Familiar vs. unique in a diachronic perspective. Case study of the rise of the definite article in North Germanic

ALICJA PIOTROWSKA (iD)

DOMINIKA SKRZYPEK (iD)

Author affiliations can be found in the back matter of this article

Abstract

The aim of the present study is to follow the development of the suffixed definite article in North Germanic, in particular taking into account the unique reference expressed by the nascent article. The study is based on the corpora of Old Swedish, Old Danish and Old Icelandic texts written between 1200 and 1550. Both qualitative and quantitative methods, such as logistic regression models, are applied. The study is grounded in the notions of familiarity and uniqueness, which we explore diachronically. The results indicate that the use of the definite article is much more frequent in familiar than in unique contexts in North Germanic in the periods studied, as a greater proportion of NPs with direct anaphors is definite in the oldest extant texts, as well as throughout the later periods, than the proportion of NPs with unique referents. NPs with unique referents are further shown to constitute a non-uniform group, where the 'more local' unique NPs (grounded in specific knowledge) appear more frequently with a definite article than the 'more global' unique referents (grounded in encyclopaedic knowledge).

1 Introductory remarks

The discussion of the meaning of definiteness dates back to Frege's classic example *The Morning Star is the Evening Star*. It has been further fuelled by a debate between Russell (1905) and Strawson (1950) and has later mainly focused on why some discourse referents may be definite: either because they are familiar or because they are unique. In languages which have developed definite article, some uses of the article may be explained by evoking familiarity, while others by invoking uniqueness. A number of attempts have been made to either subsume all uses of the definite article under one or the other (e.g. Christophersen 1939 opts for familiarity), or to reconcile the notions by widening the scope of their meaning, as in weak familiarity (Roberts 2003). Enlightening data comes from studies on languages with more than one definite article and languages in which the definite article displays different patterns of behaviour depending on the type of context and the semantic notion invoked, e.g. the German definite article which may be contracted with a preposition or remain a full lexeme, e.g. *im* (= *in* + *dem*) vs. *in dem* 'in the'. In recent years, a number of authors have suggested that the two notions cannot be reconciled and that there are two types of definiteness instead, instantiated by two definite articles: weak and strong, with distribution roughly corresponding to the contexts relying on the semantic notions of uniqueness and familiarity. This observation is grounded in empirical studies, most notably Ebert (1971a), and has been further explored in theoretically oriented works such as Hawkins (1978) and Löbner (1985). The terms weak and strong were proposed by Schwarz (2009) in his work on definite articles in German. We will discuss this in more detail in section 2.

In the following we differentiate between: a) semantic concepts of familiarity and uniqueness, b) morphosyntactic representation, i.e. definite articles and c) contexts in which the articles may or may not be used, such as anaphora, unique reference and larger situation use.

The meaning of definiteness has so far mainly been unravelled in synchronic studies. The majority of these are based on constructed examples, which are studied in detail, with possible and probable contexts constructed around them (Hawkins 1978; Heim 1988). Fewer studies have been based on actual corpus data (famously Fraurud 1990 and her later publications; cf. Löbner 2003), revealing a number of counter-intuitive facts, e.g. that nominal phrases with a definite article (defNPs) used to introduce new discourse referents, i.e. first mention definites, are in fact more frequent than definites used anaphorically (Fraurud 1990). The majority of these first mention uses were established by bridging (Hawkins' *associative anaphora*), a use that is grounded in both familiarity (via an anchor) and uniqueness (the only possible discourse referent), as in 1.

(1)   John has bought a new house. **The roof** is green.

The different uses of NPs with a definite article seem to be ordered diachronically from those in which the definite article is used because the referent can be subsumed under some notion of familiarity (e.g. anaphora) to those in which the referent can be more easily subsumed under the notion of uniqueness (e.g. prototypically unique entities in a context – *king in a country*). In their model of definite article grammaticalization De Mulder and Carlier place direct anaphora (co-referring NPs) at the onset of grammaticalization of the definite article (De Mulder & Carlier 2011; cf. Himmelmann 1997; Lyons 1999: 158ff; see also Simonenko & Carlier 2016). As the final stage of definite article grammaticalization they discuss the use with unique referents and the bridging uses as the half-way stage.

In the following discussion we will take Hawkins' typology of the major usage types of the definite article as our starting point (Hawkins 1978: 106–129). This typology includes the following subtypes, illustrated by relevant examples quoted after Hawkins:

(2)   1.   anaphoric and immediate situation uses, e.g. *Fred was wearing trousers. The pants had a big patch on them.* (Hawkins 1978: 107)[1]
         2.   larger situation uses (relying on specific or general knowledge), e.g. *The Prime Minister has just resigned.* (Hawkins 1978: 116)
         3.   associative anaphoric uses, e.g. *The man drove past our house in a car. The exhaust fumes were terrible.* (Hawkins 1978: 123)

---

[1]   Note that Hawkins (1978) groups anaphoric and immediate situation uses together; in this paper, however, these terms are not treated as interchangeable.

For the present study we are mainly concerned with the second type, i.e. larger situation uses. These are different from immediate situation uses, in which the unique referent is present, though not necessarily visible, in the situation in which the utterance is made (a requirement that the referent be visible is important for demonstratives, not for articles). Thus in larger situation use people in the same village can talk about *the church, the pub, the village green*; members of the same nation can talk about *the Queen, the navy, the Prime Minister*; all people can talk about *the sun, the moon, the planets*. These larger situations can be of varying size, but they will all have as their focal, defining point the immediate situation of utterance in which the speech act is taking place (Hawkins 1978: 115). In this sense the definite article retains its link with the original demonstrative. Depending on the size of the situation, the definite reference may be made based on specific knowledge (we are in a village we are familiar with, we know there is a church in it, we can talk about the church) or on general knowledge (we do not know the church, but it is part of our knowledge that there usually is one in a village). In all these uses the referent in question has to be unique in the situation independent of its size.

The aim of the present study is to consider uses of the incipient definite article in North Germanic, based on a corpus of Danish, Swedish and Icelandic texts written between 1200 and 1550. In the study we focus on NPs with a suffixed definite article used to refer to unique discourse referents. We argue that unique discourse referents do not form a homogenous group as the semantic concepts invoked for their resolution involve also familiarity. In our typology of definite article uses we follow Hawkins (1978) (see example 2). We regard an NP as familiar if it is co-indexed with another NP that precedes it in the text. When specified, we also use the notion of familiarity in a broader, pragmatic sense, i.e. an NP is familiar if its referent is associated with the context of the situation, but not necessarily found in the text (Jespersen's 1943 *situational basis*, cf. Lyons 1999: 254). Consider e.g. a defNP 'the king' used in a text discussing a specific country. The definite article is used because the referent is unique, however, it may be considered unique only with respect to this particular country. In this sense it is made familiar by the familiarity of the country under discussion or in which the discussion takes place and its uniqueness is 'local' rather than 'absolute'. We assume that definiteness uses can be put on a scale that goes from more familiarity (e.g. anaphora) to less familiarity (e.g. global uniqueness), with other categories (such as bridging, local uniqueness, immediate situations) falling in-between.

The main issue we wish to address is whether there is a difference in expression between the types of larger situation use that rely on specific knowledge (more 'local' uniqueness as it were) and the types that rely on general knowledge (more 'global' uniqueness) in a diachronic context. We assume that such differences exist and our hypothesis is that the more local types of reference predate the more global ones in terms of definite article presence, which can be measured by the frequency of the incipient definite article in these contexts. A significant rise in frequency of a grammaticalizing item is often considered to be an important indicator of the ongoing grammaticalization process (Bybee et al. 1994), although not a prerequisite. It can be hypothesised that certain uses of the incipient definite article will display higher frequencies earlier than others. Our hypothesis is that the uses in which the semantic concept of familiarity may be involved will be those with higher frequencies than the uses involving semantic uniqueness only.

The paper is organized as follows: we begin with an introduction of the theoretical tenets of the paper in 2; in 3 we give a short presentation of definiteness marking in modern North Germanic. In 4 the corpus and annotation method applied in the present study are presented. In 5 we present and discuss the results of the corpus study; section 6 concludes the paper.

2 Familiar vs. unique in a diachronic perspective

The observation that not all uses of the definite article can be subsumed under one category is of particular interest to a language historian. The diversity of definite article uses naturally leads to an assumption that not all uses have emerged at the same time, but rather that some predate others in the grammaticalization of the definite article.

The traditional model of definite article grammaticalization recognizes the original deictic meaning (most definite articles are derived from demonstratives or other deictic elements, see Lyons 1999) and treats the first uses of the demonstrative to point within texts rather than situations (i.e. direct anaphora) as an extension of this original meaning. It is possible to use demonstratives as

anaphoric markers in any language without necessarily stipulating that they are on the verge of grammaticalization into definite articles. The interpretation of the definite form in such contexts is based on the semantic concept of familiarity (cf. Hawkins' anaphoric and immediate situation use).

