ULLA STROH-WOLLIN

The newly born definite article spreads along two paths – a theoretical discussion and a case study in Old Scandinavian

1. Introduction

In a recently published volume on the grammaticalization path of the definite article in German, it is pointed out that the various grammaticalization models proposed in the literature (e.g. Greenberg 1978, Lehmann 1985/2015, Lyons 1999 and Himmelmann 1997, 2001) assume a strictly linear progression. But, the editors write, recent research challenges this view, which shows that we need more empirical studies on the develop-

This article is an outcome of my project "Through the forest of hypo-determination – the Scandinavian case". I am indebted to the Swedish Research Council for financially supporting the project (project number: 2017-01533). I also want to give special thanks to Veturliði Óskarsson who generously shared his time to help me in reading the Icelandic texts and two anonymous reviewers for their valuable comments on an earlier version of this paper.

Stroh-Wollin, U., Professor emerita, Department of Scandinavian Languages, Uppsala University. "The newly born definite article spreads along two paths – a theoretical discussion and a case study in Old Scandinavian". *ANF* 139 (2024), pp. 21–88.

Abstract: This article contains an investigation of the spread of the definite article in Old Scandinavian on its way towards obligatorification from the very scant use of the newly born article. On an overall level, it turns out that the process is significantly earlier in West Norse than in East Norse. The empirical analysis is preceded by a theoretical discussion of which referential categories there is reason to distinguish, alongside the (direct and indirect) anaphora, where the development is generally assumed to begin, as definite articles normally arise from demonstratives in anaphoric use. When it comes to the non-anaphoric NPs, it is argued, with inspiration from Löbner 2011, that there is reason to pay regard to whether or not the head noun expresses inherent uniqueness and/or relationality. This view is confirmed by the empirical analysis, which shows that scribes can be significantly more as well as significantly less inclined to use the definite article in NPs with such features than in NPs without. Moreover, this result suggests that definiteness marking does not progress strictly linearly along one single path, contrary to what various models in the literature presume (including the scale of uniqueness in Löbner 2011). Instead, a two-path model is proposed.

Keywords: development of definiteness marking, Old Scandinavian, definite suffix, grammaticalization, referential categories, inherent relationality, inherent uniqueness.

DOI: https://doi.org/10.63420/anf.v139i.28743

ment of definite articles. (Szczepaniak & Flick 2020) In this spirit, this paper both accounts for an empirical investigation of the spread of the (post-nominal) definite article in Old Scandinavian and contributes to the theoretical discussion.

The empirical and theoretical parts of the study are intertwined, in that the empirical analysis is based on and confirms the results of the theoretical discussion on referential categories that precedes it and thereby also leads to a conclusion of cross-linguistic relevance, namely that the definite article spreads along two paths, once it has emerged in a language.

As for the empirical investigation, the focus is entirely on the post-nominal, suffixed, definite article, which is common to all Scandinavian languages[1] and is the only way in these to mark definiteness in noun phrases without pre-nominal modifiers or restrictive relative clauses. The later development of pre-adjectival definite articles and definiteness-marking in noun phrases with relative clauses will not be followed here, for reasons explained in section 2.

Further, for the sake of clarity, the focus is not on the emergence of the article, which I will just briefly touch on, but on the subsequent extension of its use. It is generally assumed that definite articles arise from demonstratives, a good reason being that demonstratives, just like definite articles, can be used for anaphoric reference. The Scandinavian case does not give reason to question the current view on this point. It is also the case that even in the earliest documented Scandinavian we find usages that are incompatible with demonstratives, i.e. uses where a referent is not introduced beforehand in the discourse but is still accessible to the receiver on the basis of some other knowledge shared with the sender.

A fundamental question for the empirical study, which warrants the relatively extensive theoretical considerations that precede it, is how the nature of this shared knowledge matters in the process from the very scant use of the newly born article to its obligatorification. A first step in exploring this question is to establish a well-grounded view of which referential categories besides the anaphoric uses there is reason to distinguish, after which it will be possible to investigate whether the article spreads at a different pace in the different categories.

The influential systemization of the uses of definite articles in Hawkins 1978 (chapter 3) can serve as a point of departure for the discussion.

[1] All standard varieties use the post-nominal article, but the vernacular spoken in south-western Jutland in Denmark actually has a pre-posed definite article (*æ*); see further below.

Hawkins distinguishes between *anaphora*, *associative anaphora*, *immediate situation uses* and *larger situation uses*. An anaphor (or a direct anaphor) is co-referential with a preceding antecedent; see (1a–b). The antecedent (underlined below) and the anaphor (in bold face) can have the same head noun as in (1a), but this is not necessary, cf. (1b).

(1) a. *A little boy went missing yesterday, but fortunately* ***the boy*** *was found safe this morning*.
b. *A little boy went missing yesterday, but fortunately* ***the child*** *was found safe this morning*.

An associative (or indirect) anaphor has no co-referential antecedent but is still triggered by (or anchored in) some conceptually related word or phrase in the preceding context. A typical example is part-whole relations as in (2a), where *the chimney* is easily understood as the chimney of the previously mentioned house. In part-whole relations, there is always an NP anchor (underlined in 2a) to the indirect anaphor. Other kinds of indirect anaphora can be deduced on various grounds, e.g. as in (2b), by a verb phrase (underlined); here, *the money* is understood as the outcome of the selling of the house.

(2) a. *The house is mortgaged to the top of* ***the chimney***.
b. *They have sold their house and will use* ***the money*** *on a round-the-world trip*.

Unlike NPs with (direct and indirect) anaphoric reference, the immediate and larger situation uses do not depend on the preceding discourse. The immediate situation uses comprise uses where the referent is accessible in the immediate situation, i.e. the immediate physical surrounding, whether or not the referent is visible to the hearer. If the hearer can see the bucket in (3a), it would also be possible to use a demonstrative phrase: *that bucket*, but replacing *the dog* with *that dog* in the message in (3b) would be odd if found on a sign on a gate for instance.

(3) a. *Pass me* ***the bucket****, please.* Hawkins 1978:111
b. *Beware of* ***the dog****.* Hawkins 1978:112

The larger situation uses, finally, differ from the immediate situation uses in that they require the recipient not only to be able "to locate the referent within some situation" but also "to identify the correct situation" (Hawkins 1978:120). This situation, or, to put it another way, the relevant context of the discourse, can be larger or smaller. For example, *the king*

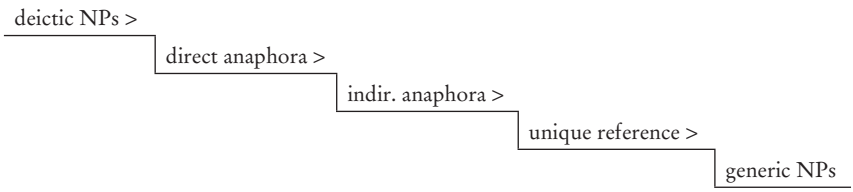

Figure 1. Grammaticalization model of how definiteness marking spreads to new referential categories, after Skrzypek 2012:49 (see also Skrzypek et al. 2021:30).

can refer to the king within a contextually given monarchy and *the fridge* to the fridge of a contextually given household etc.

Hawkins' classification is recognizable in the grammaticalization model (based on the grammaticalization theory in Heine 2002) in Skrzypek et al. 2021 (p. 30, following Skrzypek 2012:49). The different stages of the diachronic development are presented in Figure 1, where *unique reference* corresponds to Hawkins' larger situation uses.[2] Regarding the immediate situation uses, only the deictic kind is considered. Moreover, *generic NPs*, which are not considered in Hawkins 1978, are added to the model as a possible extended use of the definite article beyond the unique reference NPs/larger situational uses.

A somewhat different view of the development of definiteness marking follows from the theory of determination presented in Löbner 2011 (first outlined in Löbner 1985) and the author's scale of uniqueness (Löbner 2011:320). Löbner distinguishes four conceptual types of nouns, defined by the possible combinations of two binary features, inherent uniqueness and inherent relationality, which make some nouns more inherently definite than others. The scale of uniqueness runs from pragmatic uniqueness, which applies to definite NPs that lack inherent uniqueness (as well as relationality), to semantic uniqueness, which applies to definite NPs with inherent uniqueness. The development of definiteness marking is assumed to start at the pragmatic end and progress along the scale. (See further section 4.1.)

[2] The label *unique reference* may be inspired by De Mulder and Carlier 2011 (p. 531), where this type is described as reference to entities that are "unique within the discourse universe".

The two perspectives just mentioned differ on one very important point. While the referential classification in Hawkins 1978 and its application in the grammaticalization model in Figure 1 do not pay regard to the lexical content of the NP, it follows from the basic assumption in Löbner 2011 that the development of definiteness marking is affected by the lexically inherent features of uniqueness and relationality.

In the following, I will argue that there is reason to take into account whether definite NPs possess inherent uniqueness and/or relationality when sorting them into different referential categories. I consider it reasonable that such features are important clues for the recipient's identification of the intended referent.

Nevertheless, as will become clear, I do not find Löbner's uniqueness scale to be the ideal model for understanding the diachronic development of definite articles, among other reasons because it is unclear how indirect anaphora and NPs in larger situational uses that are not equipped with uniqueness or relationality features fit into the scale. An outcome of this study is a proposed model that allows the definite article to spread along two paths when it begins to be used in non-anaphoric noun phrases, one for NPs with inherent uniqueness and/or relationality and one for NPs without.

The rest of the article is organized as follows. Section 2 briefly presents what the reader should preferably know about the noun phrase in Old and Modern Scandinavian in order to fully understand the presentation in the following. Section 3 accounts for some points of the departure with relevance for the two following sections, of which section 4 is a theoretical discussion, based on previous research, chiefly about referential categories, while section 5 accounts for the empirical investigation. Section 6, finally, is a discussion of the methodological and, above all, theoretical implications of the study.

Before proceeding, a couple of reading guidelines are required. First, throughout the text, I will use NP as an abbreviation for "noun phrase" without reference to the abstract structure of noun phrases. After Abney 1987, it is common to use the designation DP (even outside the generative paradigm) instead, but it has also later been suggested that the DP structure applies only to article languages, while noun phrases in article-less languages would have NP structure (see e.g. Bošković 2012 and subsequent works). In this context, it is not necessary to take a position on this issue.

Second, much of the literature on definiteness and definite articles takes Hawkins 1978 and the referential categories for definite NPs presented there as the point of departure for their accounts, but the terminology varies somewhat. When Hawkins distinguishes between anaphora and associative anaphora, it is now common to talk of *direct anaphora* and *indirect*, or *bridging*, *anaphora* for the same categories. I will stick to *direct* and *indirect* anaphora here. I also prefer to say that an indirect anaphor is *anchored* in the preceding context (following Fraurud 1990), rather than *triggered* (as in Hawkins' description).

2. A brief guide to the NP in Old and Modern Scandinavian

Some knowledge concerning the Scandinavian NP and its structure, variation and diachronic development will facilitate the reading in the following. The most important facts are briefly presented below.

As mentioned above, all Scandinavian languages have developed a post-nominal, suffixed, definite article. The point of departure was a now lost demonstrative, reasonably in post-nominal position and with a tendency to cliticise to the noun. Interestingly, although the outcome was the same all over Scandinavia, the article appears to have arisen from two separate, albeit related, origins, *enn* and *hinn* (a reinforcement of *enn*), which means that the development ran in two parallel processes (and from at least two separate innovation centres), as illustrated in (4a–b) with a noun in the neuter gender, *hus* 'house'.

(4)	a.	*hus et*	>	*hus'et*	>	*husit*
	b.	*hus hit*	>	*hus'*(*h*)*it*	>	*husit*
		house DEM				house.DEF

As shown in Stroh-Wollin 2016, 2020, it started with *enn* in Iceland, but we find evidence of this origin also in some early Norwegian texts. This may be due to Icelandic influence on Norwegian writing, if (5a) shows a purely Icelandic innovation (as suggested in Stroh-Wollin 2016, 2020). But it should also not be excluded that the innovation arose in (north-)western Norway among the people who colonized Iceland around 900. For the rest of the mainland, including most of Norway, various circumstances speak for an origin in *hinn*, although we only find a few transitional forms

with preserved *h* in a single manuscript, a fragment of the Norwegian *Saga of Olav Tryggvason* (No. 4 in Codex De la Gardie 4–7), e.g. *hunzhins* 'the dog's'. (See further Stroh-Wollin 2020.)[3]

The first instances of a noun in the definite form are found in two runic inscriptions from the province of Uppland in eastern Sweden dated to the 11th century. Both inscriptions end with the same little prayer, the one that is rendered in transcribed form in (5). (See sigla U 644 and U 669 in the Scandinavian Runic-text Database.) For the sake of clarity, *and-inni* is rendered with 'the soul' in the English translation, although 'his soul' would have been more idiomatic. Note also that the article (and its demonstrative origin) appeared in different guises due to inflection for gender, number and case; cf. *-it* (neut.sing.nom./acc.) in *husit* in (4b), *-hins* (masc.sing.gen.) in *hunzhins* and *-inni* (fem.sing.dat.) in *andinni* in (5).

(5) *Guð hialpi andinni.*
God help.SUBJ soul.DEF
'May God help the soul.'

The Scandinavian demonstrative corresponding to English *that* and the related words in West Germanic have also given rise to a definite article used only in front of pre-nominal modifiers. This article is found in the mainland varieties and Faroese, but not in Icelandic. The Swedish example in (6a) shows so-called double definiteness, i.e. concomitant use of the pre-adjectival article and the definite form of the noun. The same principle applies to Norwegian and Faroese, whereas Danish takes the noun in the indefinite form in the same context, see (6b). Icelandic, on the other hand, normally can do without any pre-adjectival article in corresponding cases but takes the noun in the definite form; see (6c).[4]

(6) a. *det stora huset* (Swedish)
the big house.DEF.
'the big house'

[3] The different origins of the definite article in Iceland and (most of) Norway show that West Norse should not be perceived as a uniform variety in all respects, and this of course applies to East Norse as well. Although it may sometimes be convenient to set West Norse against East Norse, we should also consider Old Scandinavian as a dialect continuum.

[4] However, see Pfaff 2014, 2015 for more detailed accounts on NPs with adjectival attributes in present-day Icelandic and Pfaff 2019 regarding the development from the 13th century onward.

b. *det* *store* *hus* (Danish)
the big house
'the big house'
c. *stóra* *húsið* (Icelandic)
big house.DEF.
'the big house'

Thus, Scandinavian possesses two different definite articles with different etymologies. While the different varieties are uniform in their way of expressing definiteness in NPs without modifiers, they have landed in different rules when it comes to NPs with modifiers. It is also worth noting that the different modes of definiteness marking in different kinds of NPs have not developed concurrently. Furthermore, the emergence of the free article is somewhat more obscure than one might expect, as it has a precursor (with unclear function) in a companion to adjectives of the so-called weak declension, which had early lost the deictic meaning of the original demonstrative.

The pre-adjectival definite article is homonymous (stress disregarded) with the demonstrative from which it counts its origin, which means that the NPs in (6a–b) have a demonstrative reading as well: 'that big house'. (Such a reading in Icelandic requires a demonstrative: *það stóra húsið*.) However, only a demonstrative reading is possible if this determiner appears in direct contact with a noun (provided that is not followed by a restrictive relative clause, cf. below), in which case the noun takes the indefinite form in Icelandic and Danish and the definite form in Swedish and Norwegian. Thus, the meaning of the NPs in (7a–c) can only be read as 'that house', never as 'the house'.

(7) a. *það hús* 'that house' (Icelandic)
b. *det hus* 'that house' (Danish)
c. *det huset* 'that house' (Swedish, Norwegian)

In Old Scandinavian, demonstratives usually appear with nouns that are unmarked for definiteness. Demonstratives could also be post-posed as well as pre-posed (occasionally before nouns in the definite form) – irrespective of variety.

Free pre-nominal determiners also appear in Scandinavian NPs with restrictive relative clauses, in which case the noun takes the indefinite form in Danish and Icelandic, and either the definite or the indefinite form in Swedish and Norwegian – regardless of whether the noun is preceded by an adjectival attribute or not; see the Swedish examples in (8a–b). It

should be noted, though, that restrictive relative clauses are not always anticipated by the pre-nominal determiner. Also, single nouns in the definite form can combine with restrictive relative clauses, in Icelandic even when preceded by attributes; see (9a–b).[5]

(8) a. *det hus*(*et*) *som är till salu* 'the house that is for sale' (Swedish)
b. *det stora hus*(*et*) *som är till salu* 'the big house that is for sale' (Swedish)

(9) a. *huset som är till salu* 'the house that is for sale' (Swedish)
b. *stóra húsið sem er til sölu* 'the big house that is for sale' (Icelandic)

Although the pre-nominal determiners that anticipate restrictive relative clauses have the same form as the pre-adjectival article and are rendered as *the* when translated into English, they are not referred to as articles, but as determinatives (cf. Diessel 1999:135 f.). Further, in Old Scandinavian, demonstratives were often attached to the left periphery of relative clauses, i.e. after but not necessarily in direct contact with the noun. I find it reasonable to seek the origin of the phrase-initial determinative illustrated in (8a–b) in this use, but, to the best of my knowledge, this process has not been explored in any detail.

The empirical investigation that will be presented in the following is based on excerption of semantically definite noun phrases that consist of either bare nouns, nouns in the definite form or combinations of noun and demonstrative. One reason for excluding NPs with various modifiers is that Scandinavian, as we have seen, varies considerably as regards the modes of definiteness marking in NPs with (certain types of) modifiers and in NPs without, which, moreover, have not developed concurrently. Further, it is not obvious that all modes of definiteness marking have – or have always had – exactly the same formal and/or functional impact on the NP.

While NPs with adjectival attributes, relative clauses etc. were excluded from the investigation, NPs with demonstratives have nevertheless been included. This choice is warranted by the fact that demonstratives are generally possible with various cases of anaphoric reference, whether in

[5] Note, however, that the relative clauses in (9) are ambiguous between a restrictive and a non-restrictive reading. The preferences for one or the other of the variants demonstrated in (8) and (9) vary somewhat across the different Scandinavian languages. The choice is also a matter of style, e.g. in Swedish and Norwegian, *det hus som …* has a more formal flavour than *det huset som …*.

languages with definite articles or languages without. Some Old Scandinavian texts also show uses of demonstratives, where modern versions of the same texts would rather have the definite form of the noun (or simply an anaphoric pronoun). But, importantly, these uses are still always consistent with a demonstrative reading, and there is thus no reason to assume that we are seeing in them a tendency towards the emergence of yet another definite article alongside the post-nominal cliticized one.

