

## *An* from Old to Middle English

Paola Crisma, Università di Trieste, University of York  
Susan Pintzuk, University of York

### Abstract

In this pilot study<sup>1</sup> we build on the research of Crisma (2015), who outlines three stages in the development of *a(n)* in the history of English: In Stage One, *an* is simply the numeral ‘one’; in Stage Two, *a(n)* acts as an overt existential operator in certain contexts; in Stage Three, *a(n)* is reduced to an expletive. Stages One and Two are attested in the Old English period; Stage Three is the Present-Day English system. In this study we investigate the use of *a(n)* in Middle English texts, to determine whether Stage Two continues into Middle English, and whether we can see the beginning of Stage Three. We find that the distribution of *a(n)* provides evidence that Stage Two is attested in texts written in the first Middle English period (M1, 1150-1250), while characteristics of Stage Three appear in texts from the third period (M3, 1350-1420).

### 1. Introduction

Crisma (2015) outlines the development of *an* in the history of English from its original status as the numeral *one* to its present status as an ‘indefinite article’. Three different stages are recognised in this trajectory: in Stage One, attested by some early Old English (OE) West Saxon texts (Bede’s *Ecclesiastical History* and Alfred’s *Laws*), the grammar of *an* is similar to that of modern Icelandic, with *an* simply the numeral ‘one’. Stage Two, observable in OE in Ælfric’s *Lives of Saints* and, surprisingly,

---

<sup>1</sup> We thank an anonymous reviewer for helpful comments and suggestions. All errors and omissions remain our own.

in a very innovative *Orosius*,<sup>2</sup> is similar to modern Hebrew and is characterised by the fact that only *overt* operators can take wide scope; here, *an* acts as an overt existential operator whenever an indefinite noun phrase is interpreted as specific or when it takes wide scope over some other operator. Stage Three, where *a(n)* is reduced to an expletive, is represented in Crisma (2015) by Present-Day English (PDE), with an unanalysed gap of almost a millennium between the establishment of Stage Two and the current Stage Three system. The primary aim of this chapter is to replace this millennium gap with a realistic dating of the earliest evidence for Stage Three. This is a descriptive pilot study: we examine data samples from Middle English (ME) texts to determine whether 1) Stage Two characteristics persist throughout the period, and/or 2) we can see the beginning of the Stage Three PDE system in the ME data.

One striking characteristic of the use of *an* in OE is that it is totally incompatible with a generic interpretation: thus, sentences like *A dog has four legs* are never encountered in OE data from either Stage One or Stage Two. In Table 1 in Crisma (2015:133), there are 62 tokens of noun phrases headed by a singular count noun that receive a generic interpretation, none of them introduced by *an*.<sup>3</sup> This distribution was taken as evidence that the numeral *an* of Stage One was re-analysed as an existential operator at Stage Two, hence the incompatibility with the generic reading. Thus, the telltale piece of evidence that marks the change from Stage Two to Stage Three of the development of the indefinite article is the presence of *a(n)* introducing generic singular noun phrases. This is true independently of how the indefinite article is analysed in PDE, whether it is a classifier as in Borer (2005); or a marker of the count reading as in Ackles (1997), which focuses on *an* in the history of English; or an expletive as suggested in

<sup>2</sup> It is surprising because *Orosius* is a fairly early OE text, written c898, while *Lives of Saints* was written c997.

<sup>3</sup> This of course does not mean that only bare singulars could be interpreted as generics in OE. Crisma (2011) argues that bare plurals in OE had basically the same interpretive properties as in PDE, since they could be generic as well as existential. It remains an open issue whether OE noun phrases introduced by *se*, the ancestor of *the* and *that*, could also be interpreted as generic: in the singular, as in PDE, or in the plural, as in modern German, as shown in (i) and (ii) respectively. Example (ii) is from Zamparelli 2000: 165, his example (430), citing Brugger (1993); the original spelling is preserved here.

(i) **The dolphin** is a mammal

(ii) ... daß        **(die) Elefanten**   wertvolle   Zähne   haben  
       ... *that*        *(the) elephants*   *precious*   *teeth*   *have*

‘... that elephants have valuable teeth’

Crisma (2015). Although there is little consensus on the formal analysis of the indefinite article, it is still possible to determine the point in the history of English when the present situation establishes itself, with *a(n)* becoming compatible with generic singular count nouns and in fact obligatory in that context unless the definite article is used.

## 2. Selecting the Middle English Texts

For this initial study of the use of *a(n)* in ME, we used texts from the Penn-Helsinki Parsed Corpus of Middle English, second edition (PPCME2).<sup>4</sup> The PPCME2 is divided into four time periods, each covering 70-100 years: M1 (1150-1250), M2 (1250-1350),<sup>5</sup> M3 (1350-1420), and M4 (1420-1500). For most texts, the date of composition and the date of the manuscript fall within the same period; for those texts which are composed in one period with the earliest manuscript from a later period, two periods are specified, e.g. M34 (written in M3 with an M4 manuscript) for *The Book of Vices and Virtues* (see below).

We selected texts from the first three ME periods using several criteria:

- 1) size: larger texts result in larger numbers of tokens from each text, so that texts and periods can be reliably characterised with respect to the use of *a(n)*;
- 2) genre: we chose to start this study using prose texts rather than poetry;
- 3) dialect: except for *Ayenbite of Inwyrt* (Kentish), we limited our investigation to texts from the East and West Midlands.

