented this. Zaliznjak (2004: 105), working on the full set of birchbark letters available at the time, observes that there is variation, but that the use of GA was very limited in Early Old Novgorodian. It therefore seems fair to exclude the TOROT sample of birchbark letters from consideration. The *Novgorod First Chronicle* is of course also a Novgorodian text, but was retained since the NA/GA distribution was quite similar to that seen in the other two selected texts.

Even in the three selected texts, there is the possibility that the time gap between time of composition and time of manuscript may influence the choice of object case. We might expect to see more NAs in the *Novgorod First Chroni-*

27 At least, as Živov (2017: 773, 780) points out, the NA never becomes a marker of high literary style in the singular, as it does in the plural.

cle than in the *Russkaja Pravda* and the *Primary Chronicle*, since the manuscript is considerably earlier than the other two, and could also be affected by the late expansion of the GA in Old Novgorodian. As we see in Table 5, this is not the case. The *Russkaja Pravda* has the highest proportion of NA objects, but the numbers are too low for the difference between the *Russkaja Pravda* and the *Novgorod First Chronicle* to be statistically significant. However, there are significantly more NAs in the *Novgorod First Chronicle* than in the *Primary Chronicle* (p=0.01667, Fisher's exact test). We may also note that the proportion of NAs in the *Russkaja Pravda* and the *Novgorod First Chronicle* is roughly the same as in the OCS dataset, but the proportion in the *Primary Chronicle* is clearly lower. The overall number of animate masculine singular human objects is also lower – not unexpectedly, since the OES dataset is smaller.[28] This suggests that the dataset is fairly representative of the situation in early OES. We can expect to find a richer system than that observed by Timberlake (1997), since he only included the *Primary Chronicle*.

	NA	per cent	GA	per cent
Russkaja pravda	10	34.5	19	65.5
Novgorod First Chronicle	27	20.5	105	79.5
Primary Chronicle	22	10.5	188	89.5
OCS dataset	93	24.3	290	75.7

Table 5 Case by source, masculine human direct objects, ŏ-stems, jŏ-stems and *synъ* 'son'

While the OCS treebank data had full animacy annotation and was aligned at token level with a Greek source text fully annotated for information status and anaphoric relations, the OES texts had only lemmatisation, morphological annotation and syntactic annotation in the outset. For the purposes of this article, all masculine singular animate direct objects were spot-annotated for information status and anaphoric relations. In cases where the direct object belonged to an anaphoric chain, the chain was not always traced back in full, but the immediate antecedent was always identified. Note that complements of prepositions were not included in the spot annotation, but they are in the OCS dataset.

Since there was no animacy annotation, GA and NA objects had to be extracted separately. In one query, all masculine genitive singular direct objects (i.e. bearing the relation tag OBJ) which did not have a lemma ending in *-a*,

28 *Russkaja Pravda*: 4015 words, *Novgorod First Chronicle*: 27,068 words, *Primary Chronicle*: 55,414 words, *Codex Marianus*: 57,577 words, *Codex Zographensis*: 52,184 words.

-ja,[29] and did not depend on a negated verb or a supine were extracted. All lemmas in this dataset were then manually inspected, and a list of animate lemmas was created. The list was supplied with a list of animate nouns that did appear in the NA, but not the GA. On the basis of this list of lemmas, a full set of NA objects could be extracted: all masculine singular accusative direct objects with a lemma belonging to the list of animate lemmas were selected.[30] Finally, the GA data were filtered, excluding all lemmas that were not on the lemma list used to extract NA objects, in the R script that handles all the statistical modelling of the data.[31]

In Table 5 only human objects are included. There is, however, a case for including nouns denoting animals as well, and this data was collected.

	NA	GA
Russkaja pravda	5	0
Novgorod First Chronicle	1	0
Primary Chronicle	5	6

Table 6 Masculine singular direct objects denoting animals, case by source

Table 6 shows that the number of such examples is very low overall, and that only the *Primary Chronicle* has evidence of variation – the lemmas *bykъ* 'bull', *zmii* 'snake' and *konь* 'horse' all have both NA and GA occurrences. All the occurrences of *bykъ* 'bull' belong to a single storyline, where the bull is introduced in the NA (16), and then picked up three more times in the GA, as in (17), see also Timberlake (1997: 51). This is slender evidence, but suggests that the system of the text is at a stage where the NA/GA variation has spread to nouns denoting animals, and where the distribution could be governed by information status: NA for new referents,[32] GA for old – which is what Tim-

29 Note that nominalised adjectives and participles were also included. For nouns, this in practice limits the dataset to ŏ- and jŏ-stems and the noun *synъ* 'son'. The latter is historically a ŭ-stem, but consistently behaves like an ŏ-stem in this respect.

30 Note that restricting the query to nominals with the syntactic relation tag OBJ excludes any dependents of perception verbs such as *viděti* 'see', *slyšati* 'hear', as these are known to fluctuate between taking direct accusative objects and genitive obliques. All genitive-shaped animate objects/oblique arguments of such verbs are thus ambiguous, since we cannot tell whether they are genitive obliques or GA direct objects. They were therefore tagged ARG, the supertag used when the annotator could not decide whether a dependent was a direct object or an oblique argument. As it happens, there were no NA objects headed by such verbs in the OES sources.

31 All datasets and scripts, including the full list of animate lemmas, can be found at https://doi.org/10.18710/8J6V1D.

32 Note that if NA is an indefiniteness marker in (16), the indefiniteness is doubly marked, since the attributive short-form adjectives *velikъ* 'large' and *silenъ* 'strong' also mark indefiniteness.

berlake (1997: 54) suggests.

(16) *i* *nalězoša* *bykъ* *velikъ* *i* *silenъ*
and found.3PL bull.NA large.NA and strong.NA
'And they found a large and strong bull' (*Primary Chronicle* 123.9)

(17) *i* *povelě* *razdraždіti* *byka*
and ordered.3SG irritate bull.GA
'and he ordered (them) to irritate the bull' (*Primary Chronicle* 123.9–10)

The animal examples will not be discussed further, and are excluded from the data in the rest of this section (but see Klenin (1983: 92–99) for further discussion). The obtained datset is not large and allows only limited generalisations, but seems to be sufficient to point out some systematic differences from the situation in OCS.