However, in the next stage, the original demonstrative comes to be used with new discourse referents which are in some way grounded in the preceding discourse (Hawkins' associative anaphoric use). Despite several attempts to prove the contrary, demonstratives can only appear in these contexts under restricted circumstances (e.g. in Japanese, see Takamine 2014: 39; see also Diessel 1999: 24) and the fact that a form is used in the associative anaphoric context may be taken to be a symptom of its ongoing grammaticalization into a definite article. So far, the transition from direct to indirect (or associative) anaphora escapes us. The indirect anaphora itself is a heterogeneous and complex context, which seems to be partly based on familiarity, but also reliant on uniqueness. Some attempts to describe the diachronic path through indirect anaphora have been made (Skrzypek 2020), but many questions remain unanswered.

The final stage of definite article grammaticalization is the possibility (later on an obligation) to use the original demonstrative with referents which are unique within the utterance-situation, or which are inherently unique (Hawkins' larger situation use). This use is clearly at odds with the original meaning of the form, as the demonstrative is used to pick out a discourse referent from among other potential referents, while the definite article used in a larger situation use serves to ensure the hearer of the referent's uniqueness and confirms that it cannot be confused with another referent. This use is based solely on uniqueness.

The strong-weak dichotomy proposed by Schwarz (2009) classifies the definite article used anaphorically as strong, and the definite article in the larger situation use as weak. The definite articles in associative anaphoric uses are classified by Schwarz (2009) partly as strong and partly as weak. In languages with two distinct definite articles, such as North Fering (Ebert 1971b) and some German dialects (Hartmann 1967; 1978), we may assume two grammaticalizations: of two definite articles (the concept of two definite articles is also explored in Breu 2004 and Wespel 2008), see also Dahl's study on the grammaticalization of the definite articles in Swedish (Dahl 2015). In languages with one definite article which unites both familiar and unique uses, we propose instead one grammaticalization, whose sub-stages include the rise of strong and weak uses of the grammaticalizing article. Considering the etymology of the definite article, i.e. deictic element, typically a demonstrative pronoun, it is natural to assume that the grammaticalization proceeds from uses based on the semantic concept of familiarity to those based on uniqueness, which is a claim supported in this paper (but also among others in De Mulder & Carlier 2011). We will address this hypothesis with a statistical analysis of the North Germanic data in section 5.

We want further to address the issue of the bridge between the textually grounded uses of the developing definite article and the larger situation uses that seem to be free of such grounding. How exactly does the development proceed from one to the other type? Since there exist different types of larger situation uses (based on how large the situation under consideration is), which are the earliest to adopt the definite article and which lag behind? These questions will be addressed in section 5, where we present an analysis of the North Germanic data.

3 Definiteness marking in present-day North Germanic languages

An overview of the NPs in North Germanic is given in *Table 1* along with the respective form of definite and indefinite articles.

In present-day North Germanic languages there are two definite articles: postposed and preposed. The postposed definite article is a suffix, always attached enclitically to the noun (in Insular Scandinavian languages Icelandic and Faroese the article attaches to the case-inflected form of the noun), as illustrated in examples 3–5 below. The origins of the postposed definite article are to be sought in the distal demonstrative *hinn* 'yon' (Perridon 1989; Skrzypek 2012). Apart from the postposed article, in Mainland Scandinavian languages (Danish, Norwegian, Swedish) and Faroese there is a preposed definite article *den* 'this', which generally occurs in NPs where the noun is accompanied by an attribute. In the present study we are only concerned with the suffixed definite article.

Context	Type of article	Form	Languages
Single noun NP '*the house*'	Definite article postposed	hus-**et** house-DEF	Danish, Norwegian, Swedish
		hús-**ið** house-DEF	Faroese, Icelandic
NP with an adjective and noun '*the big house*'	Definite article preposed	**det** stor-**e** hus DEF big-WK house	Danish
		det stor-**e** hus-et DEF big-WK house-DEF	Norwegian
		det stor**a** hus-et DEF big-WK house-DEF	Swedish
		hið/tað stór-**a** hús-ið DEF big-WK house-DEF	Faroese
	–	stór-a hús-ið big-WK house-DEF	Icelandic
Single noun NP '*a house*'	Indefinite article	**et** hus INDF house	Danish, Norwegian
		ett hus INDF house	Swedish
		eitt hús INDF house	Faroese
	–	hús house	Icelandic

Table 1 An overview of the NPs in North Germanic.

(3)   *Swedish*
Jag   ha-r   köp-t   hus-et.
1SG   have-PRS   buy-PST   house-DEF
'I have bought the house.'

(4)   *Danish*
Jeg   ha-r   køb-t   hus-et.
1SG   have-PRS   buy-PST   house-DEF
'I have bought the house.'

(5)   *Icelandic*
Ég   hef   keyp-t   hús-ið.
1SG   have.1SG.PRS   buy-PST   house-DEF
'I have bought the house.'

The postposed definite article in North Germanic languages may be used deictically, to refer to objects in the speaker's or hearer's direct vicinity, as in 6. Another function of the definite article is the direct anaphoric reference, which can be seen as the textual form of deictic reference, as it refers back to referents already accessible in discourse, as in example 7.

(6)   *Danish*
Kan   du   give   mig   **blad-et**?
can   2SG   give   1SG.REFL   newspaper-DEF
'Can you give me the newspaper?'

(7)   *Swedish*
Han   ha-de   en   hund. **Hund-en** het-te   Bella.
3SG.M   have-PST   INDF   dog   dog-DEF   call-PST   Bella
'He had a dog. The dog was called Bella.'

The definite article is further broadly used with the indirect (or associative) anaphora, in which the referent is anchored in another discourse referent, as in 8.

(8)   *Swedish*

Jag   läs-er   en   ny   bok.   **Författar-en**   skriv-er   riktigt bra.
1SG   read-PRS   INDF   new   book   author-DEF   write-PRS   really   good
'I am reading a new book. The author writes really well.'

The last two contexts in which the definite article is used in North Germanic is the unique reference, as in *the sun, the king, the Prime Minister,* the uniqueness of the referents being established relative to space/time coordinates, and the generic reference, as in example 9. However, the definite article is not the only option for expressing generic referents, as they may be marked by the indefinite article, or they may appear in bare NPs with no articles (both in singular and plural forms).

(9)   *Swedish* (Teleman et al. 2010: 150)

Piano-t   är   ett   stränginstrument.
piano-DEF   be.PRS   INDF   string.instrument
'The piano is a string instrument.'

The case of bare nouns in North Germanic is worth exploring, as they are quite frequently used despite the existence of definite and indefinite articles (except for Icelandic, which did not develop an indefinite article). Bare nouns are used, for example, as predicates in phrases stating names of professions, where English requires an indefinite article, as in 10.

(10)   a.   *Swedish*

Lena   är   **läkare**.
Lena   be.PRS   doctor
'Lena is a doctor.'

b.   *Danish*

Jan   er   **præst**.
Jan   be.PRS   priest
'Jan is a priest.'

The use of bare nouns does not always correspond to English indefinite articles and does not necessarily involve existential interpretation per se. Consider examples 11–12. The reference of NPs in these examples is based neither on familiarity or uniqueness, but rather it is incorporated into verbs ('to house-buy', 'to elk-hunt') or nouns ('elk-tracks', 'dog-tracks'). These constructions have been previously analysed as pseudo-incorporations (cf. Asudeh & Mikkelsen 2000), and are considered to be neutral with respect to definiteness and number, i.e. they are formally singular nouns, but they do not necessarily refer to only one entity (cf. Pettersson 1976).

(11)   *Swedish*

att   köpa   **hus**   / *ett   hus   / *hus-et
to   buy   house   / INDF   house   / house-DEF
'To be house-buying' (to be looking for a house with prospect of buying it)

(12)   *Swedish*

att   jaga   **älg**   / *en   älg   /*älg-en
to   hunt   elk   / INDF   elk   / elk-DEF
'To be elk-hunting'

4 The corpus and the method

The study is based on the corpora of Old Swedish, Old Danish and Old Icelandic texts written between 1200 and 1550. We omitted the Norwegian texts since due to Norway's political situation the extant texts from 1350–1550 represent one genre only, i.e. the so-called diplomas. As these are short legal documents with formulaic openings and closings, it is difficult to extract fragments of high narrativity, which has been the goal of the project. We have compiled the corpus ourselves from the available online sources and corpora, such as the *Fornsvenska textbanken*[2] for Swedish texts, *Middelalder og renæssance*[3] for Danish texts and *The Icelandic*

---

2   *Fornsvenska textbanken,* *https://project2.sol.lu.se/fornsvenska/*.

3   *Middelalder og renæssance,* *https://dsl.dk/website?id = 32*.

