Some as an indefinite article in Present Day English: a case of paradigmatization and constructional competition

Lotte Sommerer (University of Freiburg)

Florent Perek (University of Birmingham)

Abstract

This paper investigates the different functions of *some* in Present Day English. It especially focuses on whether and to what extent *some* functions as an indefinite article for non-count and plural nouns and as such competes with the bare marking strategy (*I need to buy some milk/some cigarettes vs. I need to buy milk/cigarettes*). It is shown that next to being a quantifier, *some* has grammaticalized into an article-like element. On the basis of a sample of direct object NPs extracted from the British National Corpus (XML Edition), we analyze singular, plural and non-count NPs functioning as syntactic objects which either occur bare or have *some* as a determiner. One question is how often and in which constructional environments *some* functions as a so-called 'near-article' and which other functions it fulfills (e.g. partitive or vagueness marker). The results of a multinomial logistic regression model are used to discuss which grammatical factors are correlated with the use of *some* as an existential marker of indefiniteness. We then investigate the potential factors predicting speakers' choice between using *some* as an overt near-article or leaving the nominal bare in indefinite NPs (i.e. zero article), by means of binomial logistic regression. Theoretically, this paper contributes to the study of such mechanisms as paradigmatization, analogization and constructional competition within a usage-based, constructional model of language and its change.

"*Some* is a complicated little word" (Israel 2000: 169)

1 Introduction

This paper investigates the different functions of *some* in Present Day English (PDE). It especially focuses on whether and how often *some* functions as an indefinite article. It is shown that next to being a (partitive) quantifier, *some* has developed an additional, article-like function especially for plural and non-count nouns. It will be discussed a) to what extent this is indeed the case and b) if this function is in competition with zero marking due to 'paradigm pressure' and 'paradigmatization' caused by a shifting coding strategy in English towards marking (in)definiteness overtly and obligatorily (Sommerer 2018; Sommerer & Hofmann 2021).

It is well-known that in Modern English (in)definiteness has to be marked overtly and obligatorily with singular count nouns, and the definite article *the* and the indefinite article *a/an* function as the default markers to do so.

	DEFINITE		INDEFINITE	
	Count	**Non-count**	**Count**	**Non-count**
SINGULAR	the book	the ink	a book	Ø / some ink
PLURAL	the books		Ø / some books	

Table 1. Use of the definite and indefinite article in Present Day English (adapted from Quirk et al. 1985: 265)

(In)definiteness is a complex composite notion which has to do with referentiality and identifiability via uniqueness, familiarity and specificity (i.e. locatability in a shared set) (e.g. Christophersen 1939; Hawkins 1991, Chestermann 1991, Lyons 1999). According to Langacker (1991: 54), articles are discourse-pragmatic elements which ground the nominal in discourse. In English, nominals typically need to be grounded by a determiner to fully function in a referential NP and these determiners express the way in which a "conceptualizer makes mental contact with an instance of a nominal type" (Israel 2000: 171). At the same time, it is also well-known that English plural and non-count nouns do not follow this trend of overt marking in the sense that they can occur 'bare' in indefinite NPs (ex.1a and 1b).

1) a) *I see [**ducks**]NP on the lake.*

   b) *I need to buy [**milk**]NP.*

Although leaving the determination slot unfilled in these examples is perfectly acceptable, various researchers (e.g. Sahlin 1979; Chesterman 1993; Israel 2000; Jacobsson 2002) have suggested that *some* has become a competitor to the so-called 'zero-marking' (Ø) strategy for these contexts (2a & b):

2) a) *I see [**some ducks**]NP on the lake.*

   b) *I need to buy [**some milk**]NP*

Recruiting *some* as an overt 'near-article'[1], in a way completes the overt coding strategy and the article paradigm presented in Table 1. In other words, *some* – as a function word – is an example of a construction which is currently undergoing secondary grammaticalization (Narrog & Heine 2021: 8) because speakers of English have started to recruit it as a slot filler to fill a gap in the existing article paradigm.

If one assumes that English is currently undergoing a diachronic development towards paradigm completion, this raises the following research questions:

i.   RQ1: How often is *some* used as a quantifier or with other functions in Present Day English? How often is it used as a near-article to exclusively mark indefiniteness?

ii. RQ2: Which factors/variables increase the likelihood for *some* to be used as a bleached marker of indefiniteness?