From period M1 we included *Peterborough Chronicle* (c1150);<sup>6</sup> *Katherine Group* (c1225), consisting of five smaller texts (*Sawles Warde*, *Hali Meidhad*, *St. Katherine*, *St. Juliana*, and *St. Margaret*); and a sample of *Ancrene Riwe* (c1230). From period M2, we included a sample of *Ayenbite of Inwyrt* (1340); from period M3, two texts from Chaucer (*Tale of Melibee* and *The Parson's Tale*, both c1390) and a sample from one M34 text, *The Book of Vices and Virtues* (c1450). We had a particular point in including *Ayenbite* (M2) and *Vices* (M34): both texts are translations from

<sup>4</sup> For more detailed information about the corpus and the individual texts, see the PPCME2 website: <http://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-3/index.html>

<sup>5</sup> It should be noted that the second ME period (ME2) is not well-represented; the data from ME2 therefore should not be used without caution.

<sup>6</sup> *c* (= *circa*) indicates a date preceding or following the given date by 25 years.

the Old French *Somme le Roi*, and the samples we included from each were translated from the same passages of the French text.

### 3. Collecting the Evidence

The collection of a database of ME nominals that would serve as the empirical basis for the present study was largely modelled on the work done on OE in Crisma (2015). Differences, which are discussed below, were due to differences in the annotation of the two parsed corpora and to a slightly different perspective in interpreting the data.

The first step was to collect all nominal arguments – subjects and objects<sup>7</sup> – with a morphologically singular head noun from the selected PPCME2 texts. These were collected in two different sets: nominals in the first set were those introduced by *a(n)*, while nominals in the second set were ‘bare’ in the sense of Carlson (1977); that is, they were not introduced by numerals, articles, demonstratives, quantifiers, possessives or genitives. During the coding phase, we decided to treat separately nominals headed by *man* and *thing*; the former is extensively discussed in section 5. The nominals were then manually coded according to their denotation properties, using in most cases the same labels used in Crisma (2015), with a few notable exceptions discussed below. Using this procedure, each noun phrase in the sample that fit the criteria is associated with a composite label describing the various properties that are relevant for the present study.

The first element of the label encodes whether the noun phrase is introduced by *a(n)* (AN) or is a bare singular count noun (BSG). At this stage of the coding, bare mass nouns were coded as (BMS) and discarded. Note that deciding whether a particular instance of a singular noun is mass or count is not always straightforward. So, alongside clear cases such as (1), there are dubious cases as in (2):

- (1) clear mass noun (CODE <NPTYPE:BMS>)
- it is the endeles blisse of hevene, ther **joye** hath no  
*it is the endless bliss of heaven, where joy has no*
- contrarioustee of wo ne grevaunce  
*adversity of woe nor grievance*  
 (CMCTPARS,327.C2.1673)

<sup>7</sup> Crisma’s (2015) object data included only direct objects, since indirect objects are not annotated as such in the York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE). However, the PPCME2, unlike the YCOE, unambiguously labels indirect objects, and therefore these objects could be easily and automatically included in the database.

- (2) ambiguous between mass noun and count noun (CODE <NPTYPE:BN>)  
 He seide þæt **fulle feoht** was sett  
*He said that utter war / (a) furious battle was created*  
 betweneŋ ða cristene & þa heðene  
*between the Christians and the heathens*  
 (CMPETERB,51.275)

All dubious cases were coded BN ('bare noun'), as shown in (2), and discarded along with mass nouns, to avoid including ambiguous cases in the database of otherwise unambiguous tokens. On the other hand, clear cases of coercion, where a normally mass noun was converted to a count noun by the presence of *a(n)*, were coded AN, as in (3), and included in the database for the study.

- (3) coercion of mass noun to count noun (CODE <NPTYPE:AN-EXS>)  
 And by the manere of his speche it semedthat in  
*And by the manner of his speech it seemed that in (his)*  
 herte he baar **a crueelire**  
*heart he carried a cruel anger*  
 (CMCTMELI,218.C2.56)

The second element of the label indicates whether the nominal has an existential (EXS) interpretation, as in (3) and (4), or a generic (GNR) one, as in (5).

- (4) a. existential interpretation (CODE <NPTYPE:BSG-EXS>)  
 Eue heold iparais **long tale** wið  
*Eve held in-paradise (a) long conversation with*  
 þe neddre  
*the serpent*  
 (CMANCRIW-1,II.54.520)
- b. existential interpretation (CODE <NPTYPE:AN-EXS>)  
 As ha þeos bone hefde ibeden com **a kempe** of  
*When she this plea had made came a champion from*  
 helle on englene heowe  
*hell in angel's guise*

‘When she had made this plea, a champion came from hell in the guise of an angel’ (CMJULIA,107.187)

- (5) a. generic interpretation (CODE <NPTYPE:BSG-GNR>)  
 þu seist þt muce confort haueð **wif** of hire were  
*you say that much comfort has (a) wife from her husband*  
 ‘you say that a wife has much comfort from her husband ...’  
 (CMHALI,147.282)
- b. generic interpretation (CODE <NPTYPE:AN-GNR>)  
 Certes **a shadwe** hath the liknesse of the thyng  
*Certainly a shadow has the likeness of the thing*  
 of which it is shadwe  
*of which it is (the) shadow*  
 (CMCTPARS,292.C2.187)

Again, it was sometimes not possible to decide between the two alternative readings, existential and generic, because some sentences were genuinely ambiguous between the two, as shown in (6). Such cases were tagged AMB.

- (6) ambiguous between existential and generic (CODE<NPTYPE:AN-AMB>)  
 And therfore seith **a philosophre** in this wise  
*And therefore says a philosopher in this manner*  
 ‘And therefore a (particular) philosopher says / philosophers say in this manner ...’  
 (CMCTMELI,224.C1.277)

As in Crisma (2015), labels for nominals with an existential interpretation may encode some additional information. First, there is the possibility that overt material, such as a (reduced) relative clause, makes the noun phrase clearly specific, as in (7); this is indicated by the additional tag SPC.