5.2 Variation at lemma level

As in the OCS dataset we observe that only nine out of the 87 lemmas in the dataset are attested in variation, see Table 7, and it is worth raising the social-prominence hypothesis again. Is the NA more likely to be used with low-status persons? Živov (2017: 769–770), leaning on Zaliznjak's (2004: 105–107) results from the birchbark letters, argues that there may be a case for claiming that "social'naja aktivnost", which we may translate as "level of social activation", plays a role in OES in general.

	NA	GA
bratъ 'brother'	3	35
episkopъ 'bishop'	1	5
kъnjazъ 'prince'	1	47
mužь 'man'	13	18
otrokъ 'servant, young man'	3	1
popъ 'priest'	1	3
synъ 'son'	18	33
xlopъ 'serf'	4	2
čeljadinъ 'servant'	3	2

Table 7 OES lemmas in variation

As we see in Table 7, the list does contain some low-status persons, such as *xlopъ* 'serf' and *čeljadinъ* 'servant', but also high-status persons such as

*episkop*ъ 'bishop' and *kъnjaz*ь 'prince'. The latter has just one NA example pitted against 47 GAs (18).

(18) *a* *vъskormili* *jesmy* *sobě* *knjaz*ь
but raised are self.DAT prince.NA
'but we have raised a prince for ourselves' (*Primary Chronicle* 123.9)

The most robustly attested variation is found with the socially 'neutral' nouns *mužь* 'man' and *syn*ъ 'son'.

There are also nine lemmas that are attested exclusively in the NA in our dataset: *vidok*ъ 'witness', *dětišč*ь 'child', *koньčanin*ъ 'inhabitant of the outskirts of town', *ljudin*ъ 'man', *mladenьc*ь 'infant', *prezviter*ъ 'priest', *svojak*ъ 'brother-in-law', *sъl*ъ 'envoy', *šurin*ъ 'brother-in-law'. It should be noted that none of them occur more than twice in the dataset, so we have no evidence that the GA is not an option for these lemmas. We can note that the set contains two nouns denoting young children, who are clearly low-status, but also the presumably high-status *prezviter*ъ 'priest', and a number of socially neutral nouns.

The remaining 69 lemmas are only found in the GA. Again, the majority of them are attested just a few times and give us no real evidence as to whether NA was an option. Only seven lemmas are attested more than four times, they are (in order of frequency) *bog*ъ 'God' (37), *člověk*ъ 'man' (13), *posadьnik*ъ 'governor, posadnik' (10), *cěsar*ь 'emperor, king' (8), *igumen*ъ 'hegumen' (7), *mitropolit*ъ 'metropolite' (6), *otьc*ь 'father' (5). This list seems to support the social-prominence or social-activation hypothesis to some extent, since all these lemmas except *člověk*ъ 'man' refer to high-ranking male persons. However, one should also note that many of them will be uniquely identifiable in most contexts, which would result in GA on a discourse-prominence approach too. In particular, *bog*ъ 'God' behaves as a proper noun in most contexts, and as in OCS, all proper nouns occur in the GA.

All in all, the situation is fairly similar to what we saw in OCS, though the evidence in favour of a social-prominence account is perhaps slightly stronger. In the following, statistics for the full data set ("all"), the set of lemmas attested in the NA ("NA set") and the set of lemmas in attested variation ("variation") will be given throughout (see Table 8).

5.3 Information status analysis

When we break down our OES data in the same information status categories as in the OCS dataset (Table 8), we see that the GA outmatches the NA in every category. There is thus much less evidence to suggest that the variation is driven by information status – recall that in OCS (Table 3) there was a clear

split between new and anchored on the one hand (about 50/50 NA/GA), and old and accessible on the other hand (strong preference for GA). In the OES dataset new, specific human objects seem to pattern with old and accessible ones – there are only four examples of NA overall, and only one disputed example of *mužъ* 'man' in the variation set (19). The same sentence also contains one of the three invariant NA new specific examples (*koncjanъ* 'inhabitant of outskirts of town'), which is also disputed.

(19) *i* *ubiša* *muž̲* *prus̲* *a* *koncjanъ*
and killed.3PL man.NA? Prussian.NA? and outskirt-inhabitant.NA?
drugyi
other.NA
'and they killed a Prussian man and another man from the outskirts of town' (*Novgorod First Chronicle*, year entry for 6726/1218)

	all		NA set		variation	
	NA	GA	NA	GA	NA	GA
new	4	22	4	8	1	8
anchored	25	62	25	49	20	49
accessible	6(1)	64(2)	6(1)	19	6(1)	19
old	10(7)	89(18)	10(7)	44(16)	9(7)	44(16)
kind	0	15	0	0	0	0
non-spec. new	11(4)	48(4)	11(4)	21(4)	8(4)	21(4)
non-spec. inferred	0	4	0	1	0	1
non-spec. old	3(1)	7(1)	3(1)	4	3(1)	4
quantified	0	1	0	0	0	0

Table 8 Information status of OES masculine singular human direct objects, all objects vs. objects that occur in NA vs. objects that are attested in variation. Numbers in parenthesis indicate how many of the occurrences in a cell have a possessive pronoun dependent. For an explanation of the three datsets, see Section 5.2

While the NA interpretation of the variable lemma *muž̲* 'man' is certainly possible, the underlining in the transcription indicates an abbreviation mark, which means that the word-final character has been dropped. The character may be either <ъ> (indicating NA) or <a> (indicating GA), but the coordinated *koncjanъ* would suggest that it is NA (see Michell & Forbes (1914: 60) for this interpretation). Krys'ko (2006: 184), however, takes *koncjanъ* to be

genitive plural, which is syncretic with the NA, which leaves the interpretation of *mužь* entirely open.[33]

The remaining two invariant NA new specific examples are found in (20) and (21).

(20) (And Olga said: Do not perform a pagan burial ritual over me.)

bě *bo* *imušči* *prezvuterъ*
was.3sg for having priest.NA
'for she had a priest (with her)' (*Primary Chronicle* 68.4–68.6)

(21) *žena* *dětišťъ* *rodi* *bezъ* *očьju.* *i* *bez* *ruku*
woman child.NA bore.3sg without eyes and without hands
'a woman gave birth to a child without eyes and without hands'
(*Primary Chronicle* 165.12–165.13)

Timberlake (1997: 52) discusses example (20) and fits it into his templates by pointing out that *iměti* 'have' is an existential verb, and that the object referent is mentioned exclusively in virtue of being a priest, his other properties are irrelevant. In the PROIEL-style information status annotation he must nonetheless be tagged new, since it is clear from the progressive construction *bě imušči* 'was having' that we are talking about the specific priest who is with her at the moment. The information status annotation thus does not pick up that particular nuance of meaning.