---

[1] The terms 'near-article' or 'article-like' account for the fact that if one defines an article as a grammatical element which primarily and obligatorily marks (in)definiteness, then *some* does not deserve full article status yet. It clearly is not the default marker like *a/an* and it still competes with zero-marking.

iii.   RQ3: In its article-like function, is it a competitor to the bare option?

We will try to answer these questions by conducting a qualitative and quantitative empirical corpus study. To our knowledge, no study so far has analyzed *some* quantitatively and in a corpus-based manner, a research gap which we aim to fill. For this paper, we analyze samples of singular, plural and non-count NPs functioning as syntactic objects which either occur bare or have *some* as a determiner. The results of a multinomial logistic regression model (section 4.2) are used to discuss in what way independent variables (e.g. number of N or additional modification) increase the likelihood for *some* to be used as an existential marker of indefiniteness. On top of that, we investigate the covert or overt marking of indefiniteness in plural and non-count NPs by investigating the competition between using *some* or by leaving the nominal bare (i.e. zero-marking (Ø)) (binominal regression model in section 4.3).

Theoretically, this paper subscribes to a usage-based, constructional model of language and its change (e.g. Barðdal et al. 2015; Diessel 2019; Smirnova & Sommerer 2020; Hilpert 2021) and discusses mechanisms like analogization (Fischer 2007; Traugott & Trousdale 2013), obligatorification/ paradigmatization (Lehmann 1995[1982]) and constructional competition (Zehentner 2019). Ultimately, the investigated phenomenon is used in order to discuss to what extent observable synchronic variation in a language follows expected trajectories of language change (see Lorenz & Tizón-Couto this volume).

The paper is structured as followed: First we provide some theoretical background discussing the functional versatility of *some* (section 2). Then we present our methodology, i.e. the corpus data and the classification system of our variables and assumed functions (section 3). Section 4 presents the regressions models and their findings. The results are interpreted in a separate section (5) in which we also relate our findings to the overarching theme of this special edition. Section (6) concludes the paper with some open questions and desiderata for future research.

2   Functions of *some* in Present Day English

In this section, we discuss the various functions of *some* as they are described in the literature. In general, *some* occurs in two main syntactic functions. On the one hand, it can occur independently, as a pronoun heading its own NP (e.g. *If there is pizza left, I'd like to get <u>some</u>*). On the other hand, it can be used dependently as a pre-head element (e.g. <u>*Some people like*</u> *pineapple on their pizza).* In this paper we are exclusively concerned with *some* as a pre-head dependent in the left periphery of the head noun. In this position, *some* can trigger a variety of semantic interpretations.

In general, the functions of *some* can be understood better if one keeps its paradigmatic relations in mind. *Some* is a "relative quantifier", which can be paradigmatically positioned on a cline from *no > some > most > all* (Langacker 2016: 6). On a scale, *some* expresses a moderate quantity smaller than *most* but larger than zero (with plurals at least two). In logic-based analyses, *some* is treated as an "existential quantifier" ∃ (Partee, ter Meulen & Wall 1990: 138; Farkas 2002) and contrasts with universal quantifiers ∀ exemplified in English by *all*, *each*,

every. Downing (2015: 384) lists *some* as a "non-exact quantifier" and Langacker (2016: 8) remarks that *some* "is quite vague about quantity".

Additionally, *some* is a positive polarity item (PPI) and rarely occurs in the scope of negation. As an assertive element, *some* is found in opposition to non-assertive *any* (Israel 2000: 170).[2]

3) a) *I see **some cars** on the road.*

    b) *I don't see **any cars** on the road.*

Some can express a meaning slightly different from the indefinite singular article *a(n):*

4) a) ***Some student** is waiting in front of your office.*

    b) ***A student** is waiting in front of your office.*

Finally, another opposition has to be mentioned: *some* vs. zero marking.

5) a) *I see **some cars** on the road.*

    b) *I see **Ø cars** on the road.*

It is this opposition in plural and non-count contexts which we are especially interested in. It will be discussed below in more detail (also see section 5).

In general, the literature distinguishes two main functions of *some*: partitive vs. non-partitive construal. Often *some* invites a partitive reading where the referent is "construed as part of a larger group familiar in the discourse" (Israel 2000