- (7) existential specific nominal (CODE <NPTYPE:AN-EXS-SPC>)  
**Ayong man** called Melibeus myghty andriche  
*Ayoungman called Melibee mighty and rich*  
 bigat upon his wyf that called was Prudence a  
*begat upon his wife who called was Prudence a*  
**doghter** which that called was Sophie  
*daughter who that called was Sophie*  
 ‘A young man called Melibee, mighty and rich, begat upon his  
 wife, who was called Prudence, a daughter who was called Sophie.’  
 (CMCTMELI,217.C1b.5)

Second, the nominal may co-occur with elements potentially participating in scope interactions, such as negation, quantifiers, but also *irrealis* modality or certain verbs in the matrix clause (some epistemic verbs, commands, etc.). For these cases, the tag SCOPE or NG (for ‘negation’)<sup>8</sup> was added, followed by -nrw, -wd or -amb, depending on whether the nominal had narrow or wide scope or whether the scope was ambiguous:

<sup>8</sup> Overall, we found 10 cases of indefinites in the scope of a negative element, four introduced by *a(n)* and six bare singulars. This very low figure is expected, given that English continues to be a (strict or non-strict) Negative Concord (NC) language into the 15th century (see Ingham 2013); thus, the vast majority of subjects and objects in the scope of negation are introduced by *nan/non*, as in (i):

- (i) ach hare wununge n-aeð nan ʒete.  
 but their dwelling NEG-has no gate  
 ‘but their dwelling has no gate’ (CMANCRIW-1.II.60.609)

In the earliest texts (M1 and MX1), the form is consistently spelled *na(n)*, which might be analysed as a negated *a(n)*. However, it would not be correct to count it as a possible variant of *a(n)* alternating with bare singulars: in negative clauses displaying NC, in fact, there is no possible alternation between two forms, since nominals introduced by *na(n)* are the only option. It is only in clauses without NC that there is a choice between a bare singular and *a(n)+N*. Establishing what determines the absence of NC in the 10 cases we included in our study goes beyond the scope of the present work, but it is clear that these cases must be kept separate from nominal phrases introduced by *na(n)*. It is also worth mentioning that the problem exists only for the earliest ME stages: starting from M2, the negative determiner is consistently spelled with the vowel <o>, that is, it is consistently *no(n)* throughout all the texts included in PPCME2 with the exception of Northern texts, while the (ancestor of) the indefinite article is regularly spelt *a(n)*.

- (8) existential nominal with narrow scope (CODE <NPTYPE:AN-EXS-SCOPE-nrw>)  
 Ich chulle lete makie þe of golt **an ymage** as cwen icrunet  
*I shall let make thee of gold an image as queen crowned*  
 ‘I will have a golden image made of you as a crowned queen’  
 (CMKATHE,36.269)
- (9) existential nominal with wide scope (CODE: <NPTYPE:AN-SCOPE-wd>)  
 & seide to hire þus. haue cwen **acrune** isent te of  
*and said to her thus. have queen a-crown sent to-you from*  
 heouene  
*heaven*  
 ‘and (he) said the following to her: Queen, have a crown, sent to you from heaven’  
 (CMKATHE,38.308)
- (10) existential nominal with ambiguous scope (CODE: <NPTYPE:AN-SCOPE-amb>)  
 thanne seketh he **an ydel solas** of worldly thynges  
*then seeks he a useless consolation from worldly things*  
 ‘then he seeks a useless consolation from worldly things’  
 (CMCTPARS,313.C1.1073)

It should be noted that in scope interactions between logical connectives, a universal quantifier with wide scope is logically equivalent to an existential quantifier with narrow scope. Thus, the sentence *A dog does not like lettuce* has two logically equivalent representations, one with a universal quantifier<sup>9</sup> taking scope over the negation and the other with the negation taking scope over an existential quantifier. This equivalence proved to be particularly challenging for the coding of the ME texts, because tokens of this type were very frequent; the fact that we were not able to decide between the two alternatives meant that we would have needed to discard many potentially revealing examples. Therefore a new tag was introduced for these cases that was used in the place of EXS and GNR: NPE, for ‘No Presupposition of Existence’. This tag captures what generics and existentials with narrow scope have in common, namely that

<sup>9</sup> To be precise, the operator in question is Gen rather than the universal quantifier.

they do not presuppose the existence of the referent. It is no coincidence that in OE, at Stages One and Two in the development of *a(n)*, it is precisely in nominals that are generic or existential with narrow scope that *a(n)* is either categorically excluded (generics) or certainly disfavoured (narrow scope existentials).

- (11) nominals that are ambiguous, either generic or narrow scope existential
- a. (CODE <NPTYPE:BSG-NPE>  
 tis put he hat þt beo ilided þt **beast** þrin  
*this pit he commanded that be covered that beast therein*  
 ne falle  
*NEG fall*  
 ‘he commanded that this pit be covered lest (a) beast fall therein’  
 (CMANCRIW-1,II.48.446)
- b. (CODE <NPTYPE:AN-NPE>  
 whan a **gret lord** hap no child he may chese a pore  
*when a great lord has no child he may choose a poor*  
 mannes sone 3if he wole and make of hym his eir  
*man’s son if he will and make of him his heir*  
 bi adopcioun  
*by adoption*  
 ‘when a great lord has no child, he may choose a poor man’s son  
 if he wants and make him his heir by adoption’  
 (CMVICES4,100.63)

The total counts for nominals coded in this way are presented in Table 1, which is modelled after Table 1 in Crisma (2015). Here texts are also arranged in chronological order by date, and the left-to-right arrangement of the columns reflects the increasing saliency of existential presupposition in each group of nominals: it is absent with generics (GNR), nominals with no presupposition of existence (NPE) (the new category introduced in this study), and existential nominals taking narrow scope (EXS-SCOPE-nrw). At the right side of the table we find nominals taking wide scope (EXS-SCOPE-wd) and nominals with a designation specified in the environment (EXS-SPC), where there is an obvious presupposition of existence. As

for the columns in between, those where the scope is ambiguous (EXS-SCOPE-amb) or there is no other logical operators (EXS), they have a less obvious status with respect to existential presupposition, and they constitute a sort of ‘gray area’ which has little use for the present study.