Example (21) has a lemma meaning 'infant', and Timberlake (1997: 54) uses the example to argue that words denoting children might have had a clear NA-GA distribution based on first and subsequent mentions,[34] as he also speculates that animal nouns may have (see examples 16–17). However, the evidence is next to non-existent and cannot really be used to substantiate such claims.

No matter how we interpret these examples, it is clear that new specific masculine singular human objects are very rarely found in the NA in OES (although admittedly they are rare overall).

33 Krys'ko translates the sentence 'They killed one man of the Prussians and another of the outskirt inhabitants', interpreting the example to contain two partitive genitive plurals.

34 He uses the following example from the *Primary Chronicle* 164.14–164.15 as an example of a second mention: (A child was thrown into the river Setoml'.) *jego.*ga *že dětišča.*ga *vyvolokoša rybolove vъ nevodě* 'Him, the child, fishermen pulled out in their net' Note that this example is not picked up in our query because the noun was considered to be an apposition on the personal pronoun *jego*, which then determines the case for the whole phrase. The distribution of NA and GA with personal pronouns is quite different from that in nouns.

Interestingly, we see that the 50/50 split with anchored objects remains. Recall that an anchored referent is a specific referent which has not itself been previously mentioned, but has a dependent which has an information status that is not NEW or NONSPEC(ific) – i.e. the referent is already known or accessible. This is illustrated in (22), which is the opening sentence of a story. The servant has not been previously mentioned, but the referent of the reflexive possessive pronoun *svoi* has, as this is the speaker himself.

(22) *poslax* *otrokъ* *svoi* *v* *pečeru*
sent.1SG servant.NA REFL.POSS.NA in Pečera.ACC
'I sent my servant to the Pečera' (*Primary Chronicle* 234.25–234.26)

As already mentioned, in languages with definite articles, we see that the marking of these phrases tend to pattern either with new referents (no article, as in English *my servant*) or with old/accessible referents (with article, as in Norwegian *tjener-en min* 'servant.DEF my'). The observed pattern is therefore not unexpected. We may also note that this group contains a good number of Timberlake's core template examples. Their anchoring referent is consistently in the form of a pronominal possessive (19 *svoi*, as seen in (22), three *moi* 'my' and three *ego* 'his' (26). The object lemma is predominantly *mužь* 'man' (7 occurrences) and *syn* 'son' (8 occurrences), but there are also *brat* 'brother' (3 occurrences), *otrokъ* 'servant' (2 occurrences, see example (22)), *svojakъ* 'brother-in-law' (3 occurrences), *sъlъ* 'envoy' (2 occurrences) and *šurinъ* 'brother-in-law' (1 occurrence), i.e. a wider range of nouns than the ones assigned by Timberlake to his template. However, all of them are fully compatible with the meaning of the template, since they are all either kinship terms or terms for servants or employees, which means that they all lend themselves to an interpretation where they are not deployed fully as individuals, but merely as an extension of the properties of the subject. We also see that the most frequent verbs are indeed *posaditi* 'place' (5 occurrences) and *posъlati* (7 occurrences), as well as the related *prisъlati* 'send' (3 occurrences), again in line with Timberlake. However, other verbs are found too, such as the similar *vъdati* 'give' (2 occurrences), *dajati* 'give' (1 occurrence), *postaviti* 'place' (1 occurrence), but also the less similar *pojati* 'take' (1 occurrence), *pustiti* 'let go' (2 occurrences), *slěpiti* 'blind' (1 occurrence) and *ubiti* 'kill' (3 occurrences).

Quite a few of the examples, such as (15), have both the lexical items and semantics that Timberlake specifies for his template. However, we also see that there are a number of examples with the right semantics, but different lexical material (23), as well as examples that have both different lexical material and different semantics.

(23) (And they sent a messenger to Vsevolod to ask for a prince)
i *vъda* *imъ* *svojakъ* *svoi*
and gave them brother-in-law.NA REFL.POSS
'and he gave them his brother-in-law'
(*Novgorod First Chronicle*, year entry for 6689/1181)

One example of that is (22), where the verb is the canonical *posъlati* 'send', but the object noun is *otrokъ* 'servant'. Semantically, it is also somewhat deviant, as the narrator goes on to tell a relatively long story about how his servant fared among the Pečera (though it is possible to argue that he was only his master's eyes and ears).

The example that Timberlake uses in his article title (24) is another fairly deviant example.

(24) *čemu* *jesi* *slěpilъ* *brat* *svoi*
why are.2SG blinded brother.NA REFL.POSS.NA
'Why did you blind your brother?' (*Primary Chronicle* 263.1)

Timberlake (1997: 56) argues that the question has a strong existential flavour. While the referent of *brat svoi* 'your brother' is a specific person, Vasilko, and a part of a long story about him, the choice of that particular NP was made to indicate that it is a particularly heinous crime to blind someone if he is your *brother*. Thus, he sees this "as an occasional, but not wholly arbitrary, extension of the prototypical formula for the nominative-accusative" (Timberlake 1997: 56).

More radically deviant examples are found in (25) and (26), both with the verb *ubiti* 'kill'.

(25) (The Novgorodians began talking about war with Suzdal)
i *ubiša* *mužь* *svoi* *i* *sъvьrgoša* *i* *sъ*
and killed.3PL man.NA REFL.POSS.NA and threw.3PL him from
mosta *vъ* *subotu* *pjantikostьnuju*
bridge on Saturday of-Pentecost
'and they killed one of their own men and threw him from the bridge on Pentecost Saturday' (*Novgorod First Chronicle*, year entry for 6642/1134)

(26) *i* *ubiša* *domažira* *tъrlinicja* *i* *sъnъ* *jego*
and killed.3PL Domažir.GA Torlinič and son.NA his
'and they killed Domažir Torlinič and his son'
(Novgorod First Chronicle, year entry for 6732/1224)

In (25) it is possible to take a Timberlake-style line of reasoning – the NA was chosen because killing someone is particularly objectionable if it is one of your own. However, it does not seem likely that the NA was chosen in (26) to indicate that the reader should react in a particular way to the son being killed *because* Domažir was his father, which I think would be the outcome of Timberlake's semantics.