Parsed Historical Corpus[4] for Icelandic texts. The period 1200–1550 is the time of many rapid changes in the structure of nominal phrases in North Germanic languages, so in order to make the corpus more comparable across the three languages we have divided it into three periods, i.e. Period I (1200–1350), Period II (1350–1450), and Period III (1450–1550). The texts chosen for the corpus represent four genres, namely legal prose, religious prose, profane prose and sagas. Fragments of high narrativity are chosen to ensure examples of multiple references to salient discourse referents in different contexts. The texts of different genres are not entirely comparable across all the periods studied due to varied availability of source texts from different epochs. In the first period the majority of the texts are examples of legal prose. These are the oldest extant texts written in Nordic languages in Latin script and they are original Scandinavian texts, not translations. For these reasons they must be included in the study, even though they may be of very specific genre. The corpus for Periods II and III includes profane and religious prose, the majority of which are texts translated or adapted from foreign models, with the exception of Icelandic sagas which are original texts.

The overview of the corpus used in the study is presented in *Table 2*. The corpus consists of ca. 280 000 words. The length of the corpus for each language varies heavily, with Old Icelandic being represented by the longest texts and Old Danish by the shortest texts. This disparity is of little importance, however, as the number of NPs extracted from each part of the corpus is comparable across the three languages. We were striving to obtain ca. 3 000 NPs in each language. The total number of noun phrases annotated in the corpus amounts to 9 363.

Language	Number of words	Extracted NPs	NPs with unique reference within all NPs		NPs with direct anaphoric reference within all NPs	
Danish	33 122	3 022	317	10.5%	686	22.7%
Swedish	87 161	3 120	329	10.5%	692	22.2%
Icelandic	159 741	3 221	31	1.0%	618	19.2%
Raw total	280 024	9 363	677	7.2%	1 996	21.3%

Table 2 The overall number of NPs, NPs with unique referents and NPs with direct anaphoric referents annotated in the corpus.

The annotation included tagging the relevant NPs as U (unique). This was done under the following conditions: the NP in question had no antecedent and no anchor, it was a first mention use with no introduction in the earlier (or later) text, nor was the reading generic.

Among the annotated noun phrases there are 677 instances of unique referents, which constitute 7.2% of all of the NPs. Most of NPs with unique referents are found in Old Danish and Old Swedish where they constitute 10.5% of all NPs, while the Old Icelandic texts show a very low frequency of NPs with such referents. The number and proportion of the NPs with direct anaphoric reference is also given in *Table 2*, as the comparison between the two types of reference will be explored in section 5. Anaphoric NPs are evidently more frequent in the corpus. The languages display a similar average proportion of ca. 22% of NPs with direct anaphora within all of the NPs.

In gathering the data of NPs with unique referents it became clear that one item is overwhelmingly frequent in each language, namely the referent *gud* 'God', which continuously appears in its bare form throughout all the periods studied. If included, this item would constitute 25.2% (N = 228) of all of the instances of NPs with unique reference (N = 905) visibly skewing the results in favour of BNs, especially in Period II which includes many texts of religious nature. Given all that and the fact that *gud* 'God' is one of the few lexemes that remain bare in present-day North Germanic, we exclude all instances of this item from the present study, which leaves us with 677 NPs with unique referents (see *Table 2*).

Regarding the methods of annotating the corpus, special computer software is used to facilitate the manual annotation process. The program was tailor-made for the project at hand; it is called DiaDef (i.e. Diachrony of Definiteness). It allows one to annotate the noun phrases in a given text with all the necessary previously defined information. We manually annotated

---

4   *The Icelandic Parsed Historical Corpus, https://linguist.is/icelandic_treebank/Icelandic_Parsed_Historical_Corpus_ (IcePaHC).*

NPs on different levels of linguistic information, such as the type of article, type of reference, grammatical information (case, gender and number), syntactic roles and semantic information (e.g. animacy). The tool provides several features that improve the efficiency of use. For most annotation levels the system displays a context-sensitive list of prompts of available annotation tags, as well as tag suggestions based on previously annotated words. Each annotation decision is saved automatically in a periodically backed-up database. The tool also generates simple statistics that help us analyse the nature of noun phrases in Old North Germanic texts. The texts were annotated primarily from February 2017 to February 2018. In the analysis several statistical tests are used, such as contingency tables with a Chi-square Test of Independence and Binary and Multinomial Logistic Regression models. All tests were conducted with the help of the IBM SPSS Statistics programme.

5 Results of the corpus study

5.1 General results: raw frequencies

Tables 3 and *4* illustrate the results of the corpus study. We compare the frequencies of bare nouns (BN) and the grammaticalizing suffixed definite article (DEF) in two types of uses: in NPs with unique reference, i.e. based on uniqueness (*Table 3*), and in NPs with direct anaphoric reference, i.e. based on familiarity (*Table 4*). At this point, we are not yet differentiating between different types of larger situation use. Our main concern is to examine whether the postulated familiar-unique division in the frequency of use of the definite form can be found in the diachronic study and whether the uses of the definite article are associated with period in the given contexts.

Language	Period	DEF	BN	OTHER	Raw total
Danish	1200–1350	5.8% (10)	88.4% (153)	5.8% (10)	100.0% (173)
	1350–1450	11.0% (9)	69.5% (57)	19.5% (16)	100.0% (82)
	1450–1550	32.3% (20)	48.4% (30)	19.4% (12)	100.0% (62)
	Raw total	12.3% (39)	75.7% (240)	12.0% (38)	100.0% (317)
Swedish	1200–1350	13.2% (28)	83.0% (176)	3.8% (8)	100.0% (212)
	1350–1450	18.3% (19)	56.7% (59)	25.0% (26)	100.0% (104)
	1450–1550	46.2% (6)	23.1% (3)	30.8% (4)	100.0% (13)
	Raw total	16.1% (53)	72.3% (238)	11.6% (38)	100.0% (329)
Icelandic	1200–1350	0.0% (0)	46.2% (6)	53.8% (7)	100.0% (13)
	1350–1450	28.6% (4)	35.7% (5)	35.7% (5)	100.0% (14)
	1450–1550	50.0% (2)	50.0% (2)	0.0% (0)	100.0% (4)
	Raw total	19.4% (6)	41.9% (13)	38.7% (12)	100.0% (31)

Table 3 Bare nouns and the suffixed definite article in NPs with unique reference.

p < 0.001 for Danish,
p < 0.001 for Swedish.

As far as our annotation guidelines are concerned, the NPs annotated as unique reference have no co-referring NP in the preceding text, i.e. no antecedent (see 4). The NPs annotated as direct anaphora have a co-referring NP present in the preceding text. Apart from NPs with a definite article and bare NPs, there are other constructions that may appear in both uses (defined in the tables as OTHER), such as NPs with demonstratives, or NPs with only modifiers (such as

Language	Period	DEF	BN	OTHER	Raw total
Danish	1200–1350	12.9% (31)	40.7% (98)	46.5% (112)	100.0% (241)
	1350–1450	34.7% (87)	17.5% (44)	47.8% (120)	100.0% (251)
	1450–1550	38.1% (74)	8.8% (17)	53.1% (103)	100.0% (194)
	Raw total	28.0% (192)	23.2% (159)	48.8% (335)	100.0% (686)
Swedish	1200–1350	19.8% (52)	33.5% (88)	46.8% (123)	100.0% (263)
	1350–1450	71.5% (218)	0.0% (0)	28.5% (87)	100.0% (305)
	1450–1550	46.0% (57)	0.8% (1)	53.2% (66)	100.0% (124)
	Raw total	47.3% (327)	12.9% (89)	39.9% (276)	100.0% (692)
Icelandic	1200–1350	27.4% (86)	25.8% (81)	46.8% (147)	100.0% (314)
	1350–1450	23.2% (41)	28.8% (51)	48.0% (85)	100.0% (177)
	1450–1550	37.0% (47)	36.2% (46)	26.8% (34)	100.0% (127)
	Raw total	28.2% (174)	28.8% (178)	43.0% (266)	100.0% (618)

Table 4 Bare nouns and the suffixed definite article in NPs with direct anaphoric reference.

p < 0.001 for Danish,
p < 0.001 for Swedish,
p < 0.001 for Icelandic.

genitives, adjectives, etc.) but no explicit definite article; as we will discuss in section 5.2 the contribution of the OTHER category is not significant for NPs with unique referents.

A Chi-Square Test of Independence was performed to test the association between two variables in the tables, namely period (an ordinal variable) and type of article (a nominal variable). The null hypothesis for this test is the following:

(i) Period is not associated with the type of article in Swedish, Danish, and Icelandic.

Since the p-values reported for the tables are small, we decide to reject the null hypothesis in (i) for Swedish and Danish since the probability of Type I error is very small. There is thus a statistically significant association between period and article type in NPs with unique referents for these two languages. The test could not be performed for Icelandic because of insufficient data. There is simply not enough evidence in the Icelandic data to suggest an association between period and presence of a definite article (or lack of it) for NPs with unique referents.