PERIOD	TEXT	GNR		NPE		EXS-SCOPE-nrw		EXS-SCOPE-amb		EXS		EXS-SCOPE-wd		EXS-SPC		AMB		TOT	
		BSG	AN	BSG	AN	BSG	AN	BSG	AN	BSG	AN	BSG	AN	BSG	AN	BSG	AN	BSG	AN
M1	cmpeterb	0	0	2	1	1	0	0	0	1	2	0	0	0	6	0	0	4	9
	cmhali	9	2	12	0	1	0	0	0	1	2	0	0	0	0	0	0	23	4
	cmjulia	2	0	0	1	1	1	0	0	1	11	0	0	0	0	0	0	4	13
	cmkathe	1	0	0	0	1	4	0	0	2	7	0	1	0	2	0	0	4	14
	cmmarga	1	0	2	0	3	3	0	0	1	8	0	0	0	2	0	0	7	13
	cmsawles	1	0	0	1	0	0	0	0	0	0	0	0	0	2	0	0	1	3
	cmancriv	54	2	10	8	12	3	0	0	11	5	0	0	0	2	0	1	87	21
M2	cmayenbi	1	2	0	2	0	1	0	0	0	1	0	0	0	1	0	1	1	8
M3(4)	cmctmeli	1	10	2	17	0	6	0	0	0	9	0	0	0	4	0	4	3	50
	cmctpars	5	30	1	33	0	4	0	2	0	8	0	0	0	3	0	1	6	81
	cmvices4	0	7	0	5	0	1	0	0	0	0	0	0	0	2	0	1	0	16
TOTAL		75	53	29	68	19	23	0	2	17	53	0	1	0	24	0	8	140	232

Table 1. The distribution of bare singular (BSG) nominals and nominals with  $a(n)$  (AN) in Middle English

For the M1 period, the situation is basically that of Stage Two in OE:  $an$  acts as an overt existential operator with indefinite nominals that are interpreted as specific (EXS-SPC: 0 BSG, 14 AN) and with nominals that take wide scope over some other operator (EXS-SCOPE-wd: 0 BSG, 1 AN). For nominals in the absence of other logical operators,  $a(n)$  is favoured by about 2 to 1 over BSG (EXS: 17 BSG, 35 AN); for NPE nominal, either generic or narrow scope existential, as well as for existential nominatives taking narrow scope, BSG is favoured by about 2 to 1 over  $a(n)$  (EXSSCOPE-nrw: 19 BSG, 11 AN; NPE: 26 BSG, 11 AN). On the basis of these frequencies, we conclude that the use of  $a(n)$  in the M1 period is basically what we expect if the system is a continuation of Stage Two of OE. But in addition to these expected patterns, we also see the first real change: in two texts, *Ancrene Riwle* and *Hali Meidhad*, there are two examples each of generics (GNR) used with  $a(n)$ , as in (12).

- (12) **a lutel hurt** in þe echʒe derueð mare þen amuchel  
*a little hurt in the eye harms more than a-great-one*  
 iþe hele  
*in-the heel*  
 (CMANCRIW-1,II.88.1068)

The totals for generics at M1 are 68 BSG, 4 AN. As stated above, in OE all 62 of the singular nominals used as generics were bare nouns.

In contrast to the M1 texts, the distribution of BSG and AN for generics (GNR) in the M3 texts has sharply reversed: there are only 6 BSG and 47 AN. For nominals with no presupposition of existence (NPE), there is also a reversal, with only 3 BSG and 55 AN. Similarly for existential nominals with narrow scope (EXS-SCOPE-nrw) and existential nominals in the absence of other logical operators (EXS), with all 11 and 17 tokens, respectively, using AN. It is clear that the use of *a(n)* with singular nouns has generalised to all contexts, with very few exceptions. This is also true of the one text that we have for the M2 period, *Ayenbite of Inwyrt*; however, we have only 9 tokens from this text, which makes the generalizations we draw for this period rather tenuous.<sup>10</sup> In addition, as mentioned in footnote 5, conclusions drawn from texts in the M2 period must be treated with caution, since the period is not well represented by the given texts. Our conclusion from these distributions must be that Stage Three, the PDE pattern, was clearly established by the time of the M3 period of ME, i.e. by 1350-1420, if not already at M2, and it has remained unchanged for more than 650 years.

At this point it is natural to ask whether the reanalysis from Stage Two to Stage Three correlates with the phonological reduction which occurred in the pronunciation of the indefinite article. The OE form *ān* (variously inflected for case and gender) split into two different forms, the numeral for 'one' and the indefinite article, the latter undergoing vowel shortening and loss of final *n*. For the texts included in this study, it is impossible to determine whether the vowel of *a(n)* was long or short, because vowel quantity is generally not marked; it might be suggestive (but certainly not conclusive) that in the texts from period M3, the numeral for 'one' is often spelled *oo(n)* while the indefinite article is never spelled *aa(n)*. As for the loss of final *-n*, Crisma (2009) shows that the modern

<sup>10</sup> Since we have collected only a sample of the data from *Ayenbite*, this problem can be easily rectified in the future.

pattern, with *a* used before consonants and *an* before vowels, is already established in the *Katherine Group*<sup>11</sup> and continues at later stages. Since the syntax of *a(n)* in the *Katherine Group* is arguably at Stage Two, this pattern might indicate that there is no correlation between (this aspect of) phonological reduction and the reanalysis from Stage Two to Stage Three.