Instead, I would like to point out that what these 25 examples all have in common is their status as anchored – they all have a new object referent, but with a dependent possessive pronoun whose referent is already known. Therefore, while I think Timberlake is right in assuming that we are dealing with a template, or in other terms with a partially lexically specific construction which is able to preserve the NA after the GA has become the default in almost all contexts, I think his template is too specific, at least when we take data from more sources than the *Primary Chronicle* into consideration. There is a case for claiming that the template simply consists of an unspecified transitive verb, a common masculine singular noun denoting a human being, and a possessive pronoun or a personal pronoun in the genitive, and that it is not restricted to a non-individuating semantics, even though that type of semantics is common. In fact, the job of the possessive pronoun is to draw attention to the relationship between (usually) the subject and object, often at the expense of other properties. Moreover, there is a strong case for arguing that this syntactic template or partially specific construction has its direct origins in the information-status-driven distribution we observed in the OCS dataset: the licence to use the NA seems to have been transferred from a particular information status configuration (a NEW referent with an OLD dependent) to a syntactic template containing a particular type of dependent.

More evidence of this can be found if we look into other information status categories. Recall that only new referents were deemed to be anchored if they had a non-new dependent. However, referents of other information statuses can have such dependents as well. In Table 8 objects with anchors, where the anchors are operationalised as possessive pronouns or genitive personal pronouns, are indicated in parentheses.[35] It is unexpected on the basis of the OCS figures that there should be more OLD than NEW NA objects in the OES material, but this is what Table 8 tells us. However, as many as seven out of ten of these objects actually have an anchor, as seen in (27) and (28).

35 They are of course also included in the full number: 6 (1) means that there are six occurrences of a certain type, one of which has an anchor.

(27) (The Pechenegs had a great warrior ready, but the Rus had to search a long time to find theirs)
vypustiša *pečenězi* *mužъ* *svoi*
released.3PL Pechenegs man.NA REFL.POSS.NA
'the Pechenegs released their man'
(*Primary Chronicle* 123.18–123.19)

(28) (The Novgorodians had Vsevolod as their prince, threw him out but then asked to have him back from Gyurgi)
i *dastъ* *imъ* *opjat̲* *sn̲ъ* *svoi* *vsevolodъ*
and gave.3SG them again son.NA REFL.POSS.NA Vsevolod.NA
'and he gave them his son Vsevolod again
(*Novgorod First Chronicle*, year entry for 6731/1223)

Both of these examples have objects that are characteristically sent on someone's behalf, but they are also previously mentioned multiple times, and also clearly actors in the narrative. The object lemma in the seven old anchored examples is overwhelmingly *synъ* 'son' (six out of seven, the only exception is (27), and four out of seven have *posъlati* 'send' or *posaditi* 'place' as the main verb. They therefore seem to fit into the anchored template, regardless of their own information status.

With these seven examples accounted for, only three old NA objects remain. One is another example with *mladьnьcь* 'infant' from the *Primary Chronicle* (94.6, in the story of Moses), the other two (29, 30) are both rather odd examples from the *Novgorod First Chronicle*.

(29) *vъ* *lět̲* *6650* *epsp̲ъ* *i* *kupьce* *i* *sly*
in year 6650 bishop.NA and merchants.ACC and envoys.ACC
novgorodьskyja *ne* *puščaxu* *iz* *rusi*
Novgorodian.ACC.PL not let-out.3PL from Rus
'In the year 6650: they did not let the bishop and merchants and envoys of Novgorod out of Rus"
(*Novgorod First Chronicle*, year entry for 6650/1142)

(30) *kn̲zь* *že* *jaroslav* *togo* *ne* *uljubivъ* *pusti* *popъ*
prince PTC Jaroslav that not liked let-go.3SG priest.NA
bez *mira*
without peace
'Prince Jaroslav did not like that and let the priest go without peace '
(*Novgorod First Chronicle*, year entry for 6723)

In (29), we know from the previous year entry that the bishop, merchants and envoys have been sent to Kiev, so this is clearly not the first mention, but oddly we get the NA even though the main verb is negated, which is very rare. In (30), the priest has been previously introduced in the GA, but is here for some reason picked up in the NA.

When we turn to the non-specific new category, at first glance we notice a possible tendency for new, nonspecific human masculine singular direct objects to occur in the NA – 11 NA and 21 GA in the NA-set and 8 NA and 21 GA in the variation set. However, we soon see that the issue of anchoring raises its head here as well. There are two main types of NA objects in the variation set, on the one hand objects occurring in universal conditionals in the *Russkaja pravda* (31), on the other hand examples that fit (or almost fit) Timberlake's main template, but happen to be set in an embedding that yields a non-specific interpretation of the object (32, 33).

(31) *ašče* *poznajetъ* *kto* *čeljadinъ* *svoi* *ukradenъ*
if recognises someone servant.NA REFL.POSS.NA stolen.NA
'If someone recognises a stolen servant of his (and takes him, then he must ...)' (*Russkaja pravda* 38)

(32) *ože* *xoščeši* *poslati* *mužь* *svoi* *i* *vorotit* *sja*
if want.2SG send man.NA REFL.POSS.NA and returns REFL
volodimerъ *to* *vdam* *ti* *kotoroi* *ti* *gorodъ*
Volodimer then give.1SG you.DAT which you.DAT city.ACC
ljubъ
pleasing.ACC
'if you will send your man and Volodimer turns back, then I will give you whichever city you desire' (*Primary Chronicle* 256.17–256.20)

(33) *vypusti* *ty* *svoi* *mužь* *a* *ja* *svoi* *da*
let-out.2SG you REFL.POSS.NA man.NA and I REFL.POSS.NA so-that
sja *boreta*
REFL fight
'Let out your man and I (will let out) mine so that they can fight.'
(*Primary Chronicle* 122.17–122.18)

Both in (32) and in (33) it is clear from context that the speaker does not have a specific individual in mind. In (32) a certain Vasilij is asked to ask Vasilko if he will send an envoy – the choice of envoy is clearly up to Vasilko. In (33) two princes and their armies meet by a river, and one prince asks the other if he will bring out one of his fighters. It is again clear that the choice is up

to the other prince. They are therefore tagged as non-specific, but they are also both compatible with Timberlake's main template, which cannot be said about (31).