3. Points of departure

In this section, I make some clarifications of how I relate to previous research in some aspects of importance for this study. Section 3.1 deals briefly with the notions of definiteness and shared knowledge, especially in relation to the account in Hawkins 1978, and section 3.2 with the question of when a definite article has emerged. Section 3.3, finally, comments on a couple of previous studies that specifically focus on the development of definiteness marking in Old Scandinavian.

3.1 Definiteness and shared knowledge

The notion of definiteness has been the subject of a long discussion.[6] In the literature, one finds two main approaches based on the concepts of *familiarity* and *uniqueness* respectively. (For an overview, see e.g. Lyons 1999.) The former approach is chiefly associated with Paul Christophersen (1939), and later Irene Heim (1982) and Craige Roberts (2003), each with their own definition of familiarity. There are also other labels for essentially the same concept, e.g. *givenness* (Gundel et al. 1993). The view of definiteness as uniqueness goes back to Bertrand Russell (1905), who takes a more formal semantic approach. Both approaches have been criticized for not capturing all instances of definiteness.

Russell's uniqueness theory as well as Christophersen's familiarity theory are also critically scrutinized in John Hawkins' influential book *Definiteness and Indefiniteness* (1978). Nevertheless, Hawkins' descrip-

[6] In this section, I concentrate on the relationship between definiteness and the use of definite articles. However, one may argue that semantic definiteness is not exclusively expressed by definite articles, but can also be conveyed by other grammatical means such as word order, case and aspect; see e.g. Krámský 1972 and Leiss 2000, 2007, to which I will return in section 3.3.

tion of the use of the definite article shows influence from both traditions. The importance of "shared knowledge" between sender and receiver, for instance, springs from the familiarity concept. When it comes to uniqueness, Hawkins rather talks of inclusiveness. Uniqueness, he writes, results from a fusion of the meaning of the definite article and the "oneness" of a singular noun, but it cannot be part of the meaning of the article itself, as definite articles in combination with plural count nouns or mass nouns do not refer uniquely. Instead, plural NPs and NPs with mass nouns refer *inclusively* to the totality of pragmatically identified sets of objects or masses that fit the descriptive content of the phrase – and, in fact, this holds true for the singular phrases with count nouns as well. (Hawkins 1978:158 f.)

Hawkins summarizes the third chapter, "The Referential Meaning of Definiteness", as follows.

> The use of the definite article acts as an instruction to the hearer to locate the referent of the definite NP within one of a number of sets of objects which are pragmatically defined on the basis of different types of shared speaker-hearer knowledge and the situation of the utterance. The hearer locates the referent in the sense that he understands that the object referred to is a member of the appropriate, pragmatically identifiable, set. The definite description refers 'inclusively' to the totality of the objects satisfying the descriptive predicate within the relevant pragmatic set. (Hawkins 1978:17)

This definition takes the use of the definite article as its point of departure and is chiefly functional in nature, even though some wordings, "sets of objects" and "refers inclusively", may be associated with formal features. I prefer to keep the formal and functional aspects of definiteness apart, even though both are worth considering. From my perspective, a definite article formally restricts the reference of the NP (in a certain manner), but how to use the definite NP for successful communication is not grammatically coded.

The formal restriction that comes with the definite article is further outlined in Stroh-Wollin 2011, 2015. In short, I take it to follow from a combination of two features, or two steps in a derivation. First, a *set of selection* (S) is defined as a subset of the largest possible set given the descriptive core of the noun phrase, called *the universal set* (U). This means $S \subseteq U$ and nothing more. Then, *the set of referents* (R) is defined in a second step: R refers to the totality of S; thus, $R = S (\subseteq U)$.

This derivation may at first sight appear as a more complicated way to say that $R \subseteq U$, but this is not so. The set of selection is a means to

distinguish between definite and indefinite NPs. Even the set of referents of an indefinite NP may be restricted in relation to the universal set, by force of an indefinite article or a quantifier. (NPs without articles or quantifiers remain formally unrestricted.) Now, the restriction of R is purely quantitative of nature, but the restriction of S is a more complex matter.

This leads us to the functional side of definiteness. A definite article does not bring about any other meaning to the NP than the one captured in the formulas above. But, in a discourse, the restriction of S in relation to U must be understood in one way or another, and the interlocutors should agree on how. To that end, I find it a reasonable description that the understanding is based on shared speaker-hearer knowledge, an umbrella term that works for various kinds of referential categories (cf. Becker 2021:64: "mutual and unambiguous identifiability").

It is also worth noting that understanding the restriction of S does not entail specific knowledge of the referent, and this is as it should be. Of course, the hearer often has a personal acquaintance with the intended referent of a definite NP, but this is far from always the case. For the sake of simplicity, however, in the following I will talk about the recipient's task in terms of "identifying" the referent or "interpreting" or "understanding" the reference of the definite NP.

A point in Hawkins' description of the definite article that I have reason to return to is its function as an "instruction" to the hearer to locate the referent; as we will return to, I find it important that some referents are actually identifiable without the article.

3.2 From demonstrative to definite article. When has an article emerged?

It is generally assumed that definite articles normally arise through the grammaticalization of demonstratives. The phenomenon has received much attention (to say the least), both from a general, cross-linguistic, perspective as well as in research on individual languages or language families. (For overviews and further references, see e.g. Himmelmann 2001, De Mulder and Carlier 2011, Kuteva et al. 2019:137 ff., Szczepaniak & Flick 2020.)

One basic question is how the crucial transition comes about, i.e. how the (former) demonstrative is taken into use in contexts that are not compatible with a demonstrative reading. As demonstratives as well as definite articles are possible not only in anaphoric uses but also when talking of referents that are present in the immediate surrounding, both contexts

could serve, at least theoretically, as so-called bridging contexts (Lyons 1999:332). De Mulder & Carlier (2011:530 f) point to the recognitional uses as an explanation for the fact that definite articles usually originate from distal rather than proximal demonstratives. The recognitional use of a demonstrative is illustrated in (10), from Gundel et al. (1993:278).

(10) *I couldn't sleep.* ***That dog*** (*next door*) *kept me awake.*

The recognitional use is different from other uses of the demonstrative, first, because they are exclusively used adnominally and, secondly, because they do not have referents in the preceding discourse or in the immediate physical surrounding, but rather "activate specific shared knowledge" (Diessel 1999:105). According to Mulder & Carlier (2011:531), the distal demonstrative "becomes a definite article when the anchorage in the speech situation is lost".

When it comes to the Scandinavian post-nominal definite article, it is not possible to determine with certainty whether the origin of the article was a distal or proximal demonstrative; according to Stroh-Wollin 2020, it was rather the latter.[7] Nor is it possible to explore the first transitional phase of the grammaticalization. The loss of deictic power seems completed already in the earliest documents. These also show uses that are not compatible with demonstratives. In fact, this even applies to the very first instances, the two runic inscriptions rendered in (5). However, in the oldest texts, the anaphoric uses dominate greatly, so it is an educated guess that anaphoric uses provided a good context for reinterpretation, which does not exclude that it also took place in so-called immediate situation uses.

Now, as indicated in the introduction, the focus of this study is not on the emergence of the definite article but on its later expansion. The word *emergence* as well as the word *transition* used above imply (correctly) that I consider the reinterpretation of the demonstrative as a definite article as an initial phase of the grammaticalization. Some linguists claim that an article cannot be distinguished from a demonstrative unless it is used *consistently* in some contexts where demonstratives are not possible, so e.g. Himmelmann (2001:833), Becker (2021:42 ff.), Goldstein (2022:7).

[7] The *hinn* behind the article in Mainland Scandinavian (cf. section 2) should not, as sometimes happens, be confused with the demonstrative *hinn* used in the sense of 'the former' or 'the other' and occasionally (alongside the demonstrative corresponding to Engl. *that*) as determinative before relative clauses. The latter word appears consistently as *hinn* even in the earliest Icelandic manuscripts, while the former corresponds to *enn* in the same documents.

Also, Skrzypek et al. (2021:11) has chosen the label "incipient article" for occurrences during the development towards mandatory use. However, for instance Schlachter (2020:19) claims that the grammaticalization combines "abrupt reanalysis" and "analogical extension", a description I agree with. For the Scandinavian case, I consider the reanalysis as a fact in all preserved medieval documents, since already in the earliest texts it appears in contexts incompatible with demonstratives, which I take as a sufficient criterion for speaking of a true definite article. Thus, I do not attach importance to the fact that the article is not actually used consistently in any particular context in the examined texts.

Another sign of a completed reanalysis is that the article always occurs to the right and in direct contact with a host noun. Admittedly, the early article *enn* is occasionally written as a separate word in some West Norse (chiefly Icelandic) texts, see e.g. (11), from the Icelandic Homily Book, cited after de Leeuw van Weenen (1993:44v) but with somewhat simplified orthography. However, even if we see such written instances as evidence of a real phonetic unboundedness (which I think we can do; see further Stroh-Wollin 2020), this does not mean that the article was not an article, but suggests that the phonetic fusion (and further morpho-phonological changes) came gradually and (rather wholly than partially) after the reanalysis.[8]

(11) *sa* *þykkir* *elldr* ***enn*** *héitastr* *es* *a* *siólfom* *liggr*
DEM thinks fire DEF hottest REL on self lies
'That man finds the fire hottest who is in it himself.'

3.3 Comments on earlier studies on the development of definiteness marking in Scandinavian

The development of definiteness marking in Old Scandinavian is not an unexplored area, but I will confine myself here to commenting on a couple of more recent works with specific relevance in this context. While my interest here concerns the diachronic development as it appears in different parts of Scandinavia, previous researchers have often focused on a

[8] The non-fused article raises the question of what categorial status it should be given. One possibility would perhaps be *non-bound affix*, since it was not a free lexeme in the normal sense but, just like its later suffixed successor, needed a noun host to its left. I prefer not to talk about clitics in a formal sense either for the unbound articles or articles in an early stage of cliticization, i.e. in a purely descriptive, phonetic, sense. (For another view, see Faarlund 2009.) Lately, the concept of *clitic* has been strongly questioned, not least as an intermediate stage in grammaticalization (cf. e.g. Haspelmath 2022). However, I leave this problem aside here, as a decisive answer is not at stake in this context.

single ancient Scandinavian variety from a relatively synchronous, albeit historical, perspective. An exception, however, is Skrzypek et al. 2021, which I comment briefly below in section 3.3.1, also to clarify how my own study relates to this investigation. In section 3.3.2, I discuss Leiss 2007 on definiteness marking in Old Icelandic, which deserves special attention due to its slightly different approach to the development than is usually assumed – and to the one that I myself, after all, mainly land on.

3.3.1 *Skrzypek et al. 2021*

Skrzypek et al. 2021 deals with both definiteness marking with definite articles and indefiniteness marking with indefinite articles. The study is mainly based on statistical analyses of an annotated corpus consisting of excerpts from Swedish, Danish and Icelandic texts from three periods: 1200–1350, 1350–1450 and 1450–1550 respectively. The statistical analyses intend to clarify which factors favour the use of the definite and indefinite articles. The comparisons concern e.g. NPs with count nouns vs NPs with mass nouns, NPs in the singular vs NPs in the plural, subjects vs objects and NPs representing various referential categories. The clearest result regarding the definite article is that its use is favoured by anaphoric reference, which fits well with the general picture that the development begins in anaphoric contexts. Other factors may also play some role at some period of time or in a certain variety, but here it is more difficult to establish general patterns.

The authors take the referential categories presented in Figure 1 as the point of departure in their statistical analysis regarding the development of the definite article, but it is also suggested in a following chapter that it may be relevant to establish a more fine-grained division of referential categories. Most convincing is the follow-up investigation of the indirect anaphora, where the authors compare three sub-categories of indirect anaphora (from M. Schwarz 2000, see further below).

My study differs from Skrzypek et al. 2021 in several ways. For instance, I concentrate on a shorter time span, namely on texts up to the middle of the 14th century. At this point, one can anticipate that the post-nominal article is on the way to becoming mandatory all over Scandinavia, although the final consolidation will take some time. If the development can be considered to follow the S-curve pattern, the later slow phase has probably been reached at this time, which means that the more dramatic development takes place before 1350. It can also be added that some texts (in my study as well as in Skrzypek et al. 2021), above

all some of the early provincial laws, can partly be thought to represent 12th-century language rather than 13th-century language.

Further, I have not brought together texts or excerpts of texts in groups either with respect to their time of creation or to their regional provenance. This kind of organisation of the material and a statistical analysis may of course have its virtues. However, it can also lead to specific differences between individual texts escaping the researcher. Instead, I have manually examined all the texts individually and noted for each its date and place of creation. Not only Swedish, Danish and Icelandic but also Norwegian texts are included in my material. While Skrzypek et al. bring together shorter text extracts in their corpus, I have chosen texts or text extracts of at least 5000 words as far as possible to get an idea of individual scribes' language use. This has proven important because, as we shall see, article usage can also differ across contemporary writers within a certain region.

3.3.2 *Leiss 2007*

Leiss 2007 (see also Leiss 2000) traces the origin of a definite article in the Germanic languages to changes in their aspectual systems. This assumption is logical from a typological point of view, as there is an affinity between definiteness and perfectivity (and between indefiniteness and imperfectivity) and languages tend not to combine morphological definiteness marking with morphological aspect.

As for the Scandinavian languages, it has been objected that Scandinavian lost perfective marking long before the emergence of the definite article (see e.g. Perridon & Sleeman 2011:4), but Leiss' account is not restricted to explicit morphological marking but encompasses invisible and complex coding of determination and aspect as well. This means that (in)definiteness and (im)perfectivity can also be expressed by means such as word order and case. However, even from this point of view, one must assume successive shifts in the coding after the loss of the aspect markers and, with that, an increasing use of the original demonstrative that eventually lead to the emergence of a definite article.

The newly born article in Old Icelandic was used, according to Leiss (2007:97), to mark rhematic objects as definite, regardless of the previous discourse. Its function was not, unlike that of the demonstrative, to refer anaphorically. This claim is supported by a sequence from Snorri's *Heimskringla* where recurring mentions of a bull appear in the definite form only as objects in rhematic position, but not as clause-initial subjects. The non-marking of the subjects is explained by the inherent

definite effect of the clause-initial thematic position and demonstrates the "logic of hypo-determination", according to which, somewhat simplified, redundant definiteness marking is avoided (Leiss 2007:88).

This assumption is appealing, but it is at the same time hard to substantiate, one reason being that it is more complicated than it is presented to determine whether the article use is redundant or not with regard to the principles of "invisible syntax". For instance, the varieties of Old Scandinavian, Old Icelandic included, do not show a stable pattern when it comes to definiteness marking, but a constant development from very sparse occurrences of the definite article in early texts to an extensive use in later texts.

Moreover, one can easily find clear counterexamples to the pattern in the sequence that Leiss uses to underpin her theory, even in the same text, see e.g. (12). In this example, there are five instances of the noun *steinn* 'stone' (beside the name of the farm). The first mention (underlined) is indefinite and thus, of course, not marked for definiteness, but the four following ones (in bold face) all come in the definite form – even the last instance, which is a clause-initial subject. The noun *dvergr* 'dwarf' also appears in the definite form at the second mention, likewise a clause-initial subject.

(12) *… í austanverðri Svíþjóð heitir bœr mikill at Steini; þar er steinn svá mikill sem stórt hús. Um kveldit … þá er Sveigðir gekk frá drykkju til svefnbúrs, sá hann til* ***steinsins****, at dvergr sat undir* ***steininum****. Sveigðir ok hans menn váru mjǫk druknir ok runnu til* ***steinsins****.* ***Dvergrinn*** *stóð í durum ok kallaði á Sveigði, bað hann þar inn ganga, ef hann vildi Óðin hitta. Sveigðir hljóp í* ***steininn****, en* ***steinninn*** *lauksk þegar aptr ok kom Sveigðir eigi aptr.*

(the Ynglinga saga, ch. 12; *Heimskringla*)

'In eastern Svithiod there is a large farm called Stone; there is a stone as big as a house. In the evening … when Svegdi went from the drinking bout to the sleeping house, he looked towards **the stone** and saw that a dwarf was sitting under it. Svegdi and his men were very drunk and ran away to **the stone**. **The dwarf** stood in the doorway and called Svegdi, and asked him to go in there, if he wanted to meet Odin. Svegdi ran into **the stone**, but **the stone** immediately closed behind him, and Svegdi did not come back.'

We may also note that in (12) there are three NPs without co-referent antecedents that are unmarked for definiteness in the Old Icelandic text but are rendered as definite NPs in my English translation. In the case of

drykkju 'the drinking bout', which is not in any way anticipated in the preceding discourse, we can take the identification of the referent as just "script-based", whereas *svefnbúrs* 'the sleeping house' and *durum* 'the doorway' are identified in relation to the farm and the stone respectively (even though stones do not normally have doors). Thus, while all the marked NPs in bold face in this sequence are direct anaphora, the three unmarked definite NPs are not.

It may be, as Leiss argues, that anaphoricity did not trigger use of the definite article from the very start of its appearance. Nevertheless, the example in (12) suggests that it was associated with anaphoric reference at an early stage and that its use for other kinds of referential categories lags behind. On closer inspection, this sequence also appears to be more typical of *Heimskringla* than the one Leiss cites. So, even though the idea of early definiteness marking as governed by invisible syntax is thought-provoking, I find it justified to focus in what follows on the spread of the definite article from its anaphoric uses to other referential categories.

4. Referential categories for definite NPs: theoretical considerations

The theoretical discussion in this section is focused on the question of which referential categories there is reason to identify for definite NPs. First, section 4.1 accounts for the parts of Löbner 2011 that are important in the further reasoning, while the following section, 4.2, briefly addresses some other recent studies and their division of definite NPs into different referential categories. Section 4.3, finally, is a critical discussion of the views presented with the aim of clarifying the principles that should form the basis of the empirical analysis.