It is also interesting to note that the numbers in Table 1, while not completely categorical, do not show nearly as much variation as the OE data presented in Crisma (2015). Most syntactic changes that have been studied quantitatively involve a great deal of intra-speaker variation, frequently analysed as grammatical competition; these changes are very gradual, taking hundreds of years to go to completion (see Kroch 1989, Pintzuk 1999, Pintzuk & Taylor 2006, Wallenberg 2009, among many others). In contrast, based on the numbers in Table 1, the change from Stage Two to Stage Three in the development of *an* happens very quickly, within a period of about 100 years. The change from Stage One to Stage Two in OE also occurs during a relatively short period of time, about 150 years between the late ninth century and the early eleventh century. Although some of the OE texts show quite regular Stage One or Stage Two grammars, there are also OE texts that exhibit the familiar variation between the two grammars. We do not observe this kind of variation in our study of ME. One possibility is that this is due to our choice of texts, and that it is simply an accident that we did not include in our database ME texts that show a great deal of variation with respect to the use of *an* in the various contexts. As stated above, we are presenting results from a descriptive pilot study, and much work remains to be done. But if the texts we used are indeed representative of periods M1 through M3 of ME, it is possible that the change from Stage Two to Stage Three does not involve grammatical competition, and that there is some characteristic of one stage or the other (or both) which prevents a long-term scenario of grammatical competition. We intend to investigate this possibility in future research.

#### **4. *Ayenbite of Inwyrt* and *The Book of Vices and Virtues***

As mentioned in section 2, *Ayenbite of Inwyrt* (1340; period M2) and *The Book of Vices and Virtues* (composition c1400, manuscript c1450; period M34) are two ME translations of the same Old French text, and their existence offers the rare opportunity of comparing the rendering of the same text in two different periods of ME. The figures in Table 1, however,

---

<sup>11</sup> But not in the other texts of the M1 period.

are quite disappointing from this point of view, for the two texts look very similar. In particular, both texts seem to have generalised the use of *an* with generics and NPE, the two contexts that reveal that *an* has reached Stage Three, i.e. it is the modern ‘indefinite article’. It is true that the single bare singular generic in *Ayenbite* is in fact 33% of the number of singular generic nominals used in the text, but the numbers are so low (three in total) that the frequency is meaningless.

The data on generics shown in Table 1 are, however, surprising in one respect: *Ayenbite* has three instances of singular generics and *Vices* has seven. Since they are both close translations, we might question the reason for this difference. A detailed examination of the texts reveals that in five of the seven instances, *Ayenbite* has a singular generic introduced by the definite article; this type of nominal was not included in the database for this study. The use of the definite article with singular generics may indicate that the *Ayenbite* translator avoided the use of *an* with generics, which in turn may indicate that the grammar is still conservative (Stage Two). This conclusion, however, is weakened by the observation that *Ayenbite* has the definite article where the French original also has it; in other words, *Ayenbite* is more of a word-by-word translation than *Vices*. The *Vices* translation is given in (13a) and the *Ayenbite* translation in (13b), with the Old French original in (13c).

- (13) a. riȝt as a **chirche** is halewed to þe seruise of God  
*right as a church is consecrated to the service of God*  
 (CMVICES4,105.173)
- b. Ase þe **cherche** is yhalȝed to godes seruise  
*as the church is consecrated to God.GEN service*  
 (CMAYENBI,106.2074)
- c. ausi comme li **moustiers** est dediez au service Dieu  
*so as the minster is dedicated at.the service God*  
 (SOMME LE ROI 145, p.215)
- ‘just as a / the church is dedicated to God’s service’

Table 1 does not include two types of data that show that the two texts under discussion are indeed very different, namely, the use of pronominal *man* and the use of generic noun phrases headed by *man*. Before

we present the data, some clarification is in order. There is robust evidence that *man*, or some weakened form thereof, could be a pronoun in OE and at least in the earliest stages of ME, analogous to modern German *man*. At some point in the history of English, pronominal *man* was lost, and various hypotheses have been put forth to describe and account for this loss (see in particular Raumolin-Brunberg & Kahlas-Tarkka 1997, Rissanen 1997, van Bergen 2000, Los 2002). The PPCME2 uses the part-of-speech (POS) tag MAN for the ‘indefinite pronoun MAN’: ‘If a given text clearly uses MAN (or also [the lexeme] ME in early texts) as a pronoun, then all unmodified uses of subject MAN are tagged MAN’.<sup>12</sup> Two examples of items tagged MAN in the PPCME2 are presented in (14):

- (14) a. þet hit wæs togeanes riht þet **man** scolde setten  
*that it was against right that man should set*  
 cleric ofer muneces  
*cleric over monks*  
 ‘that it was against right to set a cleric to rule monks’  
 (CMPETERB,43.66)
- b. lutel me is hwet **me** do mid mi bodi on eorðe.  
*little me.DAT is what man do with my body on earth*  
 ‘it matters little to me what one does with my body on earth’  
 (CMMARGA,65.168)

The POS tagging, however, is not entirely consistent, as is indeed expected for an element that ‘formally wavers between a noun and a pronoun’ (Los 2002:182). Thus, for example, *man* in (15) is not tagged MAN but N (common noun) in the PPCME2:

- (15) riȝt as a chirche is halewed to þe seruise of God  
*right as a church is consecrated to the service of God*  
 where **man** schal non oþer þing do but þe seruise of God  
*where man shall no other thing do but the service of God*  
 ‘just as a church is consecrated to the service of God, where one shall do nothing but the service of God’  
 (CMVICES4,105.173)

<sup>12</sup> Citation from the PPCME2 website: <http://www.ling.upenn.edu/hist-corpora/annotation/index.html>

The reason for the N tag in (15) may well be that this is an isolated case, for *Vices* seems to have abandoned pronominal *man* altogether, the degrammaticalisation<sup>13</sup> process described in Rissanen (1997) being completed.<sup>14</sup> But this example is very similar to (14a), and it might be a relic of a competing grammar that had grammaticalised pronominal *man*.