Again we see that all three examples have a possessive pronoun dependent, i.e. they would have come out as anchored in Table 8 if the head noun had been tagged as NEW (and hence specific). This is the case for four out of eight non-specific new NA objects in the variation set (indicated in parentheses). However, three of the remaining examples are modified with another anchor-like modifier *čužь* 'someone else's', as in (34).

(34) *aže kto pereimet čjužь xolopъ i dastь věstь*
if someone captures other's.NA serf.NA and gives tidings.ACC
gsnu jego to imati jemu perejemъ gr̲i
master.DAT his then have he.DAT capture grivna
'If someone captures someone else's serf and lets his master know, then he shall receive a grivna for the capture' (*Russkaja pravda* 113)

If we also include the three examples with lemmas that only occur in the NA in the dataset, we find two further examples from universal conditional clauses in the *Russkaja pravda* (without anchors of any kind), as well as one example without anchoring but with a universal conditional from the *Primary Chronicle* (35).

(35) *ašče kto koli prinesjaše dětišťь bolenъ … prinesjaxu v*
if someone PTC brought.3SG child.NA sick.NA brought.3PL in
manastyrь
monastery.ACC
'If someone brought a sick child they would carry (it) into the monastery'
(Primary Chronicle 189.9–189.14)

In general, then, there seems to be a clear pattern for using the NA with *anchored* referents, and more specifically, with any human singular object with a possessive pronoun dependent (adjectival or genitival). I agree with Timberlake that the NA has been reduced to a constructionally conditioned marker, but I think the evidence suggests, especially when we look beyond the *Primary Chronicle*, that the construction we are dealing with is fairly general: if there is a possessive pronoun, you can have the NA, but since the GA is the default, GA is also a possibility, and in fact more frequent than the NA in any context, as Table 9 shows.

	full set		NA-set		variation set	
	NA	GA	NA	GA	NA	GA
no anchor	21	236	21	81	14	81
anchor	38	76	38	65	33	65

Table 9 Presence or absence of anchor by case, all datasets

The tendency for NAs to have anchors more frequently than GAs is statistically significant both in the NA-set and the variation set.[36] Timberlake's template semantics, I would argue, largely comes from the general semantic effect of combining a possessive pronoun with a common noun. There is very little NA usage beyond this – possible some more information-status related use with children, animals and non-specific objects, but given the scarcity of the data this is difficult to demonstrate with any certainty.

I believe that the OES data are best seen against the backdrop of the OCS dataset. What the two datasets have in common is a clear tendency to allow NA-marking of anchored objects. In the OCS dataset this fits in with a larger tendency to NA-mark new referents, whether specific or non-specific, anchored or unanchored. In the OES dataset there seems to be little left of the information-status motivation. Instead the NA has consolidated itself as a constructionally-licensed variant that is primarily found in constructions where the object is a common noun with a possessive pronoun dependent. These constructions typically go with the low-individuation semantics that Timberlake (1997) describes, but this is not likely to be a type of semantics that went with the NA in Late Common Slavic, since the OCS dataset indicated that the NA had specialised with highly individuated, referentially persistent objects, i.e. exactly the opposite development.

6 Snapshot 3: The Middle Russian spread to the plural

When we turn to the Middle Russian material, we find that the NA-GA variation is no longer present in the singular. The GA has taken over completely, and there is no room for the NA. In the singular, we are therefore already at the modern stage – animacy appears to be established as a subgender for masculine animate singular direct objects, including animals, as far as the relatively scarce data can tell us – we find only GA examples like (36) and (37),

36 NA-set: p=0.0133, variation set: p=0.002497, Fisher's Exact Test.

and no NA examples.

(36) *žena dobra veselit muža svoego*
wife good cheers-up man.GA REFL.POSS.GA
'a good wife cheers up her husband' (*Domostroj* 20, mid-16th century)

(37) *i jazъ grěšnyi privezlъ žerebъca. v ynděiskuju zemlju*
and I sinful brought.3SG horse.GA to Indian land
'and I, sinful as I am, brought a horse to the land of India'
(*Afanasij Nikitin's journey beyond three seas* 372v, c. 1475)

Instead, we can observe the next stage of development: a quick spread of the GA to the plural, aided by the merger of the nominative and accusative plural, which had up to then been separate. This merger appears to have started in the 13th century, spread quickly in the 14th century and was completed by the 15th/16th century (Kiparsky 1967: 42). We shall use the all the Old East Slavonic and Middle Russian texts in TOROT to quickly sketch this expansion.

6.1 Annotation and data extraction

From an annotation perspective, as long as there is variation per lemma, we still cannot conflate the GA with the regular accusative. As we still find variation in the plural in some of the very latest texts in TOROT, the annotation policy of taking the genitive at face value was retained throughout.

The Middle Russian TOROT data has no animacy tagging and no information status tagging. Nothing was done to compensate for the lack of the latter, but to compensate for the lack of the former, data were again extracted by list. First, all non-negated plural masculine[37] objects were extracted under the same conditions as before (see Section 5.1). A list of animate lemmas was then compiled by examining all direct objects that were attested in the GA plural and selecting the animate ones, and combining this list with the lists of human and animate objects used to sift the OES dataset. Note that this list did not only contain common nouns, but also adjectives and verbs that occurred as nominalisations in the GA. The combined list was used to sift the dataset in the R script used to analyse the data. The dataset was extracted from the full set of OES and Middle Russian texts in TOROT, which makes it possible to track the development throughout the history of Russian.[38] The method of

37 The GA also spread to feminine and marginally neuter animate plurals, but I will not examine this development here.

38 TOROT also contains a modern Russian treebank, a converted version of SynTagRus, but since the generalisation of the GA to the plural was completed well before the modern Russian stage, no data was extracted from this treebank.