4.1 Definiteness marking, concept types and the scale of uniqueness in Löbner 2011

In recent years, a number of researchers have been influenced by the theory of determination presented in Löbner 2011 (a further elaboration of Löbner 1985) and the author's "scale of uniqueness", which (among other things) offers an alternative view on the diachronic development of definite articles. As already hinted at, I do not take the scale of uniqueness

as the best model for describing the spread of definite articles. However, the set of referential categories I propose depends considerably on the very basic assumptions in Löbner's theory, which are briefly presented in this section. Moreover, the scale of uniqueness has given rise to interesting cross-linguistic, and in this context relevant, observations, to which I will return.

Löbner's theory is first and foremost based on the assumption that nouns can be distributed across four logical types which are distinguished by the possible combinations of two binary features, [U] for inherent uniqueness and [R] for inherent relationality. The four types are sortal nouns ([–U,–R]), individual nouns ([+U,–R]), relational nouns ([–U, +R]) and functional nouns ([+U, +R]).[9]

The different combinations of the U and R features are illustrated in Figure 2. Individual nouns, to begin with, encompass terms of objects, institutions etc. that may refer to particular entities depending on the context of utterance, "including a time index, a location index and a constellation of facts" (Löbner 2011:284). A clear example is *sun*, which has a global uniform reference. In addition, proper names count as individual nouns. Thus, the nouns *sun* and *John* are [+U]. They are also [–R] as they are not inherently related to any other object. Inherent relationality characterizes instead e.g. nouns denoting body parts or kinship terms. Thus, *leg* and *sister* are [+R]. Nouns such as *leg* and *sister* are also [–U] as individuals normally have more than one leg and may have more than one sister. Nouns such as *head* and *mother*, on the other hand, do not solely express inherent relationality, but also uniqueness, as individuals have only one head and only one mother. Löbner calls [+R, –U] nouns "relational nouns" and [+R, +U] nouns "functional nouns". The prototypical subtype of nouns, labelled "sortal nouns", do not express neither inherent relationality nor inherent uniqueness; *tree* and *water* are examples of sortal nouns.[10]

[9] Löbner (2011:281) explains in a footnote that he is not really happy with the term *unique*(*ness*). What he means is rather something like '*unambiguous*(*ness*)'. I will not go further into that discussion here. In addition, the author points out that the U and R features "are not meaning components in the sense of feature semantics but just descriptive abbreviations" (Löbner 2011: 282).

[10] The notion *sortal* is sometimes restricted to count nouns, but not in Löbner 2011.

	[–U]	[+U]
[–R]	**Sortal nouns** *tree*, *water*	**Individual nouns** *sun*, *John*
[+R]	**Relational nouns** *leg*, *sister*	**Functional nouns** *head*, *mother*

Figure 2. The distribution of inherent uniqueness and relationality on four types of nouns according to Löbner 2011.

Inherent relationality means "the presence of an additional 'possessor' argument" (Löbner 2011:285). A leg is somebody's leg and a mother is somebody's mother etc. The specification of the possessor is semantically mandatory but is not always explicitly specified. For instance, when using role terms such as *king*, *president*, *chairperson* etc., the identity of the possessor is very often given exclusively from the external context of the discourse.

Importantly, also, the type distinction is relativized by polysemy (Löbner 2011:282). The noun *moon*, for example, has an individual reading when referring to the moon of the earth, but it is a sortal noun if used in the sense of 'satellite to some planet'. The noun *child* has a sortal reading, 'non-adult', as well as a relational reading, 'daughter or son to somebody', etc.

Further, Löbner (2011:287 ff.) claims that uniqueness and relationality correspond to different modes of determination and that the different types of nouns differ regarding which mode is the "natural" or "congruent" one. For example, the natural determination of unique individual nouns is the definite article, whereas the natural mode of the sortal nouns is the indefinite article. However, the function of determination is, according to Löbner, to bring about sortal, individual, relational and functional "concepts" on the basis of the descriptive content of the NP. As regards the definite article, it has the function to indicate that the NP represents a unique concept (Löbner 2011:289). From this point of view, one can say that a definite article adds uniqueness to a non-unique sortal noun resulting in an individual concept as illustrated in (13a). A definite article together with an individual noun, on the other hand, is indeed used in accordance with the inherent unique meaning of the noun, but it is, precisely because of that, semantically redundant; cf. (13b). (This redundancy

is, according to this view, reflected in languages that do not use definite articles when the NP is inherently unique without; see further below).

(13) a. *the* [+U] definite article + *tree* [–U] sortal noun → *the tree* [+U] individual concept
b. *the* [+U] definite article + *sun* [+U] individual noun → *the sun* [+U] individual concept

The distinction between congruent and incongruent definite NPs is described in Löbner's theory in terms of "semantic uniqueness" and "pragmatic uniqueness" respectively. If the NP is unique because of the lexical meaning of the head noun, it is classified as semantically unique. If its uniqueness is "coerced by determination", it is classified as pragmatically unique (Löbner 2011:307).

Interestingly, there is typological evidence that supports the distinction between congruent and incongruent definites, e.g. the uses of different definite articles in some languages. A well-known example is the Frisian dialect Fering, spoken on the islands of Föhr and Amrum, with a "strong" D-article and a "weak" A-article (Ebert 1970). The strong article is used, with Löbner's terminology, in incongruent definites and the weak article in congruent definites. (14a) below shows the use of the strong definite article in an anaphoric NP in Fering, and (14b) the use of the weak article with an inherently unique noun.

(14) a. *Peetje hee jister an kü slaachtet.* (Fering; Ebert 1970:107)
Peetje has yesterday a cow slaughtered.

Jo saai, ***det*** *kü wiar äi sünj.*
One says DEF.STR. cow was not healthy.
'Peetje slaughtered a cow yesterday. One says the cow wasn't healthy.'

b. ***A*** *san skiinjt.* (Fering; Ebert 1970:71)
DEF.WK sun shines
'The sun shines.'

A distribution of different definite articles similar to that in Fering has been reported from various West Germanic dialects and is also reflected in standard German in the choice between preposition + full (strong) article, such as *in das*, *an dem* and *zu der*, and the corresponding contracted forms, i.e. *ins*, *am* and *zur* respectively. Similar shifts between a strong and a weak definite article have been suggested for non-Germanic languages as well, and a wider outlook also tells us that there are languages

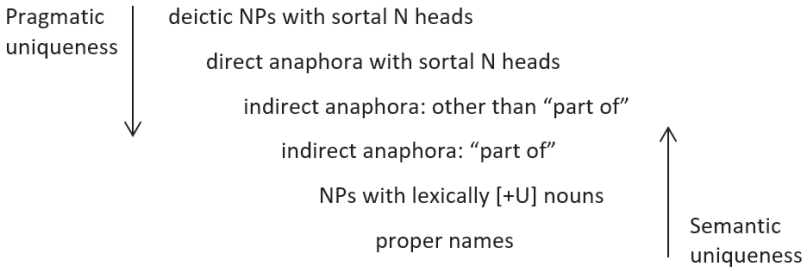

Figure 3. Simplified version of the scale of uniqueness, based on Löbner (2011:320), Ortmann (2014:314) and Czardybon (2017:26).

that differentiate between nouns with definite articles and bare nominals. (See e.g. Lyons 1999:53 f., Schwarz 2009, 2013, Ortmann 2014, but note also that Becker (2021:331) finds the strict division between strong and weak articles found in some Germanic varieties unusual.)

Löbner does not take such asymmetries concerning definiteness marking as absolute, but places different kinds of NPs along an implicative scale of uniqueness from pragmatic to semantic uniqueness. In his hierarchical model, Löbner places deictic and anaphoric NPs with sortal nouns at the upper, pragmatic, end and NPs with inherently unique nouns, proper names and personal pronouns at the semantic end at the bottom, while indirect anaphora are found in between (Löbner 2011:320). Figure 3 is a simplified model where I leave out NPs with modifiers, as here I am exclusively focusing on NPs without. I also disregard personal pronouns, which rarely take definite articles, although this does occur in some languages. Also, the division of the indirect anaphora in Figure 3 is in line with the versions of the scale in Ortmann 2014 and Czardybon 2017. Both authors refer to Schwarz 2009 regarding the indirect anaphora; see further below.

The implicative ordering of the different kinds of NPs in the model indicates that cross-linguistically languages are more apt to use definite articles with NPs higher on the scale than with the lower ones. Likewise, languages with strong and weak articles associate the strong article with pragmatic uniqueness and the weak article with semantic uniqueness. The scale of uniqueness is also supposed to indicate how languages that develop definite articles start out by stretching the use of demonstratives

for anaphoric reference with sortal nouns and gradually expand the use to new kinds of NPs along the scale.

4.2 More on referential categories for definite NPs in recent research

Current research often presents a more fine-graded set of referential categories for definite NPs than the one in Hawkins 1978. In particular, indirect anaphora now seem to be generally divided into subcategories. As mentioned, both Ortmann (2014) and Czardybon (2017) have implemented the division proposed in Schwarz 2009 in their versions of Löbner's scale of uniqueness, and Becker (2021) make a similar distinction in her study of articles in the world's languages.

Schwarz (2009; see also Schwarz 2013) focuses especially on languages that differentiate between strong and weak definite articles. According to Schwarz, it seems as though there is a typical cut-off point between indirect anaphora expressing part-whole relations and indirect anaphora expressing producer-of-product relations. Languages that distinguish between strong and weak definite articles often use the weak article for part-whole relations, but the strong for producer-product relations. In standard German, for example, the difference manifests itself when the definite article appears after a preposition, as in (15a–b). In (15a), which contains an NP expressing a part-whole relation, a reduced form of the definite article is cliticised to the preposition, but in (15b), which contains an NP expressing a producer-product relation, the full article is used (Schwarz 2009:52 f., 2013: 542).

(15) a. *Der Kühlschrank war so groß, dass der Kürbis problemlos* ***im*** (***#in dem***) ***Gemüsefach*** *untergebracht werden konnte.*
'The fridge was so big that the pumpkin could easily be stowed **in the crisper**.'

b. *Das Theaterstück missfiel dem Kritiker so sehr, dass er in seiner Besprechung kein gutes Haar* ***an dem*** (***#am***) ***Autor*** *ließ.*
'The play displeased the critic so much that he tore **the author** to pieces in his review.'

On the basis of the choice of definite article in languages with two, Schwarz speaks of "two kinds of definites". Definite NPs with the strong

article, whether in direct or indirect anaphora, are associated with anaphoricity and those that take the weak article with uniqueness.[11]

From her typological perspective, Becker (2021) does not speak of strong articles, but of anaphoric articles. Anaphoric articles roughly correspond to what has been identified as strong definite articles in some Germanic varieties, but in other languages, they are not necessarily used in opposition to a weak article, but sometimes alongside a generally functioning definite article (Becker 2021:331). Weak definite articles, on the other hand, seem to occur only in contrast to an anaphoric/strong article (Becker 2021:98).

A more complex division of indirect anaphora is found in Schwarz-Friesel 2007 (see also M. Schwarz 2000). Schwarz-Friesel proposes a primary distinction between semantic and conceptual indirect anaphora and a further distinction within these subcategories. The primary distinction is cognitively motivated; the use and interpretation of the semantic types are taken to depend on activation of lexical knowledge, but the conceptual types on activation of more general world knowledge.

The semantic subcategory encompasses NPs expressing part-of-whole relations, as e.g. between *the crisper* and *the fridge* in example (15a) above, and NPs expressing thematic roles as in (16) below, where I cite the English translation of example (6) in Schwarz-Friesel 2007. Here, the definite NP *the key* (in bold face) expresses the INSTRUMENT with which one can unlock a door, and the verb *unlock* (underlined) is taken for the anchor.

(16) *I wanted to <u>unlock</u> the door quickly, because I could hear the telephone ringing.* ***The key****, however, was buried deeply in the trolley.*

Example (17) below, which renders the English translation of example (7) in Schwarz-Friesel 2007, illustrates the kind of the conceptual indirect anaphora that depends on so-called script knowledge. Here, the mention of a restaurant is taken to evoke a conceptual frame where food and waiters are natural elements, and, thus, the noun *restaurant* (or the NP *a lovely restaurant*) functions as anchor for the indirect anaphora *the food* and *the waiter*.

(17) *I know a lovely <u>restaurant</u> in Refrath.* ***The food*** *is excellent and* ***the waiter*** *is an extremely nice guy.*

[11] Possibly such a strict division is challenged by later work of Schwarz (2019) that shows somewhat more variation in article distribution across languages.

There are also, according to Schwarz-Friesel, some conceptual indirect anaphora that do not refer to particular parts of a frame, but require more complex inferencing.

Later, Irmer (2011:235) stresses that there are many mixed cases. "In particular", he writes, "it is not clear where the borderline between thematic semantic and frame-based conceptual indirect anaphora has to be drawn". Irmer lands in a two-part classification with, on the one hand, mereological, and, on the other hand, frame-related indirect anaphora (or bridging anaphora, as is the term preferred in this work). Mereological indirect anaphora are, according to Irmer, anchored in (sets of) entities and typically express part-of and member-of relations, whereas frame-related indirect anaphora are anchored in eventualities and typically express thematic, causal and spatial relations. (Irmer 2011:236 f.)

In terms of a systematic review of different types of referential categories, Becker 2021 appears to be the most comprehensive study. Becker defines various kinds of articles in the world's languages based on which referential functions they can fulfil (see further Becker 2021:86 f.). For the definite domain, Becker (pp. 65–83) identifies nine referential functions by combining distinctions made in the literature (chiefly Hawkins 1978, Himmelmann 1997, Löbner 1985 and Schwarz 2009). In this context, we can ignore the recognitional and establishing functions. The former is hardly found in written documents, which we must rely on when investigating medieval language, and the latter is most often found in NPs with restrictive relative clauses, which are not part of this study.

This leaves us with seven referential functions for the definite domain, two of which are the generally assumed deictic and (direct) anaphoric functions and another two refer to two kinds of indirect anaphora in line with Schwarz 2009. Then we have the "situational unique uses" and the "contextually unique uses". The former of these categories largely corresponds to Hawkins' immediate situational uses, but is distinguished from deictic contexts. (Becker 2021:71 f.) The latter type corresponds to Hawkins' larger situation uses with the exception of a small set of NPs for absolutely unique entities such as the sun or the moon, which represent their own category as they cannot combine with a definite article in various languages. (Becker 2021:83, 145 ff.) Furthermore, Becker states that there is not always a clear-cut distinction between situational and contextual uniqueness (Becker 2021:74). We will return to these non-anaphoric categories below.

4.3 Towards an alternative view of the division of definite NPs into referential categories

As mentioned in the introduction, I believe that it is reasonable to assume that lexical inherent uniqueness and relationality play a role in the interpretation of NPs' reference. In this section, I discuss the impact of this assumption on the sorting of definite NPs into referential categories. This discussion is not concerned with deictic NPs and direct anaphora, which are straightforward categories in this context.

I will first consider possible subdivisions of non-deictic and non-anaphoric definite NPs. For the sake of simplicity, I will refer to these as "first mentions" (and include in this label Becker's situationally unique, contextually unique as well as absolutely unique categories). In Löbner's scale of uniqueness, we find NPs based on sortal head nouns only among the deictic uses and anaphora at the upper pragmatic part. However, even first mentions can be based on sortal nouns, and it is not obvious how to reconcile them with Löbner's model.

Secondly, I discuss how to handle the indirect anaphora. Indirect anaphora can both have sortal head nouns and head nouns with inherent relationality. In addition, Schwarz (2009) shows that some languages also make a difference in their use of articles within the latter group.

Finally, I comment on possible extensions of article use beyond the first mentions. Recall that Löbner's model has proper names as it endpoint and does not pay heed to NPs with generic reference, while the opposite applies to the model in Figure 1.

4.3.1 *First mentions*

When it comes to first mentions, I will argue that there is above all reason to make a distinction between, on the one hand, first mentions based on nouns with inherent uniqueness- and/or relationality features, because their reference is sufficiently restricted without a definite article, and, on the other hand, first mentions based on sortal nouns.

Let us first consider NPs based on nouns with inherent relationality, i.e. functional and relational nouns. For such NPs, we can actually take the relationality feature as (metaphorically) doing a job that is similar to the one Hawkins assigns to the article, i.e. to instruct the hearer to "locate the referent". Relationality assumes a "possessor", and thus we can say the feature also serves as an appeal to the receiver to "find the possessor" and thereby locate the referent. In cases where the possessor is explicit in the preceding discourse, we can classify the NP as an indirect anaphor, a

situation we will return to below. The possessor of a first mention NP, on the other hand, is identified through the *external context of the discourse*, chiefly defined with regard to when and where the discourse takes place.

Discourse-external identification is especially common with role terms. The statement in (18), for instance, if used in a monarchy with a king, would naturally concern the current king of that same monarchy without further specification.

(18) *The exhibition was opened by* ***the king****.*

When it comes to NPs based on individual nouns, i.e. nouns that do not express inherent relationality but inherent uniqueness, we have a slightly different kind of identification. The uniqueness of individual nouns is not, unlike the uniqueness of functional nouns, related to a possessor. Individual nouns express conventionalized uniqueness. This means that a unique meaning of the word is stored in the mental lexicon of the speakers of a certain speech community. For some nouns, the unique reference is obvious to all speakers of the same language. Any English-speaking person could ask another English-speaking person the question in (19a) on a cloudy day. However, conventionalization can also occur in restricted speech communities, even very small ones. A family, for instance, may have established the appellation *thingy* as an individual noun for the remote control to the TV, in which case the definite NP in (19b) would be unambiguous among the family members.

(19) a. *Where is* ***the sun****?*
b. *Where is* ***the thingy****?*

We may note that first mention NPs of the kinds demonstrated above have lasting reference within given restricted domains or the relevant speech communities. This means that we use NPs such as *the king* and *the thingy* with unaltered reference in various utterances and situations. *The thingy*, for example, would refer to the same object as *the thing*y in (19b) if someone in the same family pointed out that "the thingy needs a new battery", and *the king* would refer to the same king as in (18) if someone in the same monarchy said that "the king has abdicated" etc. Lasting reference does not mean, however, that the reference is necessarily linked to a specific individual or object over time. The individual noun *pope*, for instance, refers to different men at different points in time.

From this exposé, we can conclude that the interpretation of definite first mention NPs based on individual, functional, and relational nouns

does not in fact need to be prompted by a definite article; the reference of the head nouns is already sufficiently restricted due to the inherent relationality or uniqueness features. In this sense, it is possible to argue that relationality and uniqueness make the definite article semantically redundant.