Apart from these inconsistencies in POS tagging, there is a more general problem regarding *man*: if at Stage Two a generic nominal headed by a singular count noun could occur bare, this possibility obviously extended to generic uses of the noun *man*,<sup>15</sup> which is then not easily distinguished from pronominal *man*. It is precisely to avoid these difficulties that all noun phrases headed by the lexical item *man*, whether POS-tagged as a noun or as pronominal *man*, were excluded from Table 1. For a preliminary investigation of noun phrases headed by unmodified *man* that seemed to receive an impersonal or arbitrary interpretation, as (15) above, we labelled them MANARB. Our label MANARB, different from MAN in the PPCME2, is also used when *man* is preceded by *a*, as in (16a-b), and is not restricted to subjects, as in (16b-c):

- (16) a. And by riches may **a man** gete hym grete freendes.  
*and by riches may a man get him great friends*  
 ‘and by riches a man may get himself great friends’  
 (CMCTMELI,232.C2.605)
- b. Pride dryueþ **a man** out of felawschip  
*pride drives a man out of fellowship*  
 (CMVICES4,101.77)

<sup>13</sup> As an anonymous reviewer points out, Rissanen (1997) uses the term ‘degrammaticalization’ but modifies it: “... an unlikely development which could be called degrammaticalization ...” (Rissanen 1997:517); “... the apparent degrammaticalization of man.” (Rissanen 1997:521). The reviewer suggests that this process can be better described as ‘retraction’ in the sense of Haspelmath 2004:33-5. We intend to investigate the history of *man* in early English in future research.

<sup>14</sup> Though, as Rissanen (1997:521) points out, even after the loss of pronominal *man*, after *some, any, every, each, man* was ‘probably pronominalized to some extent’.

<sup>15</sup> Not to mention those cases where it is impossible to decide whether *man* is the generic use of the noun *man* or is rather akin to PDE *man*, something similar to ‘mankind’.

- c. Prede: dep **man** out of uela3rede.<sup>16</sup>  
*pride does man out of fellowship*  
 (CMAYENBI,102.1997)

‘Pride drives one out of the community’

When *Ayenbite* and *Vices* are compared with respect to their use of MAN and MANARB, their grammars appear to be very different, as shown in Table 2, where the data for GNR from Table 1 are also repeated.

TEXT	GNR		MANARB		MAN	TOT	
	BSG	AN	BSG	AN		BSG*	AN
cmayenbi	1	2	8	0	38	47	2
cmvices4	0	7	1	29	0	1	36

Table 2. The distribution of GNR, MANARB and MAN in *Ayenbite* and *Vices*

\* pronominal *man* (MAN) is included in the total for bare singulars (BSG)

If the nominals labelled MANARB are indeed the generic use of *man*, Table 2 shows that the grammar of *Ayenbite* is still at Stage Two, with a clear predominance of bare generics, while that of *The Book of Vices* is at Stage Three, since singular generics (almost) always occur with *an*.

It is interesting to note that in *The Book of Vices*, the POS tag MAN, used for grammaticalised pronominal *man*, does not appear at all, which may indicate that *Vices* used *man* only as a noun;<sup>17</sup> it therefore seems reasonable to explore the hypothesis that this absence may correlate with the status of *an* as a Stage Three indefinite article. However, the GNR/MANARB/MAN data from all the texts we considered, presented in Table 3, do not seem to offer any evidence for a correlation between loss of pronominal *man* and the establishing of Stage Three *an*. It can be seen

<sup>16</sup> The French original of *Ayenbite* and *The Book of Vices*, given in (i), uses *homme* as the object and not *on*, which may explain why *Ayenbite* uses the full form *man* rather than the reduced *me*.

(i) Orgueil met homme hors de compaignie  
*pride puts man out of company*  
 ‘Pride drives one out of the community’  
 (SOMME LE ROI 72, p.210)

A systematic comparison between the original and the two translations, however, goes beyond the scope of the present work, and is left for future research.

<sup>17</sup> But see the discussion around (15).

that MANARB more or less patterns with GNR with respect to the use of *an*, though the M1 frequencies of MANARB are too low to draw firm conclusions for that stage. However, whether a given text uses pronominal *man* or not seems to be independent from *an* being at Stage Two or Stage Three. Notice the contrast between the Chaucer texts and *The Book of Vices*, both arguably at Stage Three as shown by the rate of *an* with generics: in Chaucer pronominal *man* tagged MAN is extremely common (37 tokens in *Melibee*, 70 in *The Parson's Tale*), while in the *Book of Vices* the POS tag MAN is never used.