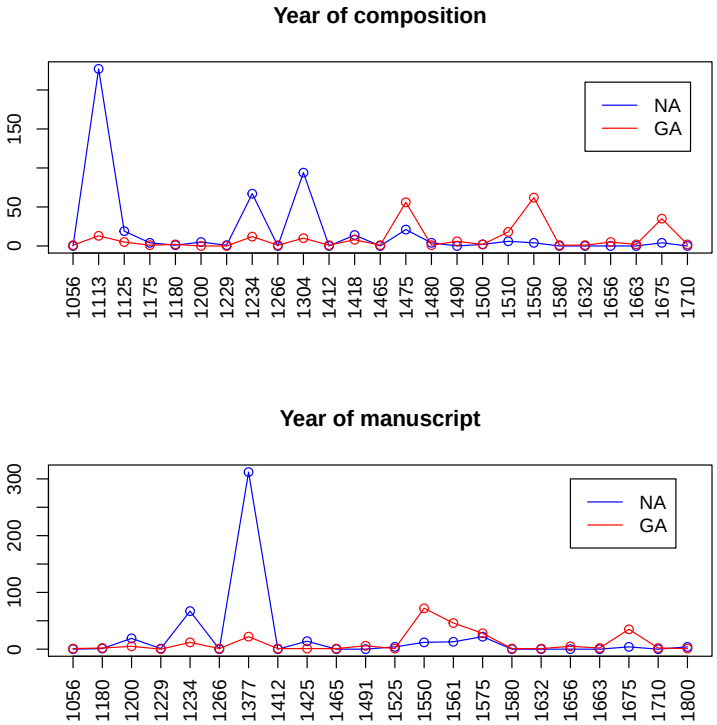

Figure 1 Plural masculine animate object case by year of composition and year of manuscript, all sources, raw figures

extraction yielded 719 occurrences (473 NA, 246 GA).

6.2 Analysis

The rise of the GA in the plural is visualised in Figures 1, 2 and 3. While the singular NA and GA remained in a situation of variation for centuries, the plurals seem to change fairly abruptly.

In each figure there is a separate graph for year of composition and year of manuscript (both often approximate). As previously mentioned, the gap between date of composition and manuscript can be more than two centuries for the earliest texts,[39] and the status of the text is therefore unclear – should

39 The extreme case is the 13th-century Tale of Igor's campaign, which has come down to us in a

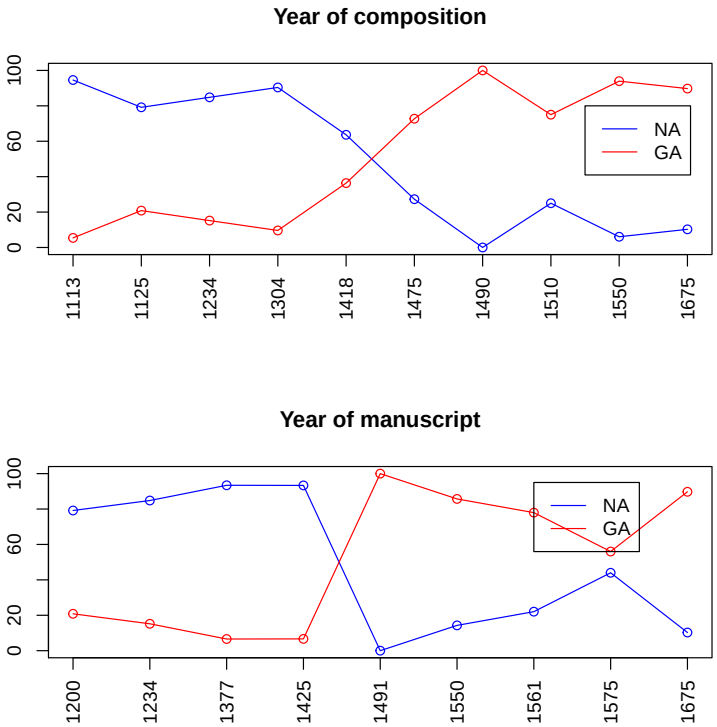

Figure 2 Plural masculine animate object case by year of composition and year of manuscript (sources with at least 6 occurrences, percentages)

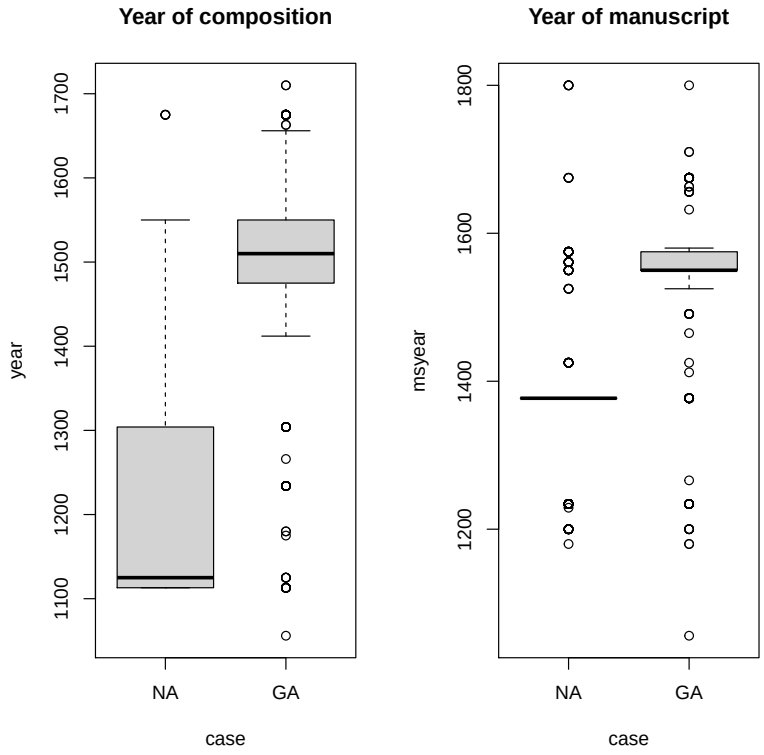

Figure 3 Plural masculine animate object case by year of composition and year of manuscript, all sources

it be taken to represent the time of composition or the time of manuscript? Notably, Klenin (1983: 91), states that "it is generally agreed in the literature that the plural human-referential genitive-accusative is first attested in the 14th century", but this is clearly based on a strict adherence to the year of the manuscript, e.g. dating both the *Primary Chronicle* and the *Suzdal Chronicle* to the 14th century on the strength of the 1377 Laurentian manuscript to which they both belong. However, this disregards the fact that the distribution of NA and GA is different in the *Primary Chronicle* and the *Suzdal Chronicle*, both in the singular and the plural, and in a way that strongly suggests diachronic change.