When it comes to first mention NPs based on sortal nouns, there is no lexical feature in the head noun that "instructs" the hearer to identify the referent in the external context of the discourse. This means that the receiver must infer in some other way which referent is intended, and we can take the definite article as an "instruction" to do so.

In some cases, the identification of the referent rests on specific shared sender-receiver knowledge. This specific knowledge may be limited to a certain circle of people, see e.g. (20a), but it can also be about "common knowledge", i.e. knowledge shared by a wider circle, see (20b).

(20) a. ***The car*** *needs to be washed.*
b. *When Eve picked* ***the apple****, she …*

First mentions based on sortal nouns may also be understood from the specific situation of the utterance, i.e. in a more narrow sense than just a specification of time and place, or by the situational frame given by the discourse-internal context. In such cases, the receiver's understanding does not rest on specific knowledge of the referent. The utterance in (21a), for instance, could be used as an excuse for a late arrival when entering a room where a meeting has already started. In this case, the first mention of *the train* is acceptable simply because it is general knowledge that people going to meetings at some distance from their home or their ordinary place of work etc. may need transport of this kind. Another case is (21b), which reproduces a part from the English translation of example (12) above. Here, *the drinking bout* is still acceptable because the hearer can infer that it was a regular event at the end of day on an Icelandic farm during the Middle Ages.

(21) a. *Sorry,* ***the train*** *was late.*
b. *In the evening … when Svegdi went from* ***the drinking bout*** *to the sleeping house …*

Also, unlike Hawkins (1978:120), I see no reason in distinguishing between first mentions depending on whether they refer to something in the immediate physical surrounding or not, provided that the referent is not visible or in some other way directly accessible to the hearer. For

instance, a sign on a gate that says "Beware of the dog" requires the same kind of inference skills from the receiver as the examples in (21a–b).

To sum up: First mentions can indeed be NPs with sortal head nouns in addition to NPs with inherent relationality and/or uniqueness. In the case of the latter type of NP, the lexical relationality and uniqueness features signal that the referent is readily identified through the external context of the discourse. First mentions based on sortal nouns, on the other hand, need the article as an appeal to the addressee to identify the referent – through specific shared speaker-hearer knowledge or from the situation implied by the utterance itself or the verbal context that it is part of.

4.3.2 *Indirect anaphora*

An indirect anaphor is, unlike a direct anaphor, not linked to a co-referential antecedent but is, unlike first mentions, understood in relation to a specific anchor, a conceptually related word or phrase, in the preceding discourse. This textual anchoring defines what an indirect anaphor is, but, as we have seen, there are also suggestions in the literature for distinguishing between different types of indirect anaphora.

We may first consider that the head nouns of indirect anaphora can be either, on the one hand, relational or functional nouns, which possess inherent relationality or, on the other hand, sortal nouns. As argued above, it is possible to see inherent relationality as a lexical means on par with the definite article "to instruct the hearer to locate the referent". Thus, one obvious hypothesis could be to make a fundamental distinction of indirect anaphora based on whether the head noun possesses a relationality feature or not. There is a problem, though, considering the distribution of strong and weak definite articles that Schwarz (2009, 2013) reports for two kinds of indirect anaphora that both express relationality.

Overall, it appears in the literature that NPs that express part-of (something) and the like stand out as a clear subcategory of indirect anaphora. In a way, this confirms that there is a difference between the two kinds of NPs discussed in Schwarz 2009, which suggests that one could talk about two kinds of relationality. Another sign of this, besides the distribution of strong and weak articles, is that the requirement for an explicit possessor is not very strong when it comes to e.g. NPs based on nouns for the producer of a product, such as *author*. The anchor of an NP of this kind is not necessarily an expression for the possessor but can be some other conceptually related expression, see (22).

(22) *We were invited to a reading aloud event, but* ***the author*** *never showed up.*

Thus, it is possible to assume that certain relational/functional nouns are similar to sortal nouns because their R feature only "half-heartedly" works as an "instruction" to locate a possessor.

To keep the two kinds of relationality apart, I will talk of prominent and non-prominent relationality. However, then the question of which relational/functional nouns should be associated with prominent relationality and which should not also needs an answer. To this end, I take some help from Schwarz 2009 (pp. 223–225, especially Figure 5.1) and suggest there are two main variants of prominent relationality. One is, not surprisingly, associated with the part-whole relations. The other concerns nouns for relations between two entities that both are part of a third. The latter kind includes e.g. kinship terms.

If this is on the right track, we can assume that the definite article is (more or less) semantically redundant in indirect anaphora based on nouns with prominent relationality, but not in indirect anaphora based on sortal nouns or nouns with non-prominent relationality. This division of indirect anaphora seems to be close to the one proposed in Irmer 2011 (cf. above).

4.3.3 *Extensions beyond first mentions*

As noted above, the model in Figure 3 places proper names at the semantic end of the scale of uniqueness, while the model in Figure 1 has the generic uses at the rightmost end. I take both extensions from first mentions as logical, but only if we distinguish between different kinds of first mentions as suggested above.

The extension from NPs with "ordinary" individual nouns to proper names as on the scale of uniqueness makes sense, because the borderline is anything but clear. This is noticeable in the fact that different languages apply different conventions when it comes to the representation of NPs on the borderline. For example, English readily uses bare nominals for *Heaven*, *Hell* and *Paradise*, often with an initial capital letter in writing, whereas the corresponding words in Norwegian and Swedish take the definite form when referring uniquely. They are also generally written with small letters in Scandinavian – even in Danish and Icelandic, where, as in English, they do not take the definite article. Also, the conventions within one and the same language may vary in a non-predictable way, as when English has *Heaven*, but *the earth*, and *God*, but *the devil*.

Place names, which are names by definition, sometimes appear with the definite article, in some languages more often than in others. Sometimes the reinterpretation from sortal noun to individual noun to name is transparent, as e.g. with *the Netherlands*. In some varieties, also personal names appear with so-called proprial articles. Proprial articles sometimes coincide with the ordinary definite article, as e.g. in German (*der Karl und die Brigitte*, cf. *Charles and Bridget*), but this is not always the case; the proprial articles found in northern Mainland Scandinavian vernaculars and in (informal) Icelandic, for instance, originate from the personal pronouns for 'he' and 'she'.

I will not dwell here on the formal differences between individual nouns and names. Suffice it to say that both categories express a conventionalized unique reference.

Generic NPs, on the other hand, are not based on nouns with inherent conventionalized uniqueness (and/or relationality), but normally on sortal nouns. Thus, it is natural to look for a link between first mention definite NPs with sortal nouns and the generic uses. Furthermore, a generic interpretation does not rest on any specific shared speaker-hearer knowledge, so we should chiefly consider the kind of first mentions whose interpretation is situation-based. We can also here identify a borderland, namely in utterances where the situation does not concern a certain event at a specific point of time, but a type situation that can occur at multiple and unspecified times.

The passage in (23) from the medieval Icelandic version of the *Alexander saga* may illustrate this idea. Here, the young Alexander is compared to a lion cub in a certain kind of situation. The anchoring in this kind of situation licenses the use of the definite NPs *the lion cub* and *the deer* in the English translation. However, in the Icelandic sequence, only *hjortinn* 'the deer' comes in the definite form; *léonshvelpr* 'lion cub' is unmarked for definiteness. The definiteness marking of one of the words in Icelandic justifies the English translation here. But, interestingly, the sentence would be equally fine with the indefinite NPs *a lion cub* and *a deer*, as the clause is not about a specific experienced event. The same potential for switching between definite and indefinite NPs also applies to so-called kind-referring generic NPs in many languages.

(23) *… ok [hann] neytir í huginum vápna sinna með snarpligum áhlaupum, sem þá, er* ***leónshvelpr*** *sér* ***hjortinn*** *fyrir sér, er hann hefir eigi tekit afl sitt, en tenn eru svá litlar, at hann má eigi bíta …*

'… and [he] uses in his imagination his weapons in bold attacks, like when **the lion cub** sees **the deer** in front of him, before he has grown strong and his teeth are so small that he cannot bite …'

It would take us too far afield in this context to discuss all kinds of generic uses (which is indeed a complicated matter, see e.g. Krifka et al. 1995) and to what extent different languages use definite NPs for various kinds of generic reference. The point is that the extension of definiteness marking of NPs in generic uses is different from the extension to definiteness marking of names. The latter is easily understood as there is no clear-cut division between individual nouns and names. But NPs with individual nouns are not a natural basis for generic uses. Instead, we find borderline cases between first mentions and generic uses of NPs with sortal head nouns.

5. Empirical investigation

This section accounts for the empirical investigation of definiteness marking in early Scandinavian. Subsection 5.1 presents the material, while the two following subsections are devoted to methodological aspects, 5.2 to the identification of the referential categories to be used and 5.3 to the excerpting of the definite NPs and the subsequent sorting of these. Subsection 5.4 accounts for the results, while 5.5 summarizes the conclusions and provides some comments.

5.1 Material

The empirical investigation is based on NPs excerpted from 22 medieval texts from Iceland, Norway, Sweden and Denmark. Only texts up to *c.*1350 are considered. For methodological reasons, the texts have been divided into two groups, one group of texts with little or moderate use of definiteness marking and one with texts where the development has gone further. In order to get a reasonably reliable picture of each of the authors' definiteness marking, it was assumed that it was desirable to excerpt samples of at least 5000 words. Some samples are considerably

longer. However, in the end, some shorter texts were also included (see below).

The former group of texts consists mainly of provincial laws. These are spread across regions all over Scandinavia, but often show a very scant use of the definite form. Moreover, bare nominals in the laws are in many cases ambiguous between a definite and an indefinite reading, which makes it impossible to calculate the proportion of semantically definite NPs that are formally unmarked for definiteness. Beside the legal texts, this group also contains a short narrative, the *Guta Saga*, from the isle of Gotland in the Baltic Sea, which like the *Guta Law*, it is written in the local variety, as well as a part of "*The Book of Herbs*" by the Danish medical writer Henrik Harpestreng (lit. 'harp string'), canon at Roskilde Cathedral on Zealand, dead 1244.

Texts from periods when the definite article has gained more ground, but is not yet mandatory, are not so numerous and unfortunately unevenly distributed. The texts chosen for this investigation include two Icelandic homilies, some narratives from Iceland, Norway and Sweden as well as a Swedish *speculum regale* and a Danish *Lucidarius*, a dialogue between a disciple and his master on religious and general matters.

Since the use of the definite form does not develop simultaneously across the entire Scandinavian area, it is not possible to describe the former group of texts as early and the latter as late in an absolute sense. In the former group, the town law from Flensburg in southern Denmark (now Germany) was written *c.*1300, whereas some of the West Norse texts in the latter group are dated as early as *c.*1200.

Table 1 provides an overview of the texts with little or moderate use of the definite article, including their provenience, suggested date of origin, manuscript and the samples investigated. In some cases, an exact dating of the text is complicated. A potential problem is always that texts are often not preserved in the original but only in later copies, but I actually take this as less problematic in most cases as long as it is possible to date the conception of the text. Normally, one can assume that a person that is copying a text has no instruction or intention of revising the morpho-syntax at the same time. This does not exclude, of course, that, for instance, some definite forms are added unconsciously when copying. With some legal texts, however, things are more complicated, because it can be assumed that the content has been revised over time. Therefore, some laws are given a rather vague date of origin in Table 1.

Table 1. Investigated texts with little or moderate use of the definite article.

Text	Provenience	Suggested date of origin	Manuscript	Sample(s)
Grágás Law	Iceland	12th century	Codex Regius *c.*1260	I: ch. 1–11; 7 934 words II: ch. 172–187; 5 376 words
Gulathing Law	Norway (west)	12th century?	E. Don. Var. 137 *c.*1250	I: ch. 1–33; 8 211 words II: ch. 103–150; 5 390 words
Eidsivathing Law	Norway (east)	12th century	AM 68 before 1325	Christian section (only remaining); 7 105 words
Uppland Law	Sweden (central)	13th century	Ups. B 12 *c.*1350	I: Church code; 7 231 w. II: Inheritance code: 5 126 w.
Guta Law	Sweden (Gotland)	*c.*1220	Holm. B 64 *c.*1350	ch. 1–39; 10 895 words
Guta Saga	Sweden (Gotland)	*c.*1250	Holm. B 64 *c.*1350	the whole text; 1 810 words
Old Västergötland Law	Sweden (south)	*c.*1225	Holm. B 59 late 13th century	the whole text; 14 636 words
Scania Law	Denmark (east)	*c.*1210	Holm. B 76 *c.*1300	on inheritance, land disputes and theft; 9 077 words
Harpestreng	Denmark (central?)	early 13th century	NKS 66 *c.*1300	the second part of the *Book of Herbs*; 6 139 words
Jutland Law	Denmark (west)	*c.*1240	Holm. C 37 before *c.*1280	book 1; 6 094 words
Flensburg Town Law	Denmark (south)	*c.*1300 (original)	Stadtbuch 2, City Archive, Flensburg	the whole text; 5 922 words

Table 2 provides information on provenience, dating etc. for the texts with more extensive use of the definite article. This group includes some texts from the so-called homily books from Icelandic and Norway, which actually contain not only homilies but also other religious texts. Both books are dated to *c.*1200, but the different texts within the books have been conceived at different times, and the use of the definite article varies considerably. Two homilies from the Icelandic Homily Book (IH) were chosen because of their extensive use of nouns in the definite form, which indicates that they came into being shortly before the book was compiled. In addition, these homilies were considered long enough for my purposes. I keep them apart by labelling them *g* and *n2* respectively in accordance with de Leeuw van Weenen 1993 (section 1:2). The *Miracles of Saint Olav* from the *Norwegian Homily Book* (NH) are also assumed to have been written in close connection with the compilation of the book (Haugen & Ommundsen 2010:15). The corpus also contains samples from another four West Norse narratives: the *Miracles of Bishop Thorlak*, *Heimskringla* and the *Alexander saga* from Iceland and the *Barlaam and Josaphat saga* from Norway.

As for East Norse, there are only three lengthy Swedish texts in prose and one lengthy Danish text representing the relevant time span. The Swedish texts are the *Old Swedish legendary* (*Fornsvenska legendariet*), a collection of biblical works, normally referred to as *MB 1* (*Medeltidens bibelarbeten 1* 'Medieval Bible Works 1'), and the *Konungastyrelsen*, the Swedish *Speculum regale*, of which parts are included in the investigation. These are the so-called Maria Saga in the *Old Swedish legendary*, a part of the introduction regarding the history of the Jews in *MB 1* and the second section of the *Konungastyrelsen*. The *MB 1* is known from two manuscripts, the later of which (dated to 1526) is actually acknowledged to be closer to the original. The *Konungastyrelsen* is known only from a print of 1634, but there is consensus that this print closely reproduces a medieval original. My dating follows Moberg 1984.[12] The Danish *Lucidarius* is dated to about 1350, the later limit of this study, and apart from the laws and works by Henrik Harpestreng, no earlier Danish manuscripts remain, except for a few too-short fragments.

For further details on the texts and editions, see Sources.

[12] The dating is further discussed in Delsing 2000. Delsing suggests an earlier dating for the bulk of the text (including the part investigated here), but this is firmly rejected in Moberg 2000.

Table 2. Investigated texts with substantial use of the definite article.

Text	Provenience	Suggested date of origin	Manuscript	Sample
IH: Homily g	Iceland	*c.*1200	Holm. 15 *c.*1200	the entire homily; 4 723 words
IH: Homily n2	Iceland	*c.*1200	Holm. 15 *c.*1200	the entire homily; 4 053 words
Miracles of Bishop Thorlak	Iceland	*c.*1200	AM 645 *c.*1220	ch. 1–30; 5 272 words
Heimskringla	Iceland	*c.*1230	AM 35 early 14th cent.	Ynglinga saga 1–26 + Harald Fairhair's saga 1–25; 11 296 w.
Alexander Saga	Iceland	*c.*1250	AM 519a *c.*1280	Book 1; 4 696 words
Miracles of Saint Olav (NH)	Norway	*c.*1200	AM 619 *c.*1200	ch. 1–17; 5 247 words
Barlaam and Josaphat Saga	Norway	*c.*1250	Holm. perg. 6 *c.*1275	pp. 1r–10v; 6 962 words
Old Sw. Leg.: Maria Saga	Sweden	*c.*1300	Holm A 34 *c.*1350	the entire saga; 6 520 words
MB 1: Intro	Sweden	*c.*1330	Holm A 1 1526	ed. Thorell 1959:1–22;7 219 words
Konungastyrelsen (speculum regale)	Sweden	*c.*1350	lost (printed 1634)	Section 2; 7 311 words
Lucidarius	Denmark	*c.*1350	AM 76 *c.*1460	Chapter 1–3; 10 676 words

As can be seen from the tables above, the West Norse texts are generally earlier than the East Norse in both groups, which is in accordance with the fact that the development of the use of the definite article in West Norse, as we shall see, is ahead of its development in East Norse.

5.2 Identifying referential categories for an authentic historical corpus

For the empirical investigation, a set of six referential categories has been identified, largely based on the theoretical considerations above, but to

some extent also with regard to what the authentic material actually looks like.

To summarize the theoretical conclusions, lexical inherent uniqueness and relationality, as introduced by Löbner (1985, 2011), must be seen as an important part of the shared sender-receiver knowledge that matters in the identification of the reference of definite NPs. However, previous research has shown that relationality seems to be decisive only in some cases (Schwarz 2009, 2013). I therefore distinguish between nouns with prominent and non-prominent relationality. Nouns with non-prominent relationality will be classed with sortal nouns, with the consequence that indirect anaphora are divided into two categories, one based on nouns with prominent relationality and one based on sortal nouns as well as nouns with non-prominent relationality. Moreover, I distinguish between first mentions based on nouns with inherent uniqueness and/or (prominent) relationality and those based on sortal nouns.

It also becomes clear that, when confronted with an authentic corpus, one must observe certain considerations in respects that theoretical discussions do not always highlight. One problem concerns the borderline between ordinary individual nouns and proper names; another concerns first mentions with sortal nouns whose reference are neither typically specific nor typically generic. As regards the former, some specific considerations have had to be observed because one should not expect the definite article with proper names or name-like nouns (see further below). However, there is very early evidence of e.g. the noun *sol* 'sun' with the definite article, which means that it does not seem relevant in this context to distinguish a specific category for inherently "absolute unique" NPs as in Becker 2021.