PERIOD	TEXT	GNR		MANARB		MAN	TOT	
		BSG	AN	BSG	AN		BSG*	AN
M1	cmpeterb	0	0	0	0	22	22	0
	cmhali	9	2	1	0	6	16	2
	cmjulia	2	0	0	0	9	11	0
	cmkathe	1	0	0	0	22	23	0
	cmmarga	1	0	0	0	14	15	0
	cmsawles	1	0	0	0	6	7	0
	cmancriw	54	2	4	2	84	142	4
M2	cmayenbi	1	2	8	0	38	47	2
M3(4)	cmctmeli	1	10	0	25	37	38	35
	cmctpars	5	30	3	111	70	78	141
	cmvices4	0	7	1	29	0	1	36
TOTAL		75	53	17	167	308	117	212

Table 3. The distribution of GNR, MANARB and MAN in Middle English texts

\* pronominal *man* (MAN) is included in the total for bare singulars (BSG)

### 5. What is 'a man'?

The data, however, are even more complex than they appear. Consider first that a strong piece of evidence for analysing impersonal *man* as a pronoun is that we have found two instances of co-referential *man* in a C-command relation. Since this configuration does not give rise to a Principle C violation, the two instances of *man* are pronouns and not

nouns for the purposes of Binding. The examples in (17a-c), adapted from Cabredo-Hofherr (2008:5) are from modern German, PDE and modern French, respectively:<sup>18</sup>

- (17) a. **Man**<sub>i</sub> erkältet sich wenn **man**<sub>i/k</sub> nicht aufpasst.  
*MAN cools himself if man NEG pay-attention*  
 ‘One catches a cold if one does not pay attention’
- b. **One**<sub>i</sub> catches a cold (easily) when **one**<sub>i/k</sub> isn’t careful.
- c. **On**<sub>i</sub> attrape un rhume si **on**<sub>i/k</sub> ne fait pas attention.  
*ON catches a cold if ON NEG do NEG attention*  
 ‘One catches a cold if one does not pay attention’

Thus it is no surprise to find pronominal *man* in this same configuration in ME:

- (18) Wel gratter þing his huanne **me** is zuo yuestned ine  
*well greater thing is when man is so established in*  
 þe loue and adrayngt in þe zuetnesse of god. þet no  
*the love and drunk in the sweetness of God that no*  
 solas ne no confort **me** ne onderuangþ: bote of him.  
*joy NOR no comfort man NEG take but of him*  
 ‘It is a much greater thing when one is so firm in the love of God and  
 inebriated by his sweetness that he does not take any joy or comfort  
 but from him’  
 (CMAYENBI,107.2082)

In the example in (18), the two instances of co-referent pronominal *man* are the reduced form *me*, tagged as MAN in the PPCME2. They correspond to *on* in the original French text:

- (19) plus est grant chose quant **on** est si afermez en  
*more is grand thing when ON is so established in*  
 l’amour et abuvrez de la douceur Dieu que nul solaz  
*the-love and drunk of the sweetness God that no joy*

<sup>18</sup> French *on* in (17c) is a reduction of *hom* ‘man’.

ne nul confort **on** ne reçoit se de lui non,  
*NEG no confort ON NEG receives if from him NEG*  
 ‘It is a greater thing when one is so firm in the love of God and inebriated by his sweetness that he does not receive any joy or comfort if not from him’  
 (SOMME LE ROI 158, p.216)

What is really surprising, however, is that the corresponding passage in *Vices*, given in (20a), has two instances of *a man* in the same configuration, and they are co-referential. This example in *Vices* is not isolated, because there are two analogous cases in the Chaucer texts, given in (20b-c):

- (20) a. and wel gretter þing is it whan **a man** is so affermed  
*and well greater thing is it when a man is so established*  
 in þe loue and so dronke in þe swetnesse of God þat no  
*in the love and so drunk in the sweetness of God that no*  
 solas ne counfort sauoureþ **a man** nouȝt but on liche of hym  
*joy nor comfort savours a man NEG but un like of him*  
 (CMVICES4,106.181)
- b. and that is whan **a man** ne douteth no travaille in tyme  
*and that is when a man NEG fear no toil in time*  
 comynge of the goode werkes that **a man** hath bigonne.  
*coming of the good works that a man has begun*  
 (CMCTPARS,313.C1.1061)  
 ‘and this is when one does not fear the future toil for the good works that he has begun’
- c. therefore sholde **a man** fle and eschue werre in as  
*therefore should a man flee and avoid war in as*  
 muchel as **a man** may goodly  
 much as a man may possibly  
 ‘therefore one should flee and avoid war as much as one may at all do so’  
 (CMCTMELI,235.C1.696)

The existence of these examples indicates that not only *man* (or some reduced form thereof) but also *a man* could be re-analysed (or grammaticalised) as an impersonal pronoun. Thus, while Table 3 may give the impression that *Vices* has abandoned pronominal *man*, this is in fact not true; it still has pronominal *man*, albeit in a totally unexpected form: *a man*.

It is very difficult to determine how many of the 29 instances of *a man* in *Vices* and the 136 in Chaucer (25 in *Tale of Melibee*, 111 in *The Parson's Tale*), are pronominal. Of course, there are also cases in which two instances of *a man* receive disjoint reference:

- (21) Eek if **a man**, by caas or aventure, shete an arwe, Or  
*also if a man by chance or accident shoots an arrow or*  
 caste a stoon, with which he sleeth **a man**, he is homicide  
*casts a stone with which he kills a man he is murderer*  
 ‘Also, if someone, by chance or accident, shoots an arrow or casts  
 a stone with which he kills someone else, he is a murderer’  
 (CMCTPARS,306.C2.762)

However, this is not an argument to consider them non-pronominal, as shown by the following French example, where the two uses of *on* receive disjoint reference:

- (22) **On**<sub>i</sub> dit maintenant qu’ **on**<sub>k</sub> doit manger 5 légumes par jour.  
*ON say now that ON must eat 5 vegetables per day*  
 ‘They say now that one must eat 5 (servings of) vegetables per day’  
 (Cabredo-Hofherr 2008:7)

In sum, examples such as those in (20) add a new piece to the puzzle of the ‘degrammaticalisation’ of *man*, showing that not only the simple form *man* and the reduced form *me* but also the complex form *a man* could be instances of pronominal *man*. Whether this complex form of pronominal *man* directly contributed to its eventual loss is a topic that we leave for future research, but it is clear that this form cannot be omitted from any account of the loss of pronominal *man* in English.