In general, the shapes of the graphs suggest that the year of composition is a better measure, i.e. that the scribes have been conservative and unwilling to change the morphosyntax of the text. Figure 1 is a simple graph of the raw occurrences of each case by year of composition, and simply shows us that the NA has the upper hand until the late 15th century (slightly later by year of manuscript), after which the GA is more frequent. Figure 2, however, which uses data from texts with more than five occurrences and is given in percentages, shows us that we are dealing with a fairly abrupt, S-curve-like change, regardless of whether we use the date of composition or of manuscript. Exactly the same thing is shown in the boxplots in Figure 3 without omitting any data. The graph using the date of composition looks more natural, as it does not contain an apparent decrease in use of the GA before the 15th-century shift, as the date-of-manuscript graph does. This is even more clear in the boxplots in Figure 3, where the year-of-composition plot shows the GA taking over from the NA relatively abruptly, while the year-of-manuscript plot fails to make much sense of the NA occurrences, modelling them as mostly outliers with a median in the late 14th century (clearly due to the Laurentian manuscript). I will therefore base my further discussion on the year of composition.

In the 11th–14th century texts we see 80–100% NA, and only occasional use of GA, while the 16th–17th century texts have the opposite pattern: 80–100% GA, and only occasional use of NA. The 15th century looks like a period of possible free variation (which has been artificially pushed to 1575 in the year-of-manuscript graph). The *Life of Sergij of Radonezh* (c. 1418/c. 1575) has 14 NA and 8 GA, while two texts from c. 1475, *Afanasij Nikitinâ□□s journey beyond three seas* and *The tale of the fall of Constantinople* (mss. c. 1550 and 1561) have 8 NA, 10 GA and 13 NA, 46 GA respectively. We sometimes find NA and GA used for the same lemma in very similar contexts in the same text, as in (38) and (39), and (40) and (41).

printed edition from 1800.

(38) *i* *prognaša* *turky* *k* *polomu* *městu*
and drove.3PL Turks.NA to ruined place
'and they drove the Turks to the ruined part (of the wall)'
(*The tale of the fall of Constantinople* 313v 16, late 15th century)

(39) *i* *tako* *prognaša* *tur'kovъ* *k* *polomu* *městu*
and thus drove.3PL Turks.GA to ruined place
'and thus they drove the Turks to the ruined part (of the wall)'
(*The tale of the fall of Constantinople* 312r 12, late 15th century)

(40) *i* *privozjatъ* *koni*
and bring.3PL horses.NA
'and they bring horses'
(*Afanasij Nikitin's journey beyond three seas*, c. 1475)

(41) *privodjatъ* *konei*
bring.3PL horses.GA
'they bring horses' (*Afanasij Nikitin's journey beyond three seas*, c. 1475)

If we look at TOROT texts from this period of variation (limited to 1412–1510), it is again clear that only a few lemmas occur in actual variation – a total of 10 out of 57 lemmas. We have already seen *turčinъ* 'Turk' (4 NA, 7 GA) and *konь* 'horse' (6 NA, 1 GA),[40] in addition there are eight more, the most frequent of which is *ljudie* 'people' (10 NA, 15 GA).[41] The remaining seven are only attested once or twice in each case. There are ten lemmas occurring in NA only (none more than three times), while the remaining 47 lemmas occur in the GA only (none more than five times). A much larger dataset would be needed to draw any firm conclusions about possible lexical diffusion of the change. Given the swift, s-curved change it also seems unlikely that any semantic or pragmatic differentiation of the two cases had time to settle.[42]

There is clearly more to be said about the development of the GA in the plural, but the aim of this section is merely to point out its obvious contrast to

40 There is reason to believe that the GA spread more slowly to animal nouns: Kedajtene (1961: 187) argues that animal GA plurals only start appearing in the 17th century, but obviously the *Afanasij Nikitin* example is much earlier.

41 Kedajtene (1961: 186–187) notes that *ljudie*, along with certain other nouns with a collective meaning, such as *děti* 'children' and *gosti* 'guests' were particularly resistant to the change. It is also worth noting that none of these nouns are ŏ-stems: *ljudie* and *gostь* are ĭ-stems, and *děti* is an old consonant stem, so morphological conditioning is possible.

42 Note that there are still some minor vestiges of NA left in modern Russian, such as *iti v gosti* 'visit (literally go in guests.NA), *vybrat' v prezidenty* 'elect president (literally elect in presidents.NA), but they are all in prepositional phrases, which are not under consideration here.

the process in the singular. While the point of departure for both processes was nominative-accusative syncretism and the need for discrimination (cf. Seržant 2019), it seems clear that the two played out very differently. The NA-GA variation in the singular lasted for centuries and was, at least to begin with, pragmatically motivated, before it lingered on as a constructionally licensed variant in OES. In the plural, on the other hand, we see a quick, sharp S-curve change – when the nominative-accusative distinction is lost, the generalised animate accusative marker is quickly extended to the plural, without a clear stage of functional specialisation.[43]

7 Conclusions

The aim of this article was twofold: first, to use enriched treebank data to make sense of a much-debated change in the history of Russian, namely the rise of the animacy category, and second, to take a methodological look at the challenges posed by this type of change to treebank annotation.

The treebank data show that the situation we observe in the OES dataset is at the tail end of a long-haul change, characteristic of alternations between functionally specialised markers. Looking at the OES data alone, we can hardly claim that the NA is an indefiniteness marker anymore, nor that the GA is anything more specialised than a default marker for singular human masculine objects, and this situation has lead e.g. Krys'ko (1994) to dismiss the idea that the alternation was ever motivated by semantic or pragmatic factors. However, if we look at the data against the backdrop of the OCS dataset, we see that the constructionally-licensed use of the NA that is still fairly frequent in the OES material appears to be directly motivated by the information-status-driven distribution that we see in the OCS dataset: anchored objects used to be NA-marked because they had not previously been mentioned, but the NA-marking became associated with the possessive pronoun anchor instead of with the status of the head noun. The change in the plural is an entirely different type of change: since nominative-accusative syncretism appeared in the plural only after the GA had been generalised as an animate subgender accusative marker in the singular, it could spread swiftly and without clear indications of functional specialisation.