Figures 1 and *2* illustrate the distribution of the types of NPs among NPs with unique reference over time.[5] In the case of NPs with unique referents (see *Table 3* and *Figures 1* and *2*), both Danish and Swedish exhibit similar developments. BNs are the most frequent in both languages and nearly all periods (with the exception of Period III in Swedish).[6] We observe that the highest proportion of BNs in Danish and Swedish occurs in Period I and decreases continuously through Periods II and III. Conversely, NPs with a definite article have a relatively low frequency with unique referents in Period I, but rise somewhat throughout Periods II and III, especially in Danish. On average only 12.3% and 16.1% of NPs with unique referents display the suffixed definite article in Danish and Swedish respectively. In the Icelandic corpus there were extremely few instances of NPs with unique referents compared to the Danish and Swedish corpora. No patterns of change or conclusions can be drawn based on the Icelandic data; we thus exclude

---

5   PERIOD is treated here as a continuous variable with 50 year intervals.

6   In the texts from Period III in Swedish there are, unfortunately, very few NPs with unique referents, so the proportions in this period are not statistically significant.
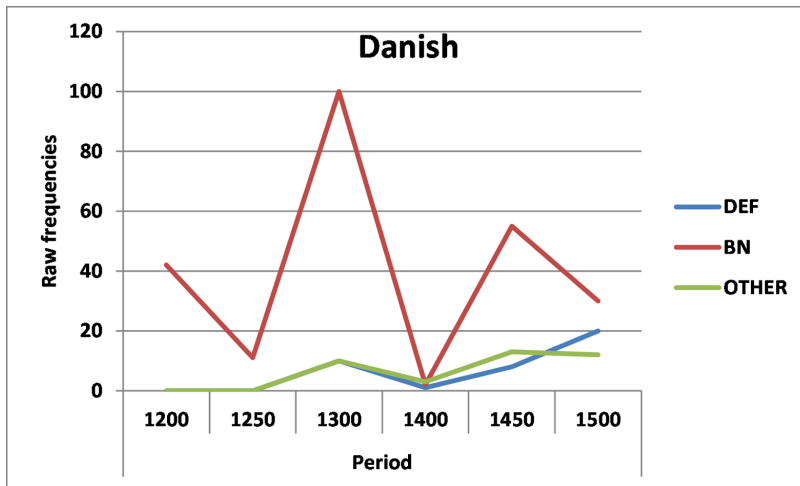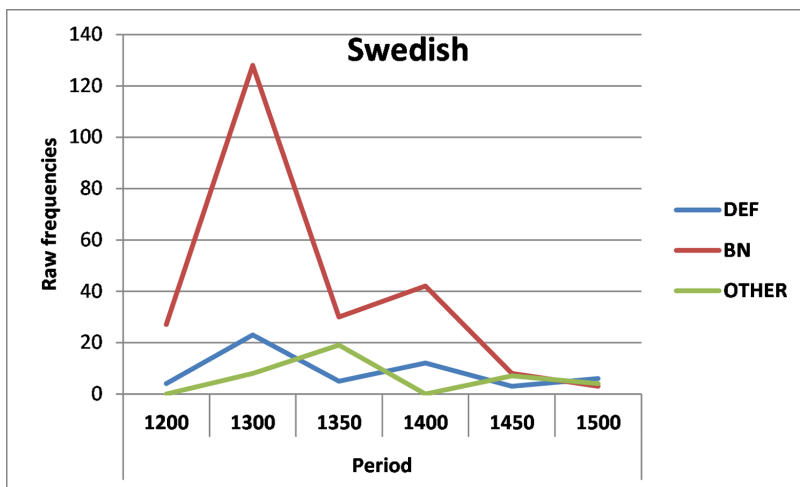
Icelandic NPs with unique referents from the present analysis. The lack of unique reference NPs is most likely due to the composition of the corpus; the Icelandic texts include only one legal text and a few religious sagas, the majority of the corpus, however, is composed of chivalric sagas. The Danish and Swedish corpora comprise many more legal and religious texts in which NPs with unique referents are found more often.[7]

As a comparison we also present the data concerning NPs with direct anaphoric reference, a use that is often given as an example of familiarity as the underlying meaning of definiteness (see *Table 4* and *Figures 3–5*). In this use we observe that the frequency of NPs with the suffixed definite article fluctuates, but it is overall gaining in frequency between 1200 and 1400 in Swedish and between 1200 and 1450 in Danish, while the use of bare nouns in direct anaphoric contexts decreases. On average, the incipient definite article is used in 34.5% of all anaphoric NPs, while bare nouns are used in 21.6% of anaphoric NPs. Compared to the average of 74.0% of NPs with unique referents that are expressed through bare nouns (here we are excluding the results for Icelandic), it is clear that BNs are quite strongly disfavoured in the context of direct anaphora in Swedish and Danish as early as in Period II. The pattern of bare NPs with anaphoric referents in Icelandic is quite the opposite, as they still constitute a large proportion of NPs relative to definite-marked NPs. The frequency results here are misleading. A closer inspection of Icelandic NPs with anaphoric referents in Period III reveals that among all of the bare NPs used anaphorically there are in fact only two lexical items: *kóngur* 'king' and *drottning* 'queen', as in 13, in which the referent is introduced in the first line as *Ríkharður kóngur* 'king Richard' and later on referred back to in a bare NP *kóngur* 'king'. Since there is a clear co-referring antecedent such examples are annotated as NPs with direct anaphora, even though such cases are ambiguous and may waver between anaphoric and unique reading.

---

7   Relative to contexts with direct anaphora, NPs with unique referents are 4.8 times more likely to appear in legal texts than in profane prose, and 2.2 times more likely to appear in religious texts than in profane prose in the corpora studied.
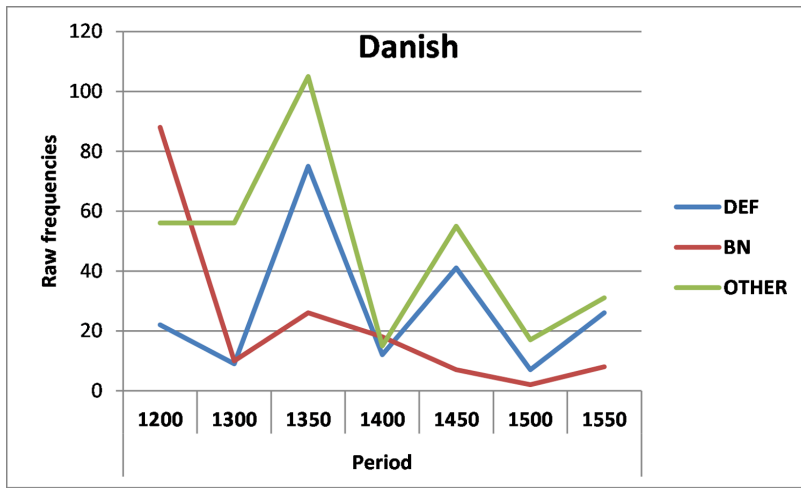
Danish

Figure 3 The raw frequencies of NP types (DEF, BN and OTHER) across time among NPs with direct anaphoric reference in Danish.

Swedish

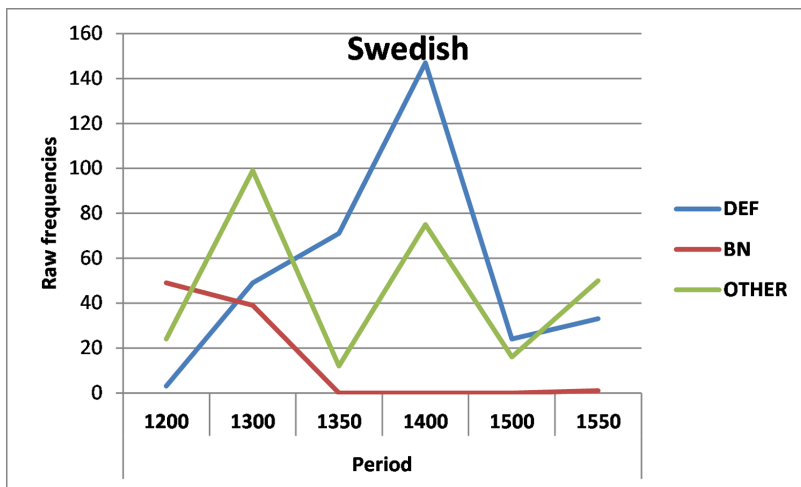**Figure 4** The raw frequencies of NP types (DEF, BN and OTHER) across time among NPs with direct anaphoric reference in Swedish.
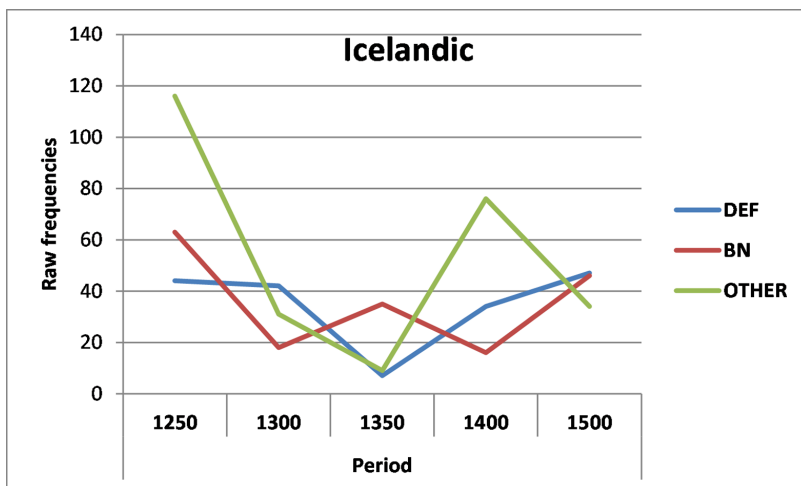
Icelandic

Figure 5 The raw frequencies of NP types (DEF, BN and OTHER) across time among NPs with direct anaphoric reference in Icelandic.