### 5. Conclusions

This pilot study presents two novel empirical findings: first, we show with quantitative evidence that the modern ‘indefinite article’ is clearly established by the end of the fourteenth century, with a puzzling lack of variation; second, we present original syntactic evidence that the complex form *a man* could also be a grammaticalised impersonal pronoun in ME.

These findings open the way for new lines of investigation. First of all, it would be desirable to understand the reasons for the lack of ME texts displaying real variation in the use of *a(n)*; this question can be addressed by enlarging the empirical base to gain a more complete picture of the diachronic development that led to the final establishment of the ‘indefinite article’ in the modern sense. Then, there is the intriguing problem of the morphosyntactic analysis of pronominal *a man* and its role in the eventual demise of pronominal *man* in English. These topics are left for future research.

### References

- Ackles, Nancy. 1997. *Historical Syntax of the English Articles in Relation to the Count/Non-count Distinction*. Ph.D. dissertation, University of Washington.
- Bergen, Linda D. van. 2000. The Indefinite Pronoun *man*: ‘Nominal’ or ‘Pronominal’? In Ricardo Bermudez-Otero, David Denison, Richard M. Hogg & Chris B. McCully (eds.), *Generative theory and Corpus Studies: A Dialogue from 10 ICEHL*, 103-122. Berlin & New York: Mouton de Gruyter.
- Borer, Hagit. 2005. *In Name Only*. Oxford: Oxford University Press.
- Brugger, Gerhard. 1993. Generic interpretations and expletive articles. *University of Venice Working Papers in Linguistics*, 3/1:1-30.
- Cabredo-Hofherr, Patricia. 2008. Les pronoms impersonnels humains - syntaxe et interpretation. *Modèles Linguistiques* tome XXIX-1, vol 57, 35-56.
- Carlson, Greg N. 1977. A unified analysis of the English bare plural. *Linguistics and Philosophy* 1, 413-456.

- Crisma, Paola. 2009. Word-initial *h-* in Middle and Early Modern English. In Donka Minkova (ed.) *Phonological Weakness in English: From Old to Present-Day English (Palgrave Studies in Language History and Language Change)*, 130-167. Houndmills, Basingstoke, Hampshire and New York, NY: Palgrave Macmillan.
- Crisma, Paola. 2011. The emergence of the definite article in English: a contact-induced change? In Petra Sleeman & Harry Perridon (eds.) *The Noun Phrase in Romance and Germanic. Structure, Variation, and Change*, 175-192. Amsterdam/Philadelphia: Benjamins.
- Crisma, Paola. 2015. The 'indefinite article' from cardinal to operator to expletive. In Chiara Gianollo, Agnes Jäger & Doris Penka (eds.), *Language Change at the Syntax-Semantics Interface*, 125-151. Berlin: De Gruyter Mouton.
- Haspelmath, Martin. 2004. On directionality in language change with particular reference to grammaticalization. In Olga Fischer, Muriel Norde & Harry Perridon (eds.), *Up and Down the Cline: The Nature of Grammaticalization (Typological Studies in Language 59)*, 17-44. Amsterdam: John Benjamins.
- Ingham, Richard. 2013. Negation in the history of English. In David Willis, Christopher Lucas, and Anne Breitbarth (eds.) *The History of Negation in the Languages of Europe and the Mediterranean*, 119-150. Oxford: Oxford University Press.
- Kroch, Anthony & Ann Taylor. 2000. Penn-Helsinki Parsed Corpus of Middle English, second edition. <http://www.ling.upenn.edu/hist-corpora/PPCME2-RELEASE-3/index.html>.
- Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1, 199-244.
- Los, Bettelou. 2002. The loss of the indefinite pronoun *man*. Syntactic change and information structure. In Teresa Fanego, Javier Pérez-Guerra & María José López-Couso (eds.), *English Historical Syntax and Morphology: Selected papers from 11 ICEHL, Santiago de Compostela, 7-11 September 2000*. Volume 1 [Current Issues in Linguistic Theory 223], 181-202. Amsterdam/Philadelphia: Benjamins.
- Pintzuk, Susan. 1999. *Phrase Structure in Competition: Variation and Change in Old English Word Order*. New York: Garland.
- Pintzuk, Susan & Ann Taylor. 2006. The loss of OV order in the history of English. In Ans van Kemenade & Bettelou Los (eds.), *Blackwell Handbook of the History of English*, 249-278. Oxford: Blackwell Publishing.
- Raumolin-Brunberg, Helen & Leena Kahlas-Tarkka. 1997. Indefinite pronouns with singular human reference. In Matti Rissanen, Merja Kytö & Kirsi Heikkonen (eds.), *Grammaticalization at work: Studies of long-term developments in English*, 17-85. Berlin: Mouton de Gruyter.

- Rissanen, Matti. 1997. Whatever Happened to the Middle English Indefinite Pronouns? In Jacek Fisiak (ed.), *Studies in Middle English Linguistics*, 513-529. Berlin & New York: Mouton de Gruyter.
- Wallenberg, Joel. 2009. *Antisymmetry and the Conservation of C-Command: Scrambling and Phrase Structure in Synchronic and Diachronic Perspective*. Ph.D. dissertation, University of Pennsylvania.
- Zamparelli, Roberto. 2000. *Layers in the Determiner Phrase*. Revised Ph.D. dissertation, University of Rochester.