From the methodological point of view, we saw that the phenomenon had been given a very conservative analysis in the PROIEL and TOROT treebanks: GA-objects were analysed as morphological genitives throughout the period under consideration, even though this was almost certainly the 'wrong' anal-

43 But see Timberlake (1997: 57–60) for the observation that the distribution of NA and GA in the plural in 15th century chancery correspondence between Polotsk and Riga was also driven by low individuation/agentivity.

ysis at any attested stage of the development. Nonetheless, it was chosen because it made automatic tagging simpler and also saved annotators from having to make difficult decisions, and made the data as retrievable as a special GA analysis would have done – in any case the relevant dataset was impossible to extract without resorting to some form of animacy annotation and without using the syntactic annotation to restrict the query. Finally, we saw that it was possible to make much more of the data when principled information-status annotation was added to the earliest data layers.

References

Aissen, Judith. 2003. Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory* 28(3).

Ariel, Mira. 1988. Referring and accessibility. *Journal of Linguistics* 24(1). 65–87.

Baerman, Matthew. 1998. Defining 'subgender'. virile and devirilized nouns in Polish. *Lingua* 104. 187–233.

Bratishenko, Elena. 2003. Genitiveâ□□accusative and possessive adjective in Old East Slavic. *Scando-Slavica* 49(1).

Comrie, Bernard. 1978. Genitiveâ□□accusatives in Slavic: the rules and their motivation. In Bernard Comrie (ed.), *Classification of grammatical categories*, Edmonton: Linguistic Research.

Corbett, Greville G. 1982. Gender in Russian: An account of gender specification and its relationship to declension. *Russian Linguistics* 6(2). 197–232.

Corbett, Greville G. 1991. *Gender* Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Eckhoff, Hanne Martine. 2015. Animacy and differential object marking in Old Church Slavonic. *Russian Linguistics* 39.

Eckhoff, Hanne Martine & Aleksandrs Berdicevskis. 2015. Linguistics vs. digital editions: The Tromsø Old Russian and OCS Treebank. *Scripta & e-Scripta* 14–15.

Haug, Dag T. T. & Marius L. Jøhndal. 2008. Creating a parallel treebank of the old Indo-European Bible translations. In Caroline Sporleder & Kiril Ribarov (eds.), *Proceedings of the second workshop on language technology for cultural heritage data* LaTeCH 2008, 27–34.

Haug, Dag Trygve Truslew, Hanne Martine Eckhoff & Eirik Welo. 2014. The theoretical foundations of givenness annotation. In Kristin Bech & Kristine Eide (eds.), *Information structure and syntax in Germanic and Romance languages*, Amsterdam: John Benjamins.

Haug, Dag Trygve Truslew, Marius Jøhndal, Hanne Martine Eckhoff, Eirik Welo, Mari Johanne Bordal Hertzenberg & Angelika Müth. 2009. Com-

putational and linguistic issues in designing a syntactically annotated parallel corpus of Indo-European languages. *Traitement Automatique des Langues* 50.

Huntley, David. 1993. Old Church Slavonic. In Bernard Comrie & Greville G. Corbett (eds.), *The Slavonic languages*, London/New York: Routledge.

Karttunen, Lauri. 1969. Discourse referents. In *Proceedings of the 1969 conference on computational linguistics* COLING '69, 1–38. Stroudsburg, PA, USA: Association for Computational Linguistics.

Kedajtene, E. I. 1961. K voprosu o razvitii form roditel'nogo-vinitel'nogo padeža (na materiale vostočnoslavjanskix jazykov). In *Issledovanija po leksikologii i grammatike russkogo jazyka*, Moscow: AN SSSR.

Kiparsky, Valentin. 1967. *Russische historische Grammatik. Band II: Die Entwicklung des Formensystems*. Heidelberg: Carl Winter.

Klenin, Emily. 1983. *Animacy in Russian. A new interpretation*. Columbus, Ohio: Slavica.

Krys'ko, V.B. 1994. *Razvitie kategorii oduševlennosti v istorii russkogo jazyka*. Moscow: Lyceum.

Krys'ko, V.B. 2006. *Istoričeskij sintaksis russkogo jazyka. Ob"ekt i perexodnost'*. Moscow: Azbukovnik.

Ledgeway, Adam, Norma Schifano & Giuseppina Silvestri. 2019. Differential object marking and the properties of d in the dialects of the extreme south of italy. *Glossa* 4(1). 51.1–25.

Lunt, Horace G. 2001. *Old Church Slavonic grammar*. Berlin: Mouton de Gruyter 7th edn.

Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.

Malchukov, Andrej & Peter de Swart. 2009. Differential case marking and actancy variations. In Andrej Malchukov & Andrew Spencer (eds.), *The oxford handbook of case*, Oxford: Oxford University Press.

Meillet, Antoine. 1897. *Recherches sur l'emploi du génitif-accusatif en vieux-slave*. Paris: É. Bouillon.

Michell, Robert & Nevill Forbes (eds.). 1914. *The chronicle of Novgorod 1016–1471*. London: Royal Historical Society.

Prince, Ellen. 1981. *Toward a taxonomy of given-new information* 223–255.

Seržant, Ilja A. 2019. Weak universal forces: The discriminatory function of case in differential object marking systems. In Natalia Levshina, Karsten Schmidtke-Bode, Ilja Seržant & Susanne Maria Michaelis (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence*, Berlin: Language Science Press.

Timberlake, Alan. 1997. Čemu jesi slěpil brat svoi: Templates and the devel-

opment of animacy. *Russian Linguistics* 21(1).
von Tischendorf, Constantin. 1889. *Novum Testamentum Graece*. Leipzig: Gieseche & Devrient.
Tomson, A. I. 1908/2006. Roditel'nyj-vinitel'nyj padež pri nazvanijax živyx suščestv v slavjanskix jazykax. In A. I. Tomson (ed.), *Izbrannye raboty po morfologii slavjanskix jazykov. razvitie kategorii oduševlennosti*, Moscow: KomKniga.
Živov, Viktor M. 2017. *Istorija jazyka russkoj pis'mennosti*, vol. 1. Moscow: Universitet Dmitrija Požarskogo.
Zaliznjak, Andrej A. 2004. *Drevnenovgorodskij dialekt*. Moscow: Jazyki slavjanskoj kul'tury.

Hanne Eckhoff
Lady Margaret Hall
Norham Gardens
Oxford
OX2 6QA
hanne.eckhoff@mod-langs.ox.ac.uk
https://www.mod-langs.ox.ac.uk/people/hanne-eckhoff