When it comes to first mentions with sortal nouns, on the other hand, I have distinguished a separate category for NPs with generic or generic-like reference, provided they would generally be rendered with the definite article in modern Scandinavian. One example is given in (24), where the NP *naturen* (lit. *the nature*) can be said to have an "inferred absolute reference", which means that it refers to some abstract concept as an indivisible entity. This all-encompassing interpretation should not be confounded with the uniqueness of individual nouns. The head nouns of these "absolute" uses are sortal nouns and the absolute meaning is not, unlike the uniqueness of individual nouns, conventionalized. This kind of NPs are rather related to kind-referring generic NPs. In English, this type often seem to appear as bare nominals, while the corresponding

phrases in Scandinavian generally come as nouns in the definite form; cf. the Norwegian example in (24) and the English translation.

(24) **Naturen** er verken ond eller god. (Norwegian)
Nature.DEF is neither evil or good.
'**Nature** is neither evil nor good.'

Another related use of the definite article appears in what we can call "iterative adverbials" in e.g. generic sentences. The Swedish example in (25) demonstrates an iterative adverbial, a PP with an NP-complement that refers to the time of a recurring event. Even NPs of this kind often come as bare nominals in English, while the Scandinavian languages seem more apt to use definite NPs here; cf. below.

(25) Jag brukar sova en stund efter **middagen**. (Swedish)
I use-to sleep a while after dinner.DEF
'I usually sleep for a while after **dinner**.'

The set of referential categories that I have landed on based on these considerations can, in principle, be defined as follows. I will return below with some clarifications.

S1: direct anaphora (which are based on sortal nouns)
S2: indirect anaphora based on sortal nouns or nouns with non-prominent relationality
S3: first mentions based on sortal nouns with specific (and definite) reference
S4: first mentions based on sortal nouns with generic or generic-like reference
R: indirect anaphora based on nouns with prominent relationality
U: first mentions with inherent uniqueness and/or inherent (prominent) relationality

The use of the letters S, R and U in the labelling of the referential categories is of course a mnemonic device, but, which should be apparent, there is no absolute one-to-one mapping with the logical types of nouns proposed in Löbner 2011.

5.3 Excerption and sorting

As already pointed out, the analysis of Old Scandinavian texts concerns semantically definite NPs without descriptive attributes. Two kinds of NPs have been excerpted in all texts of the corpus: NPs that consist of only the head noun with the post-nominal definite article and NPs con-

sisting of a noun (normally without the definite article) + a demonstrative. Recall that the post-nominal article is not related to the surviving demonstratives; the latter correspond to English *that* and *this*, whereas the article has it origin in *enn/hinn*.

Semantically definite NPs consisting only of a head noun without the definite article have been excerpted solely in the texts with more extensive definiteness marking. This is mainly due to genre differences between the provincial laws and other texts in the corpus. As concerns the legal texts, it is often too difficult to determine whether a "bare noun NP" is semantically definite or indefinite. This is usually sufficiently clear in the other texts, but in cases where the definite and the indefinite readings of a bare noun seemed equally good, the NP was not excerpted. For similar reasons, NPs for holidays have not been excerpted. Nouns for holidays often appear with the definite article in modern Scandinavian but less regularly than with "ordinary" nouns, cf. Engl. *Christmas* – Icel. *jól* or *jólin*.

A potential problem is also that bare nouns and nouns with the definite article can appear without a clear referentiality in certain constructions. In present-day English, for instance, the use of the bare noun *bed* in *go to bed* does not refer to any specific bed, while *go to the store* is ambiguous. The NP *the store* may refer to specific store, in which case we have a regular definite NP, but if *go to the store* just means 'do some regular shopping', we have a so-called weak definite (see further, e.g. Carlson et al. 2006). In a language without obligatory articles, both types of NPs are potentially ambiguous. A similar problem applies to some (potentially) frozen phrases. In all these cases too, doubt has had to guide the excerption.

As concerns NPs with demonstratives, only instances with anaphoric reference are excerpted. Other uses are disregarded, i.e. deictic uses in direct speech, instances with reference to something in the external context, such as *this world*, and cataphorical uses before enumerations, e.g. as when *these cases* means something like 'the following cases'.

Then, in the first step of the analysis, each excerpted NP was categorized as an S phrase, an R phrase or a U phrase. Normally, NPs with sortal head nouns are classified as S phrases and NPs with individual head nouns as U phrases due to their inherent uniqueness feature. NPs with relational and functional head nouns, however, are not divided with regard to whether their head nouns in addition to their inherent relationality possess a uniqueness feature or not (cf. Löbner 2011). Instead, they are classified as R phrases if the possessor is given by an explicit anchor in the preceding context, and as U phrases if the possessor is given by the external context of the discourse.

Polysemy and shifts of concept type also play a role for the classification. For instance, *jǫrð*(*in*) is polysemous and can mean either 'the earth', in which case the phrase is a U phrase, or 'the soil', in which case we have an S phrase. In addition, uniqueness and relationality features can sometimes be ignored. Role terms, for instance, are functional nouns, but may be used in NPs that represent sortal concepts, like *a king* in the very beginning of a fairy tale: "Once upon a time, there was a king who …". In similar cases, a subsequent mention of the referent is classified as a definite S phrase.

The main bulk of the U phrases refer to truly unique entities such as 'the sun', 'the world', 'the earth', 'the devil' and 'the church' in the sense of institution or point out a person that is defined by a certain role, e.g. the king or the bishop, when externally identifiable. Some potential U phrases, *himiríki* 'the kingdom of Heaven', *jarðríki* 'the kingdom of the earth', *paradís* 'Paradise', *helvíti* 'Hell' and *dómsdagr* 'Doomsday', always appear as bare nouns in the texts investigated and still so in Icelandic and Danish. They are therefore judged name-like in early Scandinavian and are not excerpted.

U phrases are normally singular NPs based on individual nouns or role terms. But there are some exceptions. The noun *heimsþriðjungr* 'part of the world' (lit. world-third), for example, appears in the *Heimskringla* in the plural (accusative) NP *heimsþriðjungana* with reference to *'the* three parts of the world' (at this time Europe, Asia and Africa). Also, equivalents to *the Father* and *the Son* are counted as U phrases when they appear in the fixed expression *the Father*, *the Son and the Holy Ghost*. Furthermore, NPs with sortal nouns that refer to geographical areas delimited by the external context of the discourse were classified as U phrases as well, e.g. *land*(*it*) when used for the province or the country where the intended readers live.

The R phrases largely consist of NPs for body parts and parts of non-animate entities, e.g. the door of a house, the sail on boat, the root of a flower etc., which all express prototypical part-of-whole relations. But NPs with abstract head nouns denoting the five senses, mental features, feelings or other abstract entities that characterize human beings and animals have also been counted as R phrases. In addition, some NPs denoting parts that together form a whole, e.g. the buildings of a farm, the inhabitants of a town and the warriors of an army, have been given the same classification. When it comes to kinship terms, which can in principle be considered R nouns, polysemy may cause problems in some cases, because a kinship may be obvious but still not very relevant in the

context. In such cases, I have classified NPs with head nouns such as *kona* 'woman', 'wife', *bondi* 'yeoman', 'master of the house', 'husband' or *barn* 'child' as S phrases rather than R phrases. On the other hand, I have equated words for 'heir' with the usual terms of kinship.

It can be noted that some R nouns come with the definite article more often in Scandinavian than in English, where a possessive is preferred in many cases. While in English somebody can lose *his/her leg*, *sight* or *life* (e.g. in an accident), Scandinavians use nouns in the definite form in the corresponding phrases, i.e. literally (*lose*) *the leg*, *the sight* and *the life*.[13]

After the primary sorting on S phrases, R phrases and U phrases, the S phrases were divided into the four types mentioned above. The S1 class, direct anaphora, contains exclusively NPs with co-referential NP antecedents. The S2 class contains NPs that are anchored in conceptually related words or phrases in the preceding context. The anchor may be a noun or a noun phrase as, for instance, when *the smoke* is anchored in (*a*) *fire*, but it is also very common that the anchor is a verb or a verb phrase, e.g. *burn* (*something*) can readily be followed by definite NPs such as *the fire* or *the smoke*. NPs that summarize or characterize the content in one or more sentences in the preceding discourse, such as *the/ these/ those questions* or *the/ this/ that quarrel* also belong to the S2 category.

We find NPs with demonstratives in both S1 phrases and S2 phrases, sometimes strikingly often in comparison with modern Scandinavian. Demonstratives also appear in adverbial phrases that may be taken as implicitly anchored in the here and now of the discourse, such as (*in*) *that parish*, *that morning* etc. Such instances have also been sorted into the S2 class in order to collect all NPs with demonstratives in the "anaphoric" S1 and S2 classes.

The S3 and S4 classes contain first mentions with sortal head nouns whose reference is not given by an antecedent or explicit anchor in the preceding context, which means that the receiver must infer an understanding of the phrase that fits in the current discourse. As for the S3 phrases, the understanding must work with the (type of) situation or milieu depicted in the text. Sometimes the sender assumes that the receiver has specific knowledge to pick the intended referent among other possible referents, e.g. when *the cross* refers to the cross on which Jesus was crucified. Sometimes no specific knowledge is required, but rather

[13] See Piotrowska & Skrzypek 2017 for a diachronic survey of different ways of expressing inalienable possession in Danish and Swedish and Piotrowska & Skrzypek 2022 for a more detailed analysis of medieval Swedish.

general knowledge and a certain ability to draw conclusions, as explained in connection to the examples in (21a–b). The classification S3 has also been used for definite NPs in utterances that describe type situations without specific temporal anchoring, as e.g. in (23) above. Another example is *láfanum* 'the barn' in (26) from a simile in the *n2* homily, which is intended to illustrate how good people (the wheat) should distance themselves from evil people (the chaff). (The example is rendered here with normalized spelling.)

(26) *Er svo og í* ***láfanum****, að kornið liggur undir sáðunum.*
'It is also so in **the barn**, that the wheat lies under the chaff.'

The S3 phrases in type situations can resemble the generic or generic-like phrases classified as S4, but are not used to characterize the referent in any absolute sense. They rather denote something that is typical in a kind of situation or a kind of milieu.

As for the S4 class, one kind of generic phrases have actually been excluded from the analysis, namely plural generic NPs denoting people or the like, such as for instance *svíar* '(the) Swedes'. This kind most often appear without the article in the texts investigated even when the definite form would be the only possibility in Modern Scandinavian. Thus, it may be that definiteness marking in this kind of NP comes rather late and should be investigated separately. A specific reason to exclude them in the present study is that they are numerous in some of the texts and thereby could have too great an impact on the result.

Other types of generic NPs are included in the S4 group. The by far most frequent head noun in these NPs is *maðr* 'male person', 'human being'; it most often comes in the singular but occasionally in the plural. NPs based on abstract head nouns with "inferred absolute reference" (cf. above) appear rather frequently in some texts. One example (with normalized spelling) from homily *g* is rendered in (27). Note that *ástarinnar* 'the love' comes in the definite form.

(27) *Nú ef vér gerum svo, þá höldum vér það boðorð*
Now if we do so then keep we DEM commandment

ástarinnar *að unna Guði betur en sjálfum oss.*
love-GEN-DEF to love God better than self us
'If we do this, we keep the commandment of **love**, to love God more than ourselves.'

The S4 category also includes "iterative adverbials" in generic sentences, as declared above.

Now, there is also the question of how to classify NPs if a referent is mentioned several times. The answer to this question partly depends on what classification the first instance was given. If the first was classified as an S2 or an S3 phrase and introduces a referent into the discourse previously unknown to the recipient, then the second mention will be an S1 phrase. If the first instance was classified as a U phrase, it is likely that the interpretation of subsequent mentions does not depend on the former. If, for instance, the NP *the pope* were to appear several times in a discourse, always referring to the current pope, then the referent would normally be understood just as 'the pope', not as 'the pope mentioned beforehand', so it will still be a U phrase in the second and further mentions. A similar reasoning can also be made for S3 phrases when the receiver can be assumed to have prior knowledge of the referent. If a first instance was classified as S4, a second is also normally labelled S4. If a first instance is an R phrase, following mentions are also seen as R phrases as long as the relation to its possessor is salient. It happens, though, that the referent starts living its own life, in which case repeated mentions are labelled S1.

5.4 Results

5.4.1 *Definiteness marking in texts with little or moderate use of the definite article*

Table 3 presents the number as well as the frequency per 1000 words of nouns with the definite article and nouns with a demonstrative in the texts with little to moderate use of definiteness marking. Forms of the original demonstrative, "demonstrative I" (cf. Engl. *that*, *those*), and forms of the reinforced variant, "demonstrative II" (cf. Engl. *this*, *these*), are separated in Table 3. Note, however, that demonstrative I was less associated with distal deixis than (present-day) English *that*.

Table 3. Definiteness marking in texts with little to moderate use of the definite article. N = number, F = Number per 1000 words.

		Words	Def.		Dem. I		Dem II	
Text sample	Prov.	N	N	F	N	F	N	F
Grágás I	Ice	7 934	65	8.2	54	6.8	0	0.0
Grágás II	Ice	5 376	108	20.1	23	4.3	0	0.0
Gulathing Law I	No	8 211	24	2.9	45	5.5	10	1.2
Gulathing Law II	No	5 390	8	1.5	25	4.6	1	0.2
Eidsivathing Law	No	7 105	26	3.7	30	4.2	1	0.1
Uppland Law I	Sw	7 231	107	14.8	43	5.9	0	0.0
Uppland Law II	Sw	5 126	49	9.6	48	9.4	1	0.2
Guta Law	Sw	10 895	15	1.4	20	1.8	1	0.1
Guta Saga	Sw	1 810	2	1.1	17	9.4	3	1.7
Old Västergötland Law	Sw	14 636	19	1.3	14	1.0	5	0.3
Scania Law[1]	Dk	9 077	93	10.2	10	1,1	0	0.0
Harpestreng[2]	Dk	6 139	25	4.1	7	1.1	21	3.4
Jutland Law	Dk	6 094	10	1.6	9	1.5	1	0.2
Flensburg Town Law[3]	Dk	5 922	13	2.2	41	6.9	3	0.5

[1] Out of 93 instances of the nouns with the definite article, 59 are based on the noun *bonde* 'yeoman', 'master of the house', 'husband'.

[2] The phrase *thænne yrt* 'this herb' appears 14 times in the text, making up two-thirds of the 21 instances of demonstrative II.

[3] Out of 13 instances of the nouns with the definite article, 8 are based on the noun *foghet* 'bailiff'.

Comparisons below between the texts in Table 3 and the regions they represent are based on the frequency figures. However, it must be immediately emphasized that this measure is very rough, since the frequency can also vary widely across texts in a language with mandatory definiteness marking.

Still, large differences across different texts, or different parts of a text, should say something, and a couple of texts stand out through higher frequency than the others, the *Grágás Law* and the *Uppland Law*. The *Scania Law* also gives the impression of a high frequency of nouns in the definite form, but almost two thirds are some form of the noun *bonde* 'yeoman', 'master of the house', 'husband'. Apart from these instances, the *Scania Law* shows a modest use of the definite article, comparable to that of Harpestreng's *Book of Herbs*.

The relatively high figures for the Icelandic law collection, *Grágás*, is especially interesting as it probably reflects a fairly early language (12th century). The substantial difference between the two samples is also noteworthy. The first sample includes the first eleven chapters of the Christian Law section of which the chapters 1 to 17 are explicitly attributed to Bishop Ketill and Bishop Þorlákur and, according to Finsen (1870:35), this part was written 1123. Even though it may have been subject to some revision, it is likely that it can mainly be dated to the first half of the 12th century. The second sample is a part of the section about land transfer and land holding. This section was probably written by someone else and has most likely been revised on several occasions. Thus, the high frequency of definite forms in this part of the law may imply that it rather reflects Icelandic from the latter half of the 12th century.

The *Uppland Law* shows comparatively high figures, too, but is later than *Grágás*. How much later is, however, difficult to judge. The law was given royal assent in 1296, but it is clear from the preface that this version is a compilation of old, partly revised as well as some new paragraphs (see further Stroh-Wollin 2023).

Otherwise, the texts in this group show rather few nouns with the definite article. This holds true for both the early Norwegian texts and the later Swedish and Danish ones. When it comes to the *Gulathing Law*, one could imagine that it reproduces an even older language than *Grágás*. It is assumed that the Gulathing as such was established at the beginning of the 10th century. In the only complete manuscript that remains of the *Gulathing Law*, the scribe marks that certain regulations are to be attributed to King Magnus Erlingsson, who carried out a revision of the law in 1164. These passages are found above all in the Christian Law section, which constitutes the first text sample, and, interestingly, it is also in these that we find the 10 instances of demonstrative II. One could take this as an indication that the rest of the text largely reflects an earlier language stage (when demonstrative II was less established for anaphoric uses), but in that case, it still remains difficult to assess how much earlier.

Even for some of the legal texts given a more specified dating in Table 1, it is quite plausible that even those partly reflect wordings from earlier stages, to some extent handed down orally from one generation to the next. However, the *Jutland Law*, which was given royal assent in 1241, was probably written just before (Stroh-Wollin 2023), and the *Flensburg Town Law* was first written in Latin in 1284 but was translated into Danish shortly afterwards, so we know for sure that *c.*1300 is a reliable dating for this text. The clearly demonstrated weak establishment of the

post-nominal definite article in this area, above all in the late Flensburg Town Law, should probably be understood in relation to the fact that the dialects of south-western Jutland, unlike all other Scandinavian varieties, came to use a pre-posed definite article instead of the post-posed (see further below).

Table 4 presents the numbers of nouns with the definite article and nouns with a demonstrative distributed across the six referential categories. Both kinds of demonstrative are counted together in this case. As expected, we find demonstratives only in the anaphoric S1 and S2 classes. NPs with demonstrative are quite frequent in some of the texts, as can be seen from Table 3, but they are never used without an antecedent or some kind of anchor in the previous context. Modern Scandinavian would sometimes have preferred a definite article or an anaphoric pronoun instead, but no single instance can be judged as completely impossible.

Table 4. The number of nouns with definite article and nouns with a demonstrative distributed across the referential categories.