(13)   *Icelandic* (Vilhjálms saga Sjóðs, 1543)

Ríkharður   kóng-ur   hélt   mikla   skemmtun   á   að
Rikardur   king-NOM hold.PST great.F.ACC.SG enjoyment.F.ACC.SG on   to
fara   á   skóg […].   Þess   er   getið   eitt   sinn
travel to   forest.ACC.SG this.N.GEN be.PRS inform one.N time.N
að   **kóng-ur**   var   á   skóg   far-inn […].
that king.NOM be.PST in forest.ACC.SG gone-PTCP
'King Rikardur greatly enjoyed going to a forest […]. This is to inform of one time when the king was travelling in a forest.'

Lastly, *Tables 5* and *6* illustrate the make-up of the OTHER category in both contexts, namely NPs with unique reference and NPs with anaphoric reference.

Language	Period	Possessives	Demonstratives	Adjectives	Raw total
Danish	1200–1350	60.0% (6)	20.0% (2)	20.0% (2)	100.0% (10)
	1350–1450	50.0% (8)	50.0% (8)	0.0% (0)	100.0% (16)
	1450–1550	41.7% (5)	58.3% (7)	0.0% (0)	100.0% (12)
	Raw total	50.0% (19)	44.7% (17)	5.3% (2)	100.0% (38)
Swedish	1200–1350	100.0% (8)	0.0% (0)	0.0% (0)	100.0% (8)
	1350–1450	80.8% (21)	11.5% (3)	7.7% (2)	100.0% (26)
	1450–1550	75.0% (3)	25.0% (1)	0.0% (0)	100.0% (4)
	Raw total	84.2% (32)	10.5% (4)	5.3% (2)	100.0% (38)

Language	Period	Possessives	Demonstratives	Adjectives	Raw total
Danish	1200–1350	46.4% (52)	31.3% (35)	22.3% (25)	100.0% (112)
	1350–1450	36.7% (44)	52.5% (63)	10.8% (13)	100.0% (120)
	1450–1550	41.7% (43)	48.5% (50)	9.7% (10)	100.0% (103)
	Raw total	41.5% (139)	48.5% (148)	14.3% (48)	100.0% (335)
Swedish	1200–1350	54.5% (67)	27.6% (34)	17.9% (22)	100.0% (123)
	1350–1450	62.1% (54)	31.0% (27)	6.9% (6)	100.0% (87)
	1450–1550	45.5% (30)	53.0% (35)	1.5% (1)	100.0% (66)
	Raw total	54.7% (151)	34.8% (96)	10.5% (29)	100.0% (276)
Icelandic	1200–1350	8.2% (12)	61.9% (91)	29.9% (44)	100.0% (147)
	1350–1450	32.9% (28)	63.5% (54)	3.5% (3)	100.0% (85)
	1450–1550	20.6% (7)	61.8% (21)	17.6% (6)	100.0% (34)
	Raw total	17.7% (47)	62.4% (166)	19.9% (53)	100.0% (266)

Both uses, namely NPs with unique referents and NPs with direct anaphoric referents, differ in their use of OTHER constructions, i.e. neither the suffixed definite article nor bare NPs. As regards the OTHER category among NPs with unique referents (*Table 5*), it is quite homogenous as it almost exclusively includes definite-marked NPs other than NPs with the suffixed article. The data here is, however, too scarce to draw any conclusions. The OTHER category among NPs with direct anaphora (*Table 6*) displays a similar make-up, although more adjectival NPs with no determiners can be found here, especially in the first period.

The results of binary logistic regression are presented in the next section.

5.2 Time period as a factor influencing the presence of article type: binary logistic regression

To see how big an impact the time period has on the presence of a given article type (BN, DEF and OTHER) in NPs with unique referents and NPs with direct anaphoric referents, the results of binary logistic regression are reported in this section. We fitted binary logistic regression models with PERIOD as a single continuous independent variable.[8] All three languages are taken here together, since excluding the Icelandic data did not change the results in any significant way. This method allows one to build a predictive model that shows how great the probability of a particular article type is given the increase in the PERIOD variable (this information is provided by odds ratios). We use binary logistic regression instead of, for example, linear regression, as the dependent variable is categorical and dichotomous rather than continuous or ordinal.

The B Coefficient should be interpreted as the rate of change. As regards the model for BNs among NPs with unique referents (see *Table 7*), as the PERIOD variable increases by one unit (by one year), the log-odds of bare NPs occurring in this context decrease by 0.8%. Contrary, as PERIOD increases by one year, the log-odds of NPs with a definite article occurring among NPs with unique reference increase by 1.006 times, or in other words by 0.6%. We observe the same exact result for OTHER NP types in this context. The log-odds presented here are very small because the PERIOD variable is coded as continuous and thus the model estimates the log-odds for difference in one-year intervals. The overall trend is that the log-odds for BNs occurring continuously decrease by 0.8%, while the log-odds for DEF occurring continuously increase by 0.6% each year.

Regression model	Estimate (B Coefficient)	Std. Error	Significance	Odds ratios	Model accuracy
Intercept$_{BN}$	11.703	1.426	0.0001	120964.503	
Period$_{BN}$	−0.008	0.001	0.0001	0.992	73.9%
Intercept$_{DEF}$	−10.431	1.732	0.0001	0.00003	
Period$_{DEF}$	0.006	0.001	0.0001	1.006	85.5%
Intercept$_{OTHER}$	−10.661	1.813	0.0001	0.000023	
Period $_{OTHER}$	0.006	0.001	0.0001	1.006	87.0%

Table 7 Binary logistic regression models for NPs with unique referents.

We observe the same overall tendency among NPs with direct anaphoric referents (see *Table 8*). Firstly, as the PERIOD variable increases by one year, the log-odds of bare NPs occurring in this context decrease by 0.7%; the result here is thus nearly identical with the result for NPs with unique referents, although the decrease in log-odds for BNs occurring is marginally less steep here. Secondly, as PERIOD increases by one year, the log-odds of NPs with a definite article occurring among NPs with anaphoric reference increase by 1.004 times, or in other words by 0.4%. The increase of log-odds here is smaller than for NPs with unique referents. The ratios for the OTHER NP type are not statistically significant in this context.

Regression model	Estimate (B Coefficient)	Std. Error	Significance	Odds ratios	Model accuracy
Intercept$_{BN}$	7.911	0.844	0.0001	2727.947	
Period$_{BN}$	−0.007	0.001	0.0001	0.993	78.7%
Intercept$_{DEF}$	−6.413	0.660	0.0001	0.002	
Period$_{DEF}$	0.004	0.001	0.0001	1.004	63.7%
Intercept$_{OTHER}$	−0.637	0.609	0.296	0.529	
Period $_{OTHER}$	0.001	0.001	0.518	1.000	56.1%

Table 8 Binary logistic regression models for NPs with direct anaphoric referents.

8 We fitted binary regression models to three dependent variables in each context. The binary dependent variables correspond to the use/non-use of a given article type, for instance, all bare NPs in the dataset are coded as 'yes' in the BN variable, all other NPs are coded as 'no' in that variable. The same holds for the remaining variables: DEF and OTHER. The independent variable (PERIOD) is coded as a continuous variable.