		S1		S2		S3	S4	R	U
Text sample	Prov.	Def.	Dem.	Def.	Dem.	Def.	Def.	Def.	Def.
Grágás I	Ice	44	39	12	15	0	1	5	3
Grágás II	Ice	89	19	18	4	0	0	0	0
Gulathing Law I	No	15	43	6	12	0	0	1	2
Gulathing Law II	No	6	23	2	3	0	0	0	0
Eidsivathing Law	No	19	26	6	5	0	1	0	0
Uppland Law I	Sw	53	33	31	10	6	1	4	12
Uppland Law II	Sw	26	39	6	10	3	0	14	0
Guta Law	Sw	9	15	2	6	3	0	1	0
Guta Saga	Sw	0	14	0	6	0	0	1	1
Old Västergötland Law	Sw	6	16	9	3	1	0	1	2
Scania Law	Dk	90	9	3	1	0	0	0	0
Harpestreng	Dk	9	24	4	4	1	0	11	0
Jutland Law	Dk	10	8	0	2	0	0	0	0
Flensburg Town Law	Dk	3	36	0	8	0	0	0	10

It is also clear from Table 4 that we find explicit definiteness marking mainly in the S1 and S2 categories, i.e. where we expect the development to start. In the other referential categories, nouns with the definite article are scarcer. Does this mean that definiteness marking in non-anaphoric NPs is still uncommon or is it that these phrase types as such are unusual in the law genre? It may be that there is a predominance for anaphoric phrases among the definite NPs in most of these texts, but normally one can expect a number of S3 phrases and often some R and/or U phrases as well. (In a survey of definiteness marking in modern Swedish prose, Fraurud (1990:405 ff.) actually finds more definite forms in first mentions than in "subsequent-mentions".) The S4 category, on the other hand, is not necessarily represented in all texts.[14]

There is reason to look a little closer at some of the texts investigated. Presumably, the very few cases of S3 phrases with the definite article suggest that definiteness marking in non-anaphoric NPs with sortal nouns lags behind the anaphoric ones. Potential cases are not lacking, see e.g. the bare nouns in bold face in (28) from the Christian Law section of Grágás, which correspond to nouns with the definite article in the English translation. (This and the following example are rendered with normalized spelling after Einarsson et al. 1993.)

(28) *En meðaldaga alla um jól er rétt að moka undan* ***fé*** *og reiða á* ***vøll****, þann hluta vallar er nær er fjósi, ef hann hefur eyki til, og velta þar af. Ef maður dregur mykju út og hefur eigi eyki til og skal þá færa í* ***haug****.*

'But in all days between Christmas and Twelfth Day, it is permitted to clean out the dung among **the cattle** and throw it on **the grassland** – that part of the grassland that is closest to the cattle shed – if one has a draught animal – and dump it there. If one shoves dung out but has no draught animal, then one shall throw it on **the kitchen midden**.'

In contrast, (29) demonstrates definiteness marking in U and R phrases. This passage from the same text sample explains how to know how long it was allowed to work on Saturdays, which is until the sun was "shaft high". The sun is mentioned four times, twice without the article (*sól*) and twice with (*sólina*). Beside these U phrases, there are two R phrases

[14] The law text samples exhibit some NPs in iterative adverbials with the definite article, e.g. *um daginn* 'in the day' in *Grágás I*, with reference to every day a priest sings mass in a certain church. There is one similar example in the *Eidsivathing Law* and one in the Church Code of the *Uppland Law*.

with the definite article in (29), *oddurinn* 'the tip' and *spjótsskaftshalinn* 'the end of the spear shaft', both related to the possessor *spjót* 'a spear' (underlined below).

(29) *Þá er skafthá* ***sól*** *ef maður stendur í fjöru þar er mætist sjór og land* [...] *og mætti hann sjá í haf út þá er* ***sól*** *gengur að vatni enda sýnist honum svo, ef spjót væri sett undir* ***sólina***, [...] *að* ***oddurinn*** *tæki undir* ***sólina*** *en* ***spjótsskaftshalinn*** *á sjó niður* [...].

'Then, **the sun** is shaft high, if a man is standing on a beach where land and sea meet [...] and he can look out to the sea as **the sun** sets against the water and it seems to him, if a spear were placed under **the sun**, [...] that **the tip** would reach up to **the sun** and **the end of the spear shaft** down to the sea [...]'

The sun is actually mentioned 12 times in the whole text sample, three times with the article and nine times without. Thus, it does not seem like the noun *sól* 'sun' is stored with the article in the scribe's mental lexicon. Some R and U phrases are always rendered in the indefinite form in *Grágás I*, e.g. NPs with head nouns such as *faðir* and *sonr* (for God and Jesus when mentioned with the Holy Ghost), *biskup* 'bishop', *frændi* 'relative' and *erfingi* 'heir'. This means that most of the R and U phrases are not marked for definiteness in *Grágás I*. However, a rough estimate shows that this also applies to about three quarters of the S1 and S2 phrases in this sample.

The frequency of S1 and S2 phrases with the article is significantly higher in *Grágás II* than in *Grágás I*. This outcome is logical if, as assumed above, this sample reflects a somewhat later stage of development, but in that case, one might ask why the definite article in R and U phrases is completely absent here. However, the answer is simply that the entire sample only exhibits one single U phrase (*sol* 'the sun') and no R phrases at all. This means that, unfortunately, it is impossible to use this text sample to evaluate definiteness marking in R and U phrases at the time it represents.

Besides *Grágás*, also the *Uppland Law* shows more than sporadic use of the definite article in R and U phrases, but the compilatory nature of the *Uppland Law* makes it difficult to draw conclusions regarding the definiteness marking at any specific time. Even though there are more R and U phrases with the definite article in this law than in the others, most of its R and U phrases still consist of bare nouns, and, as in *Grágás*, this also applies to the S1 and S2 phrases. The R phrases in Harpestreng's *Book*

of Herbs refer mostly to parts of plants, while the numerous instances of body parts in this text almost always appear in the unmarked form.

A couple of the other early texts are also worth a comment. One is the *Flensburg Town Law*, where ten out of thirteen NPs with the definite article are found in U phrases. The king's representative in the town (the bailiff) is referred to as *foghdæn*, i.e. with the definite article, eight times (vs. *foghæt*, i.e. without the article, 23 times), and there are two *bysins* 'the town's' with reference to Flensburg. Otherwise, this text shows an abundant use of demonstrative I, *þæn*, as definiteness marker in S1 and S2 phrases. It has been suggested that the pre-posed definite article *æ* in the vernacular spoken in south-western Jutland has its origin in this demonstrative, and its use in the *Flensburg Town Law* has been assumed to foreshadow such a development (but see Møller 1974 and Ejskjær 2005 for a different view). However, just as in the other early texts investigated, the demonstrative is not used in this text except in direct and indirect anaphora.

In a way, the definiteness marking in the *Guta Saga* shows a pattern similar to that in the *Flensburg Town Law*. It has an abundant use of the demonstrative in S1 and S2 phrases, but only three instances of a noun in the definite form, which can all be considered semantically redundant. Besides *þan wegin* 'that way', where the redundancy is due to the presence of the demonstrative, we have *faroyna* 'the sheep island', i.e. a U phrase (or possibly a name), and *drytningina* 'the queen', presented as the spouse of the king of Greece, i.e. an R phrase.

To conclude so far: The – relatively speaking – extensive use of the definite article in *Grágás* speaks for an early development in Iceland. The very limited use of the definite article as late as around 1300 in the *Flensburg Town Law*, on the other hand, suggests that (southern) Jutland was peripheral to the mainland Scandinavian centre of innovation, and this probably also applies to the isle of Gotland. However, it is difficult to locate the mainland centre (or centres?) of innovation more clearly with the help of the remaining provincial laws, as these were written down at different times. The varying frequency of definite forms in the *Uppland Law* as it appears in the compilation from *c.*1300 suggests that development started to take off during the 13th century in this province, but other laws were recorded in the first half of the century at the latest and the Norwegian ones considerably earlier.

The definite article is mostly found in S1 and S2 phrases in these early texts. It is also used with some frequency in R and U phrases in some of the laws, whereas it occurs more exceptionally in S3 and S4 phrases.

5.4.2 *Definiteness marking in texts with substantial use of the definite article*

When it comes to the group of texts with a more extensive use of the definite article, it has been possible to excerpt the bare noun NPs with semantically definite reading in addition to those with explicit definiteness marking. Thus, the following account of the result is based on the percentage distribution within the different referential categories between NPs that are marked for definiteness and those that are not.

To avoid very frequent types of NPs having too great an impact on the result, some NPs are not included in the figures to be presented. To these belong the numerous instances of *konungr* 'the king' for Harald Fairhair in *Heimskringla*, for Alexander in the *Alexander saga* and for the father of Josaphat in the *Barlaam and Josaphat saga* as well as *konungssunr* 'the king's son' for Josaphat himself. The NPs that refer to the protagonists, i.e. Harald, Alexander and Josaphat, normally come as bare nouns, whereas Josaphat's father is referred to exclusively as *konunginn* 'the king' with the definite article. The Danish *Lucidarius* also causes some problems, as some head nouns appear quite frequently in different parts of the texts. In order to reduce the risk of an overly random outcome, a rather large sample was excerpted. Nevertheless, it still seemed necessary to disregard a large number of NPs based on *menneske* 'human being' with generic reference, which (unlike most other S4 phrases) normally appeared without the article.

The absolute numbers of instances as well as the percentage distribution for each text are given in the Appendix. However, as the number of instances in some texts is small for some referential classes, I will focus on the results of combined classes below. I take this proceeding to be defensible from the point of view that the referential categories can be brought together in groups of two with regard to the primary kind of "shared knowledge" that is activated when identifying the referent. In the case of NPs with sortal head nouns, a basic distinction can be made based on whether or not the referent is identified through the preceding discourse. Such text-internal identification applies to the anaphoric S1 and S2 phrases, and the difference between these classes is normally small (cf. Table B in the Appendix). Only the *Maria saga* shows (for some reason) a fairly high percentage of unmarked S2 phrases (43 % vs. 12 % for the S1 phrases). In the case of S3 and S4 phrases, the referent is not given or implied in the preceding discourse, meaning that the addressee must infer the reference of the NP through either specific shared speaker-hearer knowledge or general world knowledge. Some of the texts show slightly

more definiteness marking in the S3 than in the S4 phrases, but it is clear that even the S3 category lags behind the anaphoric categories. Concerning the R and U phrases, the referent is readily accessible due to the head nouns' lexically inherent relationality and uniqueness features. As far as it is possible to judge, there are no remarkable differences between the R and U classes when it comes to definiteness marking. With these remarks in mind, I find the procedure justified.

Table 5 shows the absolute numbers and Table 6 the percentage distribution by NP type within each combination of classes. In the tables, the texts have been arranged so that the Icelandic texts come first, then the Norwegian, the Swedish and finally the Danish one. Within each language area, the texts are arranged chronologically. The Icelandic texts have also been divided into two subgroups: homilies and narratives. The Norwegian and Swedish text samples are all narrative in nature except the *Konungastyrelsen*. The latter is a moral and strategic instruction for a future king. The Danish *Lucidarius* is also a didactic text, but, contrary to *Konungastyrelsen*, in the form of a dialogue.

Table 5. The number of "noun with definite article" (Def.), "noun + demonstrative" (Dem.) and "bare noun" (Ø) in semantically definite NPs classified as S1 or S2, S3 or S4, R or U respectively in eleven Old Scandinavian texts.

			S1 + S2 (N)			S3 + S4 (N)		R + U (N)	
Text	Prov.	Dating	Def.	Dem.	Ø	Def.	Ø	Def.	Ø
IH: Homily g	Ice	*c.*1200	21	23	2	50	16	26	10
IH: Homily n2	Ice	*c.*1200	67	6	12	51	33	23	29
The miracles of Bishop Thorlak	Ice	*c.*1200	155	31	20	2	23	27	9
Heimskringla	Ice	*c.*1230	116	38	69	22	76	25	20
Alexander Saga	Ice	*c.*1250	39	38	11	39	14	24	29
The miracles of Saint Olav	No	*c.*1200	49	35	17	7	18	18	22
Barlaam and Josaphat Saga	No	*c.*1250	24	33	1	43	15	17	16
Old Sw. Leg.: Maria Saga	Sw	*c.*1300	135	21	32	11	25	28	35
MB 1: Intro	Sw	*c.*1330	75	20	13	33	10	32	6
Konungastyrelsen	Sw	*c.*1350	26	8	4	55	12	8	4
Lucidarius	Da	*c.*1350	68	17	3	90	34	110	21

Table 6. The percentage of "noun with definite article" (Def.), "noun + demonstrative" (Dem.) and "bare noun" (Ø) in semantically definite NPs classified as S1 or S2, S3 or S4, R or U respectively in eleven Old Scandinavian texts.

			S1 + S2 (%)			S3 + S4 (%)		R + U (%)	
Text	Prov.	Dating	Def.	Dem.	Ø	Def.	Ø	Def.	Ø
IH: Homily g	Ice	*c.*1200	46	50	4	76	24	72	28
IH: Homily n2	Ice	*c.*1200	79	7	14	61	39	44	66
The miracles of Bishop Thorlak	Ice	*c.*1200	75	15	10	**8**	92	**75**	25
Heimskringla	Ice	*c.*1230	52	17	31	**22**	78	**56**	44
Alexander Saga	Ice	*c.*1250	44	43	13	**74**	26	**45**	55
The miracles of Saint Olav	No	*c.*1200	49	35	17	28	72	45	55
Barlaam and Josaphat Saga	No	*c.*1250	41	57	2	**74**	26	**52**	48
Old Sw. Leg.: Maria Saga	Sw	*c.*1300	72	11	17	31	69	44	56
MB 1: Intro	Sw	*c.*1330	69	19	12	77	23	84	16
Konungastyrelsen[1]	Sw	*c.*1350	68	21	11	82	18	(67)	(33)
Lucidarius	Da	*c.*1350	77	19	3	73	27	84	16

[1] The figures for the R/U phrases in the *Konungstyrelsen* are based on only 12 instances in total, cf. Table 5.

As can be seen from the figures in the shaded columns in Table 6, which refer to bare noun NPs, the proportion of anaphoric NPs (S1 + S2) that completely lack definiteness marking is at most 17 % in all but one of the texts. For some reason, *Heimskringla* stands out with 31 %. In the other two groups, bare noun NPs are more common. In this context, we do not need to delve into the distribution between the definite article and the demonstrative in the S1/S2 phrases. Most demonstratives are not used in a remarkable way, although, as mentioned, they sometimes appear in Old Scandinavian texts where a modern reader would prefer the definite article. Thus, so far, the results confirm that definiteness marking gains ground first in anaphoric NPs and that this development has already gone quite far in the texts presented here.

Instead, let us compare the outcome for the R and U phrases, on the one hand, and the S3 and S4 phrases, on the other, as a test of the relevance of distinguishing between non-anaphoric NPs based on whether they pos-

sess relationality and/or uniqueness features or not. As Table 6 shows, the proportion of definiteness marked NPs sometimes varies greatly between the S3/S4 and R/U categories. According to a Chi-Square test based on the figures in Table 5, the difference is in some cases statistically significant, which is marked in bold in Table 6 for the numbers that indicate the proportion of phrases with the definite article. Although the total number of instances is limited and the classification is to some extent subjective, the significance in these cases is so high (p-values ≤ 0.029) that it cannot be ignored. This suggests that the definite article actually spreads to these groupings independently of each other, in the sense that a high share in one grouping does not automatically mean a high share in the other.

Interestingly, in some texts, the definite article is more frequent in the R/U phrases than in the S3/S4 phrases, but in other texts, it is the other way around. This also applies to the texts where the difference is statistically significant, among which two follow the former pattern and two the latter. Furthermore, when comparing the Icelandic narratives, we can see that the proportion of R/U phrases decreases from the oldest to the youngest text, which is remarkable. No similar decrease is seen in the Norwegian and Swedish texts. But also the early *Miracles of Saint Olav* and the *Maria saga* have more R/U phrases than S3/S4 phrases (even though the difference is not statistically significant in these texts), while the *Barlaam and Josaphat saga* has more S3/S4 phrases than R/U phrases.

The variable that seems most clearly consistent with the expectations of a diachronic development is the proportion of S3/S4 phrases, which, as long as we stick to the narratives, generally increases between the earlier and the later texts within each language group. If we compare the language groups, however, the development in Swedish appears to be significantly later than in West Norse. *MB 1: Intro* (about the history of the Jews) from *c.*1330 shows roughly the same proportion of S3/S4 phrases as the *Alexander Saga* and the *Barlaam and Josaphat Saga*, both from *c.*1250, and the proportion of the *Maria Saga* from *c.*1300 is roughly on par with that of *The Miracles of Saint Olav* from *c.*1200, but somewhat higher than in the earliest Icelandic narratives. The natural conclusion of this result is that the development in Swedish lags behind the development in West Norse. This lag probably applies to East Nordic in general, even if it cannot be proved, owing to the lack of early Danish texts.

It is more puzzling that the two Icelandic homilies show such a high percentage of S3/S4 phrases, regarding the *g*-homily almost the same as that of the considerably younger *Alexander Saga*. Spontaneously, the question arises whether we are seeing a genre difference here, but most of

the homilies in the homily books show little or modest use of the definite article, and I take the difference in definiteness marking in those texts as primarily a question of time of creation. Still, these two homilies cannot be younger than the manuscript itself, i.e. from *c.*1200. For the moment, I see no other explanation for the difference in definiteness marking between the homilies and the early Icelandic narratives than the circumstance that they were written during a period of rapid change with some individuals at the forefront of the development (and others lagging behind).

In Figure 4 below, the texts are sorted based on three assumptions. First, the spread of the definite article follows a similar developmental process all over Scandinavia, although not simultaneously across the entire area. Second, contemporary individuals, even within a limited geographical area, may exhibit such differences in their definiteness marking as to indicate that they may be at different stages of the development towards obligatorification of the definite article. Third, for the period covered by the texts in this investigation, the best measure of what stage the author of a certain text is at is this person's inclination to use the definite article in S3/S4 phrases.