Further, to check if PERIOD may have a greater impact on a given context (i.e. NPs with unique referents or NPs with direct anaphoric referents) in the selection of the definite article, we build a binary logistic model with DEF as a dependent variable and PERIOD (continuous variable) and CONTEXT (nominal variable: anaphoric vs. unique) as independent variables. Firstly, a simple scatter plot is presented with regression lines for both contexts in *Figure 6*. The y-axis corresponds to the predicted probability of the definite article occurring and the x-axis corresponds to the PERIOD variable; the data points represent the predicted probability of actual cases (as many cases have the same value, the dots are superimposed onto each other resulting with the plot with what seems like few data points). Both regression lines display the same trend, already presented in *Tables 7* and *8*, namely that as the years increase so does the probability for the definite article to occur. The slopes of the regression lines appear to be very similar, indicating that the context does not influence the outcome significantly, or in other words that PERIOD has the same effect on the presence of the definite article irrespective of the context.
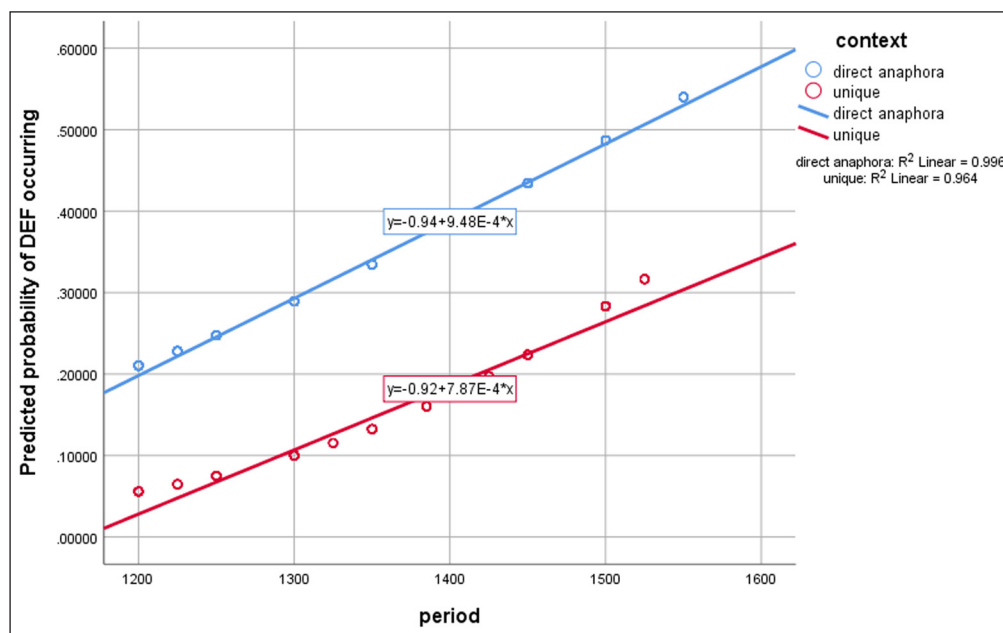

Figure 6 Scatter plot with regression lines for predicted probability of DEF by PERIOD in two contexts: NPs with unique referents and NPs with anaphoric referents.

Secondly, to check if the effect of PERIOD on the definite article is independent of context, an interaction term PERIOD*CONTEXT is added to the logistic regression model to check if there is a significant interaction between the two. The results are reported in *Table 9*.

Regression model	Estimate (B Coefficient)	Std. Error	Significance	Odds ratios	Model accuracy
Intercept	−10.431	1.732	0.0001	0.001	69.2%
Period	0.006	0.001	0.0001	1.006	
Context$_{DIR-A}$	4.018	1.854	0.03	55.575	
Period*Context	−0.002	0.001	0.118	0.998	

Table 9 Binary logistic regression model with DEF as a dependent variable and PERIOD, CONTEXT and an interaction term PERIOD*CONTEXT as independent variables.

The results indicate that, if PERIOD had a value of 0 (0 years) then NPs with direct anaphoric referents would be over 55 times more likely to occur with the definite article than the NPs with unique referents. The interaction term PERIOD*CONTEXT is the difference between the log-odds ratio corresponding to an increase in PERIOD by one year amongst NPs with direct anaphora and the log-odds ratio corresponding to an increase in PERIOD by one year amongst NPs with unique reference. The interaction between PERIOD and CONTEXT is not statistically significant ($p = 0.118$), proving that there is no statistically significant difference in regression slopes presented in *Figure 6*. Thus, even though CONTEXT is a significant variable on its own and the slopes illustrated in *Figure 6* have different constants (the line for NPs with anaphoric referents is consistently higher than that for NPs with unique referents), the PERIOD variable does not have a greater impact on either of the contexts.

Lastly, we will closely examine NPs with unique referents. As *Table 3* (see 5.1) illustrates, the number of bare NPs drops so significantly across the three periods in both Danish and Swedish (while at the same time the overall number of NPs with unique referents also decreases), that the proportion of NPs with the suffixed definite article is on the rise even though the raw numbers do not change much or actually drop (see *Figures 1* and *2*). Because NPs with a definite article are few compared to the number of bare NPs, finding lexical items that display a clear development from being predominantly bare in Period I to being predominantly definite in Period III is not feasible. For each lexical item among NPs with unique reference that occurs in the corpora more than once, such as *king, heaven, hell, bishop*, etc., bare NPs constitute a majority in each period. The only exceptions are lexical items *djævel* 'devil' and *wæruld* 'world', both of which are predominantly definite in the corpus.

(14)   *Danish* (Skriftemålsbøn, 1300)
thin   mykl miskundæligh […] at   thu fræls mik   af **diafl-s**   wald
2SG.POSS great mercy   that 2SG save 1SG.REFL of devil-GEN power
'In your great mercy […] that you save me from the devil's power.'

(15)   *Danish* (Aff Sancte Kerstine, 1450)
At   han   skal ether frelsse fran **deffuell-en**.
that 3SG.M shall 2.PL   save   from devil-DEF
'That he shall save you from the devil.'

(16)   *Swedish* (Codex Bureanus, 1300)
iosep egipt-æ   lan-z   hærra ok **wæruld-enna** helsara
Joseph Egypt-OBL land-GEN lord   and world-DEF.GEN saviour
'Joseph, the land of Egypt's lord and the world's saviour.'

With the exception of the two abovementioned lexical items, the NPs with unique referents are predominantly bare NPs.

Overall, the two contexts explored here, namely NPs with anaphoric and unique referents, show corresponding patterns, i.e. over time there appear more NPs with the definite article and fewer bare NPs. The NPs with the definite article gain in frequency with more or less the same pace for both contexts, as illustrated by the regression slopes in *Figure 6*. However, the frequency of NPs with a definite article is consistently higher in the context of direct anaphora than in the context of unique referents. The definite article was more frequently used already in Period I in anaphoric contexts than in unique contexts. While we cannot claim that the use of the definite article with anaphoric referents predates its use with unique referents, the differences in frequencies are quite striking. These differences may be due to the fact that the anaphoric context, while based on familiarity, may also satisfy uniqueness conditions, and would thus be more likely to appear as definite at the time of language change (see also Simonenko & Carlier 2020). This assumption is further explored in 5.3 and we will return to it in the concluding section.

5.3 Further analysis of NPs with unique referents: multinomial logistic regression

In this section we explore NPs with unique referents in more detail to see which factors apart from time period affect the presence of the suffixed definite article. Firstly, all the NPs with unique referents in our dataset can be divided into three domains based on what they refer to. This division is based on the semantics of the head noun. Each NP was thus annotated for the variable DOMAIN with one of the three categories. These categories are: nature (for example, *sun, earth, air, nature*, etc.), religion (*church, Bible, devil, heaven, hell, faith*, etc.), and law (*law, king, mayor, emperor*, etc.). The majority of these belong to the category of larger situation use (in terms of Hawkins 1978), which relies on general (and not specific) knowledge. Some NP referents, such as *the sun* or *the Pope*, may also be considered to be absolute uniques, unvarying for all people (see Lyons 1999: 8). The aim is to see if any of the three domains shows a significantly higher frequency of definite forms.

Further, in each domain we can discern referents that are either absolute uniques, namely uniques that rely on general encyclopaedic knowledge (what we call here 'global' uniques), and

those that rely on specific knowledge ('local' uniques). We annotated each NP with a unique referent for a variable UNIQUE TYPE with two categories, namely local and global uniques. To operationalize the annotation of this variable we assumed that all NP referents common for at least the whole country (for instance, *the king, the religion, the earth*) are global uniques, and those that are common for smaller groups of people, such as towns or villages are local uniques (for instance, *the chieftain, the priest, the forest*).

We hypothesize that, since the definite article was not fully grammaticalized in the unique context in the time periods we study, it would occur more often in contexts that rely on both familiarity and uniqueness semantics (see also 5.2 and 6). NPs with local unique referents, while clearly unique in the sense of having no direct or indirect antecedents, can still fulfil the familiarity condition in the sense of pragmatic familiarity, in which the referent is linked to the utterance situation. We expect that NPs with local uniques will occur in definite form more frequently than those with global uniques, since they rely on the familiarity with the context. As for which domain might attract the definite form more strongly, we hypothesize that the domain of law will exhibit a higher frequency of definite form. This hypothesis is, however, based again on the local/global distinction; the domain of law has the highest proportion of local unique referents compared to other domains,[9] thus the variable of DOMAIN if significant might be epiphenomenal to the variable of UNIQUE TYPE.

To check these hypotheses we use a multinomial logistic regression model with ARTICLE TYPE as the dependent variable (with three categories: BN, DEF and OTHER). Multinomial regression allows us to choose a reference category of the dependent variable to which the other two categories are compared; here we choose BN as the reference group. The model tests the probability of definite forms (DEF) occurring relative to BNs, and then OTHER forms occurring relative to BNs, as a function of different predictor factors such as LANGUAGE, GENRE, DOMAIN and UNIQUE TYPE. What is important is that the regression model provides information on which of the variables is the most prominent, while it simultaneously controls for all other factors in the model. *Table 10* illustrates the regression model for NPs with unique referents. Statistically significant results are in bold.