Based on these assumptions, the texts in Figure 4 are not arranged by provenance, nor do they follow a strict chronological order. The order is instead determined in principle by the proportion of S3/S4 phrases with the definite article. (There is one minor exception though. Due to its high frequency of definiteness marking in R/U phrases, *Lucidarius* is placed to the right of *Barlaam*, which I find defensible, as the difference is negligible as regards the S3/S4 phrases.) For the sake of clarity, only the R/U and S3/S4 groupings are compared in the figure. (The proportion of definiteness marked NPs in S1/S2 phrases is, as already stated, very or relatively high for all the texts.)

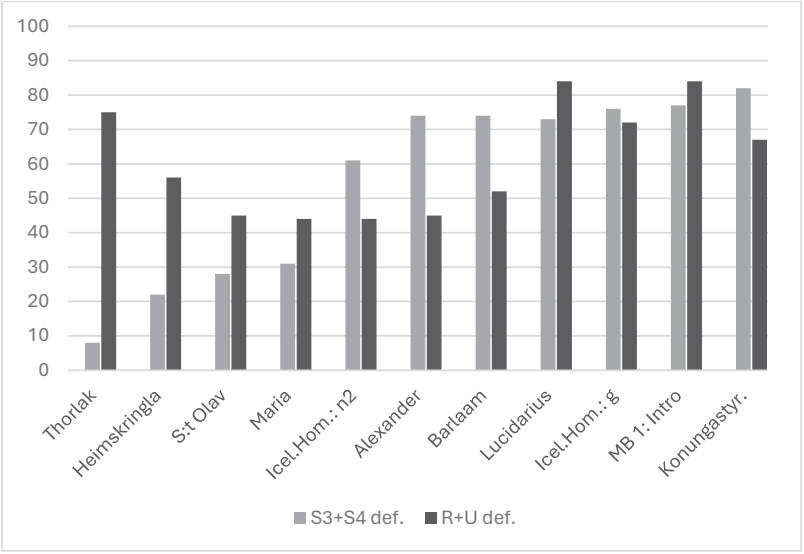

Figure 4. The percentage of NPs with the definite article in semantically definite NPs classified as S3 or S4 and R or U in eleven Old Scandinavian texts.

The visualization in Figure 4 provides a basis for reflection on how the definite article gains more ground when it is already fairly well established for anaphoric reference. If the premise that the proportion of S3/S4 phrases with the definite article in a text reflects its stage of development is on the right track, the texts in the figure can be divided into three groups. The four texts to the left constitute the first group. Here, we see a gradual increase of definiteness marking in the S3/S4 phrases from a very modest level to around 30 %. Somewhat surprisingly, however, the definite marking in the R/U phrases seems to decrease at the same rate as the increase in the S3/S4 phrases. It is legitimate to ask, of course, whether this really reflects the real course or whether there is something in the material that conveys a false impression.

It might be of some relevance that the R/U phrases in the *Thorlak* sample are all of the R type and in 30 cases of 36 refer to body parts, 23 of which appear with the definite article (but also four of the six remaining R phrases has the article). However, in the *Heimskringla* sample, the R/U phrases are distributed across NPs for body parts, other R phrases and U phrases and mark a majority of both the body-part phrases and

U phrases with the definite article, so the combined proportion of 56% definiteness marking seems realistic in this case. I take this as support for assuming that a spread across different kinds of R and U phrases in *Thorlak* would still have given a high percentage in this text sample as well. The *Miracles of Saint Olav* and the *Maria saga*, on the other hand, actually show a more even distribution of marked and unmarked phrases referring to body parts – and in R/U phrases in general. There is thus nothing to suggest that the outcome would not characterize the relevant texts in a fairly accurate way.

Does the outcome then depend on regional variation, so that the high presence of the definite article in R and U phrases in the former texts applies to early Icelandic, and the slightly lower proportion in the Norwegian and Swedish texts reflects the language on the mainland? However, the different kinds of R/U phrases in the Icelandic *Alexander saga* appear with roughly the same distribution of marked and unmarked NPs as the R/U phrases in the *Miracles of Saint Olav* and the *Maria saga*, so the answer is probably no. This leaves us with two possibilities. Either the definite article really loses some ground in the R/U phrases when it first starts to assert itself in the S3/S4 phrases, or the very high proportion of R/U phrases in the *Thorlak* and *Heimskringla* samples is just due to the preferences of the individual authors. Regardless of which alternative is correct, the conclusion should be that the definite article has begun to be used to a relatively large extent in R/U phrases before it takes firm hold in the S3/S4 phrases.

The second text group consists of text 5–7 in Figure 4. These texts show a substantially increased use of the definite article in S3/S4 phrases, while the proportion of definiteness marking in R/U phrases more or less remains the same as in the *Miracles of Saint Olav* and the *Maria saga*. Recall that the difference between the referential categories is statistically significant as regards the *Alexander saga* and the *Barlaam and Josaphat saga*. This outcome suggests that once the definite article appears more regularly in S3/S4 phrases, it quickly gains ground in this grouping, while the use of the article in R/U phrases seems less affected during this intermediate period.

The four texts on the right in Figure 4 constitute the third group, which is characterized by an extensive use of the definite article in NPs of all referential categories. It appears that this stage is reached after a second wave of increased use of the definite article in the R/U phrases.

5.5 Conclusions and comments

Although the empirical investigation described above is based on a limited number of texts, it leads to some clear conclusions regarding the spread of the definite article in the Scandinavian languages. First, the development is undoubtedly considerably earlier in West Norse than in East Norse. However, the extent to which there may be chronological differences within these respective language areas is difficult to establish.

The very sparse use of the definite article in the provincial laws from Norway and the substantially higher frequency in *Grágás* could indicate a later start in Norway, at least in the Gulathing and the Eidsivathing areas, than in Iceland. On the other hand, the difference may also be due to the possibility that these laws were created at different times, as the later Norwegian and Icelandic narratives point to a relatively simultaneous development.

As concerns East Norse, the few definite articles in the provincial *Jutland Law* and the *Flensburg Town Law* indicate that Jutland was peripheral in relation to the innovation centre of the mainland. This conclusion is also logical from the point of view that the vernacular of south-western Jutland developed its own pre-posed definite article (cf. above). This suggests that the suffixed article never conquered all of Jutland in the Middle Ages. However, the Danish *Lucidarius* from *c.*1350, possibly written in northern Jutland (Kjær 2009), shows an extensive use of the definite article, similar to that in the fairly contemporary Swedish texts *MB 1* and the *Konungastyrelsen*.

The difficulties of accessing the regional distribution of the definite article are probably due, at least in part, to the rapidity of the development once it takes off. Judging from *Grágás*, it seems that this take-off can be dated to sometime during the 12th century in Iceland, whereas the provincial law from Uppland points to sometime during the 13th century as regards that part of Sweden. The later texts suggests that a generally high use of the definite article was reached, say, somewhat later than the middle of the 13th century in West Norse, if we trust the Icelandic and Norwegian narratives, and around or a little before the middle of the 14th century in East Norse.

However, it is puzzling that the Icelandic homilies, supposedly written about 1200, exhibit a use of the definite article far beyond that in the contemporary *Miracles of Bishop Thorlak* (written by Páll Jónsson, born 1155) and also that in *Heimskringla* (by Snorri Sturlason, born *c.*1179). It is in one case on a par with the mid-century *Alexander saga* and in the

other case even more frequent. It is difficult to see any other explanation why individual variation of this size is possible than that we are dealing with a very rapid development. In that case, we must also be aware that the general dating of the process can only be approximate.

The periods I have just tried to specify, after all, refer to the intermediate rapid stage of a change that can be assumed to develop according the so-called S-curve model. Now, we have evidence of the definite article from as early as the 11th century in the two runic inscriptions from Uppland mentioned above. This means that we also have to count on a long and slow start to the process, around two hundred years in this part of Scandinavia. If we imagine an equally long prehistory to the use of the definite article in Grágás, it is possible that it emerged in Iceland fairly soon after the colonization around 900 – or even came with the first settlers.

When it comes to the texts in the present study that show a high proportion of the definite article in all types of semantically definite NPs, the overall frequency is between 80 and 90 percent. It is reasonable to imagine that the development slows down from here. Just as we have a long start-up, the final phase towards obligatorification is most likely also protracted.

Finally, it is also possible to draw some conclusions regarding the spread of the definite article across the different referential categories. As expected, the development seems to start in the anaphoric S1 and S2 phrases, and, when the definite article really begins to assert itself, these categories reach a high proportion of definiteness marking quite quickly. The S3 and S4 phrases lag behind, but show the same rapid development once the article is taken into use more than sporadically. The R and U phrases, on the other hand, seem to come in two waves. They appear very early with the definite article now and then, and all of the "later" texts investigated show a proportion of over 40 %. An increase in the use of the definite article thus starts earlier in the R and U phrases than in the S3 and S4 phrases. However, the increase of definiteness marking in R and U phrases seems to pause when the article starts to be used more regularly in the S3 and S4 phrases, and it does not pick up again until we find the article in about 70 % of the latter categories. The implications of this outcome of the study will be further discussed in the next section.

6. Implications

The empirical investigation accounted for in the previous section confirms that it is reason to keep apart the different groupings of non-anaphoric definite NPs. A couple of the investigated texts show a significantly higher use of the definite article in R and U phrases than in S3 and S4 phrases, while a couple show a significantly higher frequency in S3 and S4 phrases than in R and U phrases. This outcome has methodological as well as theoretical implications.

From a methodological point of view, there is reason to emphasize the relevance of focusing on individual authors' language usage. This approach has led to a completely different result than the one that would have been obtained if the analysis had been based on assembled subcorpora for different regions and periods of time. If, for instance, the samples from the *Miracles of Bishop Thorlak* and *Heimskringla* and the two homilies had been grouped together to represent Icelandic in the early 13th century, the difference between the narratives and the homilies, which is indeed remarkable, would have been completely invisible, as would the pattern that appears in Figure 4.

From a theoretical point of view, two conclusions can be drawn from the fact that the difference in article use between the groupings of non-anaphoric definite NPs is statistically significant in some of the texts. First, it shows that the groupings are mentally relevant to the individual. I assume this applies even when no significant difference can be measured on the individual level, but it is the significant cases that provide the proof. Second, it is reasonable to assume that the definite article spreads along two paths once it is also used in non-anaphoric NPs. This is illustrated in the model in Figure 5. The transition from deictic to (direct and indirect) anaphoric uses in the figure follows the traditional view of the development, albeit with the modification that what I call R phrases are not equated with indirect anaphora with sortal head nouns. From here, however, two separate paths arise. One applies to NPs whose reference is restricted due to the head noun's R or U feature. The other applies to S3 and S4 phrases, for which the identification of the referent is up to the receiver to infer from the discourse situation.

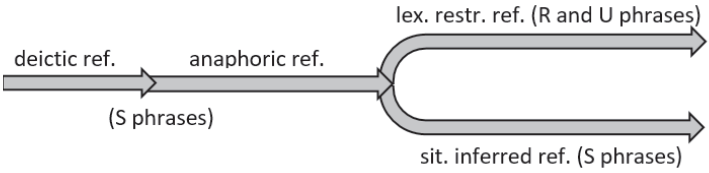

Figure 5. Model of the spread of definiteness marking from deictic NPs to direct and indirect anaphora based on sortal nouns and beyond along two separate paths, one for NPs whose reference is lexically restricted due to the inherent relationality or uniqueness of the head noun and one for non-anaphoric NPs based on sortal nouns for which the receiver must infer the intended referent.

With this model, it is completely logical that the development as concerns R and U phrases, on the one hand, and S3 and S4 phrases, on the other, does not have to proceed at the same pace, and that the distribution of the article across the referential groupings can vary during the course of development.

A virtue of the model is also that the extensions of the use of definite articles to generic NPs and to proper names can be distributed on the different paths. The use in generic NPs falls out naturally as an extension of the use in first mention NPs with sortal head nouns, while proper names are related to U phrases.

So far, the theoretical conclusions and the model in Figure 5 should be cross-linguistically applicable. However, the model as such does not predict if the R/U grouping should be ahead of the S3/S4 grouping or vice versa. One question that naturally arises from the Old Scandinavian evidence is then how we are to understand the relatively frequent early use of the definite article in R and U phrases in these languages, the later preference for the S3 and S4 phrases, and whether the course of development is a Scandinavian peculiarity or if it could apply more generally.

Spontaneously, one might think that a non-obligatory article starts to be used where it is not redundant before being introduced in the redundant cases (cf. Leiss 2007), i.e. that we should rather find it earlier in non-anaphoric NPs with sortal head nouns than in R and U phrases. On the other hand, the initial establishment in the anaphoric NPs means that the article first takes hold in phrases for which there is already a clear clue to the referent, namely an antecedent or an anchor. If, on this basis,

language users begin to see the article as a way of just making definiteness formally explicit, then the extension to R and U phrases is logical. Recall that Löbner (2011: 287 ff.) talks of redundant determination as "natural" or "congruent". Thus, one could say that if redundancy is one side of the coin, congruence is the other.

In the Nordic languages, after a while, the definite article begins to be used increasingly to mark definiteness in non-anaphoric NPs with sortal head nouns as well, and it appears that at a certain point in time there is a shift towards a higher proportion of articles in the S3 and S4 phrases than in the R and U phrases. Possibly, this is a sign that a new generation has drawn a different conclusion than the previous one, namely that the function of the article is primarily to coerce definiteness, not to add a formal marker of definiteness to phrases that are already semantically definite by virtue of a lexically inherent R or U feature. In the end, however, the article becomes mandatory in all semantically definite NPs.

I conclude with this proposal for how to understand the different steps of the development from definiteness marking in the Scandinavian languages. I leave it to future research to fill in the details and to explore whether it applies more generally.

Sources

Alexander Saga after Codex AM 519a, ed. by Andrea de Leeuw van Weenen (diplomatic and normalized), available at the Medieval Nordic Text Archive (Menota): https://clarino.uib.no/menota/page. Further information: Pettersson 2009.

Barlaam and Josaphat Saga after Codex Holm. perg. 6, ed. Rindal 1981, available at the Medieval Nordic Text Archive (Menota): https://clarino.uib.no/menota/page.

Eidsivathing Law after Codex AM 68 4°, ed. Halvorsen & Rindal 2008.

Flensburg Town Law after Stadtbuch 2, City Archive, Flensburg, available on the website of the Danish Language and Literature Society (Det Danske Sprog- og Litteraturselskab): https://tekstnet.dk/flensborg-stadsret/1.

Grágás Law after Codex Regius (Konungsbók), ed. Finsen 1852. Parts of the text are available on the website Wikisource: http://is.wikisource.org. Further information: Karlsson, Sveinsson & Árnason 1992:ix–xxxiii.

Gulathing Law after Codex Rantzovianus (E Don Var 137), ed. Keyser & Munch 1846, available on the website Heimskringla: https://heimskringla.no/wiki/. Further information: Eithun, Rindal & Ulset 1994.

Guta Law after Codex Holm. B 64, ed. Schlyter 1852, available at *Fornsvenska textbanken* (the Old Swedish text bank): https://project2.sol.lu.se/fornsvenska/. Further information: Peel 2015.

Guta Saga after Codex Holm B 64, available on the website Heimskringla: https://heimskringla.no/wiki/. Further information: Peel 2015.

Harpestrengs skrifter (The works of Henrik Harpestreng) after NKS 66, ed. by Marita Akhøj Nielsen (2006), available on the website of the Danish Language and Literature Society (Det Danske Sprog- og Litteraturselskab): https://tekstnet.dk/books/harpestreng-h_skrifter-nks66/.

Heimskringla ('The circle of the world') after Codex AM 35, ed. by Finnur Jónsson 1893–1900, available on the website Heimskringla: http://www.heimskringla.no/wiki/.

Icelandic Homily Book after Codex Holm. 15 (Isl. 15), ed. de Leeuw van Weenen 1993 (facsimile and diplomatic); Einarsson et al. 1993 (normalized), available on the website Heimskringla: http://www.heimskringla.no/wiki/.

Jutland Law after Codex Holm 37, facsimile available on the website of the Royal Danish Library: http://www5.kb.dk/permalink/2006/manus/41/dan/

Konungastyrelsen (a Swedish *speculum regale*) after print from 1634, ed. Moberg 1964 (facsimile), digitized at *Fornsvenska textbanken* (the Old Swedish text bank): https://project2.sol.lu.se/fornsvenska/. Further information: Moberg 1984.

Lucidarius after Codex AM 76, ed. Ejder & Skog 1993.

Medieval Bible Works 1 (*Medeltidens bibelarbeten 1*) after Codex Holm A 1, ed. Thorell 1959, available at *Fornsvenska textbanken* (the Old Swedish text bank): https://project2.sol.lu.se/fornsvenska/.

Miracles of Bishop Thorlak after AM 645, ed. Vigfússon 1858 (diplomatic); Egilsdóttir 2002 (normalized).

Norwegian Homily Book after AM 619, ed. Indrebø 1931, available at the Medieval Nordic Text Archive (Menota): https://clarino.uib.no/menota/page. Further information: Haugen & Ommundsen 2010.

Old Swedish Legendary (*Fornsvenska legendariet*) after Codex Holm A 34 (Codex Bureanus), ed. Wiktorsson 2020; Stephens 1847 available at *Fornsvenska textbanken* (the Old Swedish text bank): https://project2.sol.lu.se/fornsvenska/. Further information: Jansson 1936, Wiktorsson 2020.

Old Västergötland Law after Codex Holm. B 59, ed. Schlyter 1827, available at *Fornsvenska textbanken* (the Old Swedish text bank): https://project2.sol.lu.se/fornsvenska/.

Scania Law after Codex Holm. B 76, ed. Schlyter 1859, available at *Fornsvenska textbanken* (the Old Swedish text bank): https://project2.sol.lu.se/fornsvenska/.

Uppland Law after Codex Ups. B 12, ed. Schlyter 1834, available at *Fornsvenska textbanken* (the Old Swedish text bank): https://project2.sol.lu.se/fornsvenska/.

References

Abney, Steven P. 1987. *The English Noun Phrase and Its Sentential Aspect*. PhD dissertation, MIT.

Becker, Laura. 2021. *Articles in the World's languages*. Berlin, Boston: De Gruyter.

Bošković, Željko. 2012. On NPs and clauses. G. Grewendorf & T.E. Zimmermann *(eds.), Discourse and Grammar: From Sentence Types to Lexical Categories*. Boston: Walter de Gruyter. 179–242.