Regression model	Independent variables	Estimate (B Coefficient)	Std. Error	Significance	Odds ratios
DEF vs. BN	Intercept	0.596	0.558	0.286	–
	LANGUAGE = Danish	−0.645	0.312	**0.039**	**0.525**
	LANGUAGE = Icelandic	0.437	0.772	0.571	1.548
	GENRE = legal	−0.791	0.934	0.397	0.453
	GENRE = religious	1.159	0.534	**0.030**	**3.186**
	DOMAIN = law	−0.714	0.743	0.337	0.490
	DOMAIN = religion	−1.453	0.415	**0.0004**	**0.234**
	UNIQUE TYPE = local	1.530	0.357	**0.0001**	**4.618**
OTHER vs. BN	Intercept	−1.505	0.694	0.030	–
	LANGUAGE = Danish	−0.689	0.318	**0.030**	**0.502**
	LANGUAGE = Icelandic	1.076	0.694	0.121	2.933
	GENRE = legal	−24.782	0.000	0.998	0.000
	GENRE = religious	0.807	0.537	0.133	2.242
	DOMAIN = law	0.738	0.752	0.327	2.091
	DOMAIN = religion	0.828	0.534	0.121	2.288
	UNIQUE TYPE = local	1.724	0.627	**0.006**	**5.608**

Table 10 Multinomial logistic regression model for NPs with unique referents with the following reference categories: LANGUAGE = Swedish, GENRE = profane, DOMAIN = nature, UNIQUE TYPE = global.

---

9   In the domain of law 22.5% (87 out of 387) of all NPs are annotated as local. In the domain of religion 4.9% (12 out of 243) of all NPs are annotated as local. In the nature domain 6.4% (3 out of 47) of all NPs are annotated as local.

Overall, all the predictor variables entered into the model are statistically significant.[10] The factor of PERIOD is also included and controlled for, but we omit it in *Table 10* so as not to redundantly repeat the results presented in 5.2.

As regards NPs with a definite article relative to bare NPs, they are less likely to occur in Danish than in Swedish corpus texts. They are also over 3 times more likely to appear in religious texts than in profane texts. As for DOMAIN, while all the other factors are held constant, NPs assigned to the religion category are significantly less likely to be definite than NPs in the nature category, to be exact there are 76.6% less odds of the definite article occurring with the religion category than with the nature category relative to BNs. Lastly, NPs favouring definite forms rather than bare forms are 4.6 times more likely to occur with local unique referents than with global unique referents.

As regards OTHER NPs relative to BNs, only two variables are selected as significant here. OTHER NPs relative to BNs are less likely to occur in Danish texts than in Swedish texts. OTHER NPs also strongly favour local unique referents; they are 5.6 times more likely to occur in local contexts than in global contexts.

The model's predictive accuracy is 74.9%; the model is very good at predicting BNs (95.9% of cases classified correctly), but not as felicitous at predicting DEF or OTHER (respectively 18.4% and 20.5% of cases predicted correctly). As far as the relative variable importance is concerned, GENRE proves to be the strongest predictor factor in the dataset (legal texts very strongly favour BNs to other article types, which confirms the somewhat archaic character of these texts and the specific features of the genre as such, which is characterized by, among others, high frequency of BNs in Modern Swedish as well, cf. Gunnarsson 1982), followed by DOMAIN, PERIOD (see section 5.2), and lastly UNIQUE TYPE.

In conclusion, the variable of DOMAIN is not epiphenomenal to the variable of UNIQUE TYPE (the local/global distinction), as it is the category of nature that strongly favours definite forms and this category is in no way correlated with local uniques (i.e. another category that strongly favours definite forms). In the nature domain we find a relatively significant amount of NPs with a definite article with referents that are considered to be absolute uniques, common to all people, such as *the sun, the earth, the world*, etc. These lexical items, in particular *the earth* and *the world*, are high frequency NPs; together they constitute 48.9% (23 out of 47 examples) of all examples within the nature category, which in itself is a relatively small category in the dataset. The nature group displays thus a relatively high frequency of the definite article with global unique referents, but since the category is a small part of the dataset, it does not change the fact that local unique referents relative to global referents strongly favour NPs with a definite article.

While the model controls for the variable of DOMAIN, the variable of UNIQUE TYPE is highly significant confirming the hypothesis that NPs with unique referents fulfilling both familiarity and uniqueness conditions (i.e. local uniques) are significantly more likely to appear in the definite form or OTHER forms (which predominantly include other determiners as we have mentioned in section 5.1) than in bare forms. We now turn to a closer examination of a particular context in the corpus texts, namely co-ordinated NPs of the type *to the king, the bishop and the district,* which are regularly found in legal prose, stipulating to whom taxes or fines were to be paid. Since the payments were typically divided three-ways, such co-ordinated NPs join a number of referents which are 'differently' unique. In Period I NPs with unique referents tend to be unmarked and appear as BNs, irrespective of the scope of their uniqueness, as in 17, where two local unique referents are followed by a global referent.

(17)   *Swedish* (Äldre Västgötalagen KB: 7, 1225)
       Uerdher kyrkia brut-in   oc   mæss-u   fat   stol-en […] þat er
       be.PRS   church break-PTCP and mass-OBL plate steal-PTCP this be.PRS
       niv   march-a sak **kyrky** swa **hæreþe** sva **konogge**.
       nine mark-PL fine church so   district so   king
       'If a church is broken into and the Mass plate stolen […]. It is nine marks fine, to
       church, and so to district and so to king.'

---

10   Since BN is a reference category we do not see the significance and odds ratios for BNs in *Table 10*, but this does not mean that independent variables are not significantly predicting the appearance of BNs as well. Hence even the factors that are not selected as significant for DEF or OTHER may still significantly contribute to the model.

While the unique referent 'king' remains bare throughout all the periods and only sporadically appears with a definite article, more local referents, such as *höfding* 'chieftain' or *prest* 'priest' show higher frequencies of definite forms. Local unique referents, which in certain contexts may also be considered to be indirect anaphors (and thus they satisfy familiarity semantics as well as uniqueness semantics), appear with a definite article more frequently, while more global referents appear in bare forms, as in 18.

(18)   *Swedish* (Yngre Västgötalagen, KB II, 1280)
      en houothtinda skal skipta i thre lyte   en lot-en
      one tithe   shall divide in three part.PL one part-DEF
      a **biscuper** annan a **kirkia-n**.   thrithia **præst-en**
      to bishop   other to church-DEF third   priest-DEF
      'A tithe shall be divided in three parts, one part goes to the bishop, another to the church and the third to the priest.'

The statistical analysis presented here reveals a crucial fact about NPs with unique referents in North Germanic, namely, that the definite article favours NPs that provide a bridging context between familiarity and uniqueness semantics to those that rely solely on uniqueness. Here by familiarity we do not mean that an NP has an antecedent (Novelty-Familiarity-Condition in terms of Heim 1988), but rather familiarity with the context of the situation, where the NP referent is associated with the specific knowledge shared by the speaker and hearer (in terms of Christophersen 1939 and Jespersen 1943). We argue that NPs with local unique referents provide a bridging context between associative anaphors and global unique referents, as they clearly rely on a degree of familiarity and are directly linked to the utterance situation, which is not the case for globally unique referents. We have shown that NPs with unique referents are not a homogenous group, but rather that within this group the grammaticalization of the definite article also proceeds from familiar to unique semantics.

6 Discussion and conclusions

There is a clear difference in the frequency of the definite article in NPs with different types of referents in North Germanic historical texts. NPs with referents that are familiar, textually grounded, namely those with direct anaphoric reference, show a higher frequency of definite forms than NPs with unique referents. The higher frequency is observed already in the first extant texts (from ca. 1200) and continues throughout the later periods, which confirms the hypothesis that the grammaticalization of the definite article proceeds from co-referring NPs to unique NPs. Naturally, the development does not proceed in discrete steps, namely in one text some unique referents may already be definite while some direct anaphors may still appear as BNs, however, the frequencies indicate that a higher percentage of the direct anaphors will be definite than that of the uniques. Among the uniques, those from different domains and of different scope of uniqueness display partly different patterns. The local unique referents, reliant on uniqueness but also a degree of familiarity, appear as definites more frequently than those that are globally unique. These NPs provide a bridging context in the later stages of the grammaticalization, as they are not exclusively reliant on uniqueness and thus retain their links to the original demonstrative.

In section 5.2 we noted that the higher frequency of the definite article in anaphoric contexts than in unique contexts could be due to the fact that the anaphoric contexts, while based on familiarity, also satisfy uniqueness conditions. This is the crucial difference between the definite article and the demonstrative from which it has developed. The demonstrative can be used to distinguish the intended referent from similar referents (e.g. *this professor* vs. *that professor*), but with the definite article there must only be one referent for the defNP, e.g. *the professor*, to be unambiguous (Hawkins 1978: 157). The direct anaphoric contexts are based on familiarity, but the uniqueness condition must be satisfied as well, thus we may expect that in the course of definite article formation such contexts will be earlier to adopt the incipient definite, as it still retains its demonstrative features and acquires the definite article ones.

Based on the results in 5.2 we may state that the grammaticalization of the definite article proceeds from uses that are grounded in familiarity to those grounded in uniqueness. However,

it is important to bear in mind that while the definite article in direct anaphoric uses differs from the demonstrative by putting a uniqueness condition on the referent, the direct anaphoric use is still firmly grounded in the text. In 5.3 we considered the larger situation uses, which are not grounded in the text, but in which new referents may be introduced as defNPs. Within the larger situation uses, the more local unique referents (i.e. those based on specific knowledge), which could have been interpreted as indirect anaphors and thus textually anchored, seem to have served as a bridge from textual to non-textual definiteness in the grammaticalization of the definite article. The culmination of the grammaticalization is the use of the definite article with globally unique referents, a usage that is at odds with the original meaning of the demonstrative delimiting the referent from other potential referents.

Abbreviations

1, 2, 3 = 1st, 2nd, 3rd person, ACC = accusative, DEF = definite article, F = feminine gender, GEN = genitive, INDF = indefinite article, M = masculine gender, N = neuter gender, NOM = nominative, PL = plural, POSS = possessive, PRS = present tense, PST = past tense, PTCP = participle, REFL = reflexive, SG = singular, WK = weak.

Acknowledgements

Funding information

Competing interests

The authors have no competing interests to declare.

Author affiliations

Alicja Piotrowska *orcid.org/0000-0002-9868-2964*
Adam Mickiewicz University, Poznań, al. Niepodległości 4, 61–874, Poznań, Poland
Dominika Skrzypek *orcid.org/0000-0002-9277-4622*
Adam Mickiewicz University, Poznań, al. Niepodległości 4, 61–874, Poznań, Poland

Source texts

Aff Sancte Kerstine hennis pyne. In Carl J. Brandt (ed.), *De hellige Kvinder, en Legende-Samling*. Copenhagen, 1859, 38–51. Manuscript Cod. Holm. K4.

Äldre Västgötalagen. In Hans S. Collin & Carl J. Schlyter (eds.), *Samling af Sweriges gamla lagar* (Vol. 1). Stockholm: Haeggström, 1827. Manuscript Holm B 59.

Skriftemålsbøn. Marita Akhøj Nielsen (ed.). Copenhagen: Det Danske Sprog- og Litteraturselskab, 2015. Manuscript K 48.

Vilhjálms saga Sjóðs. In Agnete Loth (ed.), *Late Medieval Icelandic Romances IV*. Copenhagen: Munksgaard, 1964.

Yngre Västgötalagen. In Hans S. Collin & Carl J. Schlyter (eds.), *Samling af Sweriges gamla lagar* (Vol. 1). Stockholm: Haeggström, 1827. Manuscript Holm B 58.

References

Asudeh, Ash & Line Hove Mikkelsen. 2000. Incorporation in Danish: Implications for interfaces. In Ronnie Cann, Claire Grover & Philip Miller (eds.), *A Collection of Papers on Head-Driven Phrase Structure Grammar*, 1–15. Stanford: CSLI Publications.

Breu, Walter. 2004. Der definite Artikel in der obersorbischen Umgangssprache. In Christian Sappok & Marion Krause (eds.), *Slavistische Linguistik 2002*, 9–57. Munich: Otto Sagner.

Bybee, Joan L., Revere D. Perkins & William Pagliuca. 1994. *Evolution of grammar: Tense, aspect, and modality in the languages of the world.* Chicago: University of Chicago Press.

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

Dahl, Östen. 2015. *Grammaticalization in the North: Noun phrase morphosyntax in Scandinavian vernaculars.* Language Science Press. DOI: *https://doi.org/10.26530/OAPEN_559871*

De Mulder, Walter & Anne Carlier. 2011. The grammaticalization of definite articles. In Heiko Narrog & Bernd Heine (eds.), *The Oxford Handbook of Grammaticalization*, 522–535. Oxford: Oxford University Press. DOI: *https://doi.org/10.1093/oxfordhb/9780199586783.013.0042*

Diessel, Holger. 1999. The morphosyntax of demonstratives in synchrony and diachrony. *Linguistic Typology* 3. 1–49. DOI: *https://doi.org/10.1515/lity.1999.3.1.1*

Ebert, Karen. 1971a. Zwei Formen des bestimmten Artikels. In Dieter Wunderlich (ed.), *Probleme und Fortschritte der Transformationsgrammatik*, 159–174. Munich: Hueber.

Ebert, Karen. 1971b. *Referenz, Sprechsituation und die bestimmten Artikel in einem nordfriesischen Dialekt (Fehring).* Kiel: Christian-Albrechts-Universität zu Kiel dissertation.

Fraurud, Kari. 1990. Definiteness and processing of noun phrases in natural discourse. *Journal of Semantics* 7. 395–433. DOI: *https://doi.org/10.1093/jos/7.4.395*

Gunnarsson, Britt-Louise. 1982. *Lagtexters begriplighet* [The comprehensibility of legal texts]. Stockholm: LiberFörlag.

Hartmann, Dietrich. 1967. *Studien zum bestimmten Artikel in 'Morant und Galie' und anderen rheinischen Denkmälern des Mittelalters.* Giessen: Wilhelm Schmitz Verlag.

Hartmann, Dietrich. 1978. Verschmelzungen als Varianten des bestimmten Artikels? In Dietrich Hartmann, Hansjürgen Linke, & Otto Ludwig (eds.), *Sprache in Gegenwart und Geschichte*; *Festschrift Für Heinrich Matthias Heinrichs*, 68–81. Cologne: Böhlau.

Hawkins, John. 1978. *Definiteness and Indefinitenss: A study in Reference and Grammaticality Prediction.* London: Croom Helm.

Heim, Irene. 1988. *The Semantics of Definite and Indefinite Noun Phrases.* New York & London: Garland Publishing, Inc.

Himmelmann, Nikolaus P. 1997. *Deiktikon, Artikel, Nominalphrase: Zur Emergenz syntaktischer Struktur* (Linguistische Arbeiten 362). Tübingen: Niemeyer. DOI: *https://doi.org/10.1515/9783110929621*

Jespersen, Otto. 1943. *A modern English grammar on historical principles: Part 7: Syntax.* Copenhagen: Munksgaard.

Löbner, Sebastian. 2003. *Definite Associative Anaphora.* Düsseldorf: Heinrich-Heine-Universität, Ms.

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

Perridon, Harry. 1989. *Reference, definiteness and the noun phrase in Swedish.* Amsterdam: University of Amsterdam dissertation.

Pettersson, Thore. 1976. Bestämda och obestämda former [Definite and indefinite forms]. In Eva Gårding (ed.), *Kontrastiv fonetik och syntax med svenska i centrum* [Contrastive phonetics and syntax with Swedish in focus]. Lund: Liber Läromedel.

Roberts, Craige. 2003. Uniqueness in definite noun phrases. *Linguistics and Philosophy* 26. 287–350. DOI: *https://doi.org/10.1023/A:1024157132393*

Russell, Bertrand. 1905. On denoting. *Mind* 14. 479–493. DOI: *https://doi.org/10.1093/mind/XIV.4.479*

Schwarz, Florian. 2009. *Two types of definites in natural language.* Amherst: University of Massachusetts Amherst dissertation.

Simonenko, Alexandra & Anne Carlier. 2016. The evolution of the French definite article: from strong to weak. (*https://www.academia.edu/28473104/The_evolution_of_the_French_definite_article_from_strong_to_weak*) (Accessed 2019-02-10).

Simonenko, Alexandra & Anne Carlier. 2020. Between demonstrative and definite: A grammar competition model of the evolution of French l-determiners. *Canadian Journal of Linguistics* 65(3), 393–437. DOI: *https://doi.org/10.1017/cnj.2020.14*

Skrzypek, Dominika. 2012. *Grammaticalization of (in)definiteness in Swedish.* Poznań: Wydawnictwo Naukowe UAM.

Skrzypek, Dominika. 2020. Indirect anaphora in a diachronic perspective: The case of Danish and Swedish. In Robert Van Valin & Kata Balogh (eds.), *Nominal Anchoring*, 171–193. Berlin: Language Science Press.

Strawson, Peter F. 1950. On Referring. *Mind* 59: 320–344. DOI: *https://doi.org/10.1093/mind/LIX.235.320*

Takamine, Kaori. 2014. DP-decomposition Analysis of Japanese Demonstrative so. In Gianina
Iordăchioaia, Isabelle Roy & Kaori Takamine (eds.), *Categorization and category change*, 33–58.
Cambridge: Cambridge Scholars Publishing.

Teleman, Ulf, Staffan Hellberg & Erik Andersson. 2010. *Svenska Akademiens Grammatik: 3 Fraser* [The
Swedish Academy Grammar: 3. Phrases]. Stockholm: Norstedts.

Wespel, Johannes. 2008. *Descriptions and their domains: the patterns of definiteness marking in French-related
creoles.* Stuttgart: University of Stuttgart dissertation. *https://www.sfb732.uni-stuttgart.de/documents/
files/sinspec2_disswespel.pdf*.