Carlson, Greg, Rachel Sussman, Natalie Klein & Michael Tanenhaus. 2006. Weak definite Noun Phrases. Christoffer Davis, Amy Rose Deal and Youri Zabbal (eds.), *Proceedings of NELS 36*. Amherst, MA: GLSA. 179–196.

Christophersen, Paul. 1939. *The articles. A study of their theory and use in English*. Copenhagen: Munksgaard.

Codex De la Gardie 4–7. Manuscript at the Uppsala University Library.

Czardybon, Adrian. 2017 [2021]. *Definiteness in a Language without Articles – A Study on Polish*. PhD diss. Düsseldorf: Düsseldorf University Press.

de Leeuw van Weenen, Andrea (ed.). 1993. *The Icelandic Homily Book. Perg. 15 4° in the Royal Library, Stockholm*. Reykjavík: Stofnun Árna Magnússonar á Íslandi.

Delsing, Lars-Olof. 2000. Konungastyrelsens ortografi och tillkomsthistoria. Lars-Erik Edlund (ed.), *Studier i svensk språkhistoria 5*. Umeå: Institutionen för litteraturvetenskap och nordiska språk vid Umeå universitet. 117–138.

De Mulder, Walter & Anne Carlier. 2011. The grammaticalization of definite articles. Bernd Heine & Heiko Narrog (eds.), *The Oxford Handbook of Grammaticalization*. Oxford: Oxford University Press. 522–535.

Diessel, Holger. 1999. *Demonstratives*. Amsterdam/Philadelphia: John Benjamins.

Ebert, Karen H. 1970. *Referenz, Sprechsituation und die bestimmten Artikel in einem nordfriesischen Dialekt* (*Fering*). (*Studien und Materialen 4*.) Kiel: Nordfriisk Institut.

Egilsdóttir, Ásdís. 2002. Jarteinabók Þorláks byskups in forna. *Íslenzk fornrit XVI. Biskupa sögur II*. Reykjavík: Hið íslenzka fornritfélag. 101–140.

Einarsson, Sigurbjörn, Guðrún Kvaran & Gunnlaugur Ingólfsson (eds.). 1993. *Íslensk Hómilíubók. Fornar Stólræður*. Reykjavík: Hið íslenska bókmenntafélag.

Eithun, Bjørn, Magnus Rindal & Tor Ulset (eds.). 1994*: Den eldre Gulatingslova*. (*Norrøne tekster 6*) Oslo: Riksarkivet.

Ejder, Bertil & Inge Skog (eds.). 1993. Lucidarius. Sigurd Kroon et al. (eds.), *A Danish teacher's manual of the mid fifteenth century: (Cod. AM 76, 8°) Vol 1 Transcription and facsimile*. Lund: Lund University Press. 126–365.

Ejskjær, Inger. 2005. The definite article. Oskar Bandle et al. (eds.), *The Nordic Languages. An International Handbook of the History of the North Germanic Languages. Vol. 2*. Berlin: Walter de Gruyter. 1733–1735.

Faarlund, Jan Terje. 2009. On the history of definiteness marking in Scandinavian. *Journal of Linguistics* 45. 617–639.

Finsen, Vilhjálmur (ed.) 1852. *Grágás. Islændernes lovbog i Fristatens tid. 1. Text 1*. Kjøbenhavn: Det nordiske Literatur-Samfund.

Finsen, Vilhjálmur. 1870. *Grágás. Islændernes lovbog i Fristatens tid. 3. Oversættelse 1*. Kjøbenhavn: Det nordiske Literatur-Samfund.
Fraurud, Kari. 1990. Definiteness and the processing of noun phrases in natural discourse. *Journal of semantics* 7. 395–433.
Goldstein, David. 2022. *Correlated grammaticalization. The rise of articles in Indo-European*. Diachronica Online-First articles. Published online: 1 June 2022. https://doi.org/10.1075/dia.20033.gol
Greenberg, Joseph H. 1978. How does a language acquire gender markers? Joseph H. Greenberg, Charles A. Ferguson & Edith Moravcsik (eds.), *Universals of Human Language. Volume 3: Word Structure*. Stanford, CA: Stanford University Press. 47–81.
Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.
Halvorsen, Eyvind Fjeld og Magnus Rindal (eds.). 2008. *De eldste østlandske kristenrettene. Borgartings og Eidsivatings eldre kristenretter*. (*Norrøne tekster* 7.) Oslo: Riksarkivet.
Haspelmath, Martin. 2022. Is cliticization an intermediate stage between free lexeme and affix status? Abstract from ICHL 25 (Oxford).
Haugen, Odd Einar & Åslaug Ommundsen. 2010. *Vår eldste bok. Skrift, miljø og biletbruk i den norske homilieboka*. Oslo: Novus.
Hawkins, John A. 1978. *Definiteness and indefiniteness: a study in reference and grammaticality prediction*. London: Croom Helm.
Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. PhD diss. University of Massachusetts.
Heine, Bernd. 2002. On the role of context in grammaticalization. Ilse Wischer & Gabriele Diewald (eds.), *New Reflections on Grammaticalization*. Amsterdam: John Benjamins. 83–101.
Himmelmann, Nikolaus P. 1997. *Deiktikon, Artikel, Nominalphrase: Zur Emergenz syntaktischer Struktur*. Tübingen: Niemeyer.
Himmelmann, Nikolaus P. 2001. Articles. Martin Haspelmath, Ekkerhard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language Typology and Language Universals*. Berlin/New York. Mouton de Gruyter. 831–841.
Indrebø, Gustav (ed.). 1931. *Gamal norsk homiliebok. Cod. Am. 619 4*o. Olso: Kjeldeskriftfondet.
Irmer, Matthias. 2011. *Bridging Inferences: Constraining and Resolving Underspecification in Discourse Interpretation*. Berlin/Boston: De Gruyter Mouton.
Jansson, Valter. 1934. *Fornsvenska legendariet. Handskrifter och språk*. (*Nordiska texter och undersökningar utgivna i Uppsala av Bengt Hesselman* 4.) Stockholm: Hugo Gebers förlag.
Karlsson, Gunnar, Kristján Sveinsson & Mörður Árnason (eds.). 1992: *Grágás. Lagasafn íslenska þjóðveldisins*. Reykjavík: Mál og menning.
Keyser, Rudoph & Peter Andreas Munch (ed.). 1846. *Norges gamle love indtil 1387*. 1 bind. Christiania: Norsk historisk kjeldeskrift-institutt. Den rettshistoriske kommisjon.
Kjær, Iver. 2009. Lucidarius. *Den Store Danske*. https://denstoredanske.lex.dk/Lucidarius.

Krámský, Jirí. 1972. *The article and the concept of definiteness in language*. The Hague: Mouton.

Krifka, Manfred, Francis Jeffry Pelletier, Gregory N. Carlson, Alice ter Meulen, Godehard Link & Gennaro Chierchia. 1995. Genericity: An Introduction. Gregory N. Carlson & Francis Jeffry Pelletier (eds.), *The Generic Book*. Chicago/London: The University of Chicago Press. 1–124.

Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog & Seongha Rhee. 2019. *World Lexicon of Grammaticalization* (2nd ed.). Cambridge: Cambridge University Press.

Lehmann, Christian. 2015 [1982]. *Thoughts on grammaticalization*. Berlin: Language Science Press.

Leiss, Elisabeth. 2000. *Artikel und Aspekt. Die grammatischen Muster von Definitheit*. Berlin/New York: Walter de Gruyter.

Leiss, Elisabeth. 2007. Covert patterns of definiteness/indefiniteness and aspectuality in Old Icelandic, Gothic, and Old High German. E. Stark et al. (eds.), *Nominal Determination. Typology, Context Constraints, and Historical Emergence*. Amsterdam: John Benjamins. 73–102.

Löbner, Sebastian. 1985. Definites. *Journal of Semantics* 4. 279–326.

Löbner, Sebastian. 2011. Concept Types and Determination. *Journal of Semantics* 28. 279–333.

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

Moberg, Lennart (ed). 1964. *En nyttigh bok om konnunga styrilse och höfdinga. Johannes Bureus utgåva 1634*. (*Samlingar utg. av Svenska Fornskriftsällskapet* 69:1.) Uppsala.

Moberg, Lennart. 1984. *Konungastyrelsen. En filologisk undersökning*. (*Samlingar utg. av Svenska Fornskriftsällskapet* 69:2.) Uppsala.

Moberg, Lennart. 2000: Nytt om Konungastyrelsen. *Språk & stil* 10 (n.f.). 243–247.

Møller, Kristen. 1974. Vestjyske Artikelproblemer. *Dialektstudier 3*. København: Institut for dansk dialektforskning. 74–102.

Ortmann, Albert. 2014. Definite Article Asymmetries and Concept Types: Semantic and Pragmatic Uniqueness. Gamerschlag, Thomas, Doris Gerland, Rainer Osswald & Viebke Petersen (eds.), *Frames and Concept Types: Application in Language and Philosophy*. (*Studies in Linguistics and Philosophy* 94.) 293–321.

Peel, Christine (ed.). 2015. *Guta Lag and Guta saga. The Law and History of the Gotlanders*. London: Routledge.

Perridon, Harry & Petra Sleeman. 2011. The noun phrase in Germanic and Romance. Common developments and differences. Petra Sleeman & Harry Perridon (eds.), *The Noun Phrase in Romance and Germanic*. Amsterdam: John Benjamins. 1–21.

Pettersson, Jonatan. 2009. *Fri översättning i det medeltida Västnorden*. (*Stockholm studies in Scandinavian philology* N.S. 51.) Acta Universitatis Stockholmiensis.

Pfaff, Alexander. 2014. Inside and outside – before and after: Weak and strong adjectives in Icelandic. Petra Sleeman, Frek van de Velde & Harry Perridon (eds.), *Adjectives in Germanic and Romance*. Amsterdam: John Benjamins. 217–244.

Pfaff, Alexander. 2015. *Adjectival and Genitival Modification in Definite Noun Phrases in Icelandic – A Tale of Outsiders and Inside Jobs*. Ph.D. dissertation. University of Tromsø.
Pfaff, Alexander. 2019. Reunited after 1000 years: The development of definite articles in Icelandic. *Nordic Journal of Linguistics* 42(2). 165–207.
Piotrowska, Alicija & Dominika Skrzypek. 2017. Inalienable possession in Swedish and Danish – a diachronic perspective. *Folia Scandinavica* 23. 25–45.
Piotrowska Alicija & Dominika Skrzypek. 2022. Alienability Splits in Swedish from a Diachronic Perspective. GEMA Online Journal of Language Studies Volume 22(1). 57–79. http://doi.org/10.17576/gema-2022-2201-04
Rindal, Magnus (ed.). 1981. *Barlaam ok Josaphats saga.* (*Norrøne tekster* 4.) Oslo: Norsk historisk kjeldeskrift-institutt.
Roberts, Craige. 2003. Uniqueness in definite noun phrases. *Linguistics and Philosophy* 26. 287–350.
Russell, Bertrand. 1905. On Denoting. *Mind* 14 (56). 479–493.
Scandinavian Runic-text Database 2020. http://www.nordiska.uu.se/forskn/samnord.htm
Schlachter, Eva. A complex grammaticalization scenario for the definite article. Renata Szczepaniak & Johanna Flick (eds.), *Walking on the Grammaticalization Path of the Definite Article*. Amsterdam: John Benjamins. 17–41.
Schlyter, Carl Johan (ed.) 1827. *Westgötalagen*. (*Samling af Sweriges gamla lagar* 1.) Stockholm: Haeggström.
Schlyter, Carl Johan (ed.). 1834. *Uplandslagen.* (*Samling af Sweriges gamla lagar* 3.) Stockholm: P. A. Norstedt & Söner.
Schlyter, Carl Johan (ed.) 1852. *Gotlandslagen*. (*Samling af Sweriges gamla lagar* 7.) Lund: Berlingska boktryckeriet.
Schlyter, Carl Johan (ed.) 1859. *Skånelagen*. (*Samling af Sweriges gamla lagar* 9.) Lund: Berlingska boktryckeriet.
Schwarz, Florian. 2009. *Two Types of Definites in Natural Language*. PhD diss. The Graduate School of University of Massachusetts Amherst.
Schwarz, Florian. 2013. Two Types of Definites Cross-linguistically. *Language and Linguistic Compass* 7. 534–559.
Schwarz, Florian. 2019. Weak vs. strong definite articles: Meaning and form across languages. Ana Aguilar-Guevara, Julia Pozas Loyo & Violeta Vázquez-Rojas Maldonado (eds.), *Definiteness across languages*. Berlin: Language Science Press. 1–37.
Schwarz, Monika. 2000. *Indirekte Anaphern in Texen. Studien zur domänengebundenen Referenz und Kohärenz im Deutschen*. Tübingen: Niemayer.
Schwarz-Friesel, Monika. 2007. Indirect anaphora in text. A cognitive account. Schwarz-Friesel, Monika, Manfred Consten & Mareile Knees (eds.), *Anaphors in Text. Cognitive, formal and applied approaches to anaphoric reference*. Amsterdam/Philadelphia: John Benjamins. 3–20.
Skrzypek, Dominika. 2012. *Grammaticalization of (in)definiteness in Swedish*. Poznan: Wydawnictwo Naukowe UAM.
Skrzypek, Dominika, Alicja Piotrowska & Rafal Jaworski. 2021. *The Diachrony of Definiteness in North Germanic*. Leiden/Boston: Brill.

Stephens, Georg (ed.). 1847. *Ett fornsvenskt legendarium.* (*Samlingar utg. af Svenska fornskriftsällskapet* 7:1–2.)
Stroh-Wollin, Ulla. 2011. A semantic approach to noun phrase structure and the definite–indefinite distinction in Germanic and Romance. Petra Sleeman & Harry Perridon (eds.), *The Noun Phrase in Romance and Germanic*. Amsterdam: John Benjamins. 127–140.
Stroh-Wollin, Ulla. 2015. Understanding the gradual development of definiteness marking the case of Swedish. *Working Papers in Scandinavian Syntax* 95. 11–32.
Stroh-Wollin, Ulla. 2016. The emergence of definiteness marking in Scandinavian – new answers to old questions. *Arkiv för nordisk filologi* 131. 129–169.
Stroh-Wollin, Ulla. 2020. *Hinn* and *hinn*: Early Icelandic as the clue to the history and etymology of two Old Scandinavian words. *Nordic Journal of Linguistics* 43. 205–228.
Stroh-Wollin, Ulla. 2023. Om tidsskikt i Upplandslagen och Jyske lov. Vad säger bruket av substantivets bestämda form? *Från Skånske lov till Vasabrev. Stockholmsstudier i östnordisk filologi.* (*Selskab for østnordisk filologi* 5.) 39–53.
Szczepaniak, Renata & Johanna Flick. 2020. Introduction. Renata Szczepaniak & Johanna Flick (eds.) *Walking on the Grammaticalization Path of the Definite Article*. Amsterdam: John Benjamins. 1–13.
Thorell, Olof. 1959. *Fem Moseböcker på fornsvenska: Pentateuchparafrasen efter MB I B.* (*Samlingar utg. av Svenska fornskriftsällskapet* 60.)
Vigfússon, Guðbrandur. 1858. Jarteinabók Þorláks biskups, sú er Páll lét lesa upp á alþingi 1199. *Biskupa sögur I*. Copenhagen: Hinu íslenzka bókmentafélagi. 333–356.
Wiktorsson, Per-Axel (ed.). 2020: *Fornsvenska legendariet I–IV*. Skara: Skara stifts historiska sällskap.

Appendix

The number (Table A) and percentages (Table B) of "noun with definite article" (Def.), "noun + demonstrative" (Dem.) and "bare noun" (Ø) in semantically definite NPs classified as S1, S2, S3, S4, R, U in samples from eleven Old Scandinavian texts.

Table A

	S1 (N)			S2 (N)			S3 (N)		S4 (N)		R (N)		U (N)	
Text sample	Def.	Dem.	Ø	Def.	Dem.	Ø	Def.	Ø	Def.	Ø	Def.	Ø	Def.	Ø
Isl. Hom.: g	15	11	0	6	12	2	16	2	34	14	25	3	1	7
Isl. Hom.: n2	54	6	9	13	0	3	24	12	27	21	5	5	18	24
Thorlak	130	15	11	25	16	9	2	23	0	0	27	9	0	0
Heimskringla	79	16	46	37	22	23	20	66	2	10	16	15	9	5
Alexander	31	30	9	8	8	2	29	8	10	8	21	28	3	1
S:t Olav	40	23	11	9	12	6	7	15	0	3	10	17	8	5
Barlaam and Josaphat	22	12	1	2	21	0	11	5	32	10	3	4	14	12
Maria	124	16	20	11	5	12	5	23	6	2	20	28	8	7
MB 1: Intro	70	8	13	5	12	0	16	5	17	5	17	2	15	4
Konungastyr.	24	7	4	2	1	0	14	4	41	8	6	4	2	0
Lucidarius	45	8	3	23	9	0	68	15	22	19	31	2	79	19

Table B. Percentages based on fewer than 20 instances are rendered in italics. No percentage is given when the total number of instances is less than 10.

	S1 (%)			S2 (%)			S3 (%)		S4 (%)		R (%)		U (%)	
Text sample	Def.	Dem.	Ø	Def.	Dem.	Ø	Def.	Ø	Def.	Ø	Def.	Ø	Def.	Ø
Isl. Hom.: g	58	42	0	30	60	10	89	11	71	29	89	11	*	*
Isl. Hom.: n2	78	9	13	81	0	19	67	33	56	44	50	50	43	57
Thorlak	83	10	7	50	32	18	8	92	*	*	75	25	*	*
Heimskringla	56	11	33	45	27	28	23	77	17	83	52	48	64	36
Alexander	44	43	13	44	44	11	78	22	56	44	43	57	*	*
S:t Olav	54	31	15	33	44	22	32	68	*	*	37	63	67	38
Barlaam and Josaphat	63	34	3	9	91	0	69	31	76	24	*	*	48	52
Maria	78	10	12	39	18	43	18	82	*	*	42	58	53	47
MB 1: Intro	77	9	14	29	71	0	76	24	77	23	89	11	79	21
Konungastyr.	69	20	11	*	*	*	78	22	84	16	*	*	*	*
Lucidarius	80	14	5	72	22	0	82	18	54	46	94	6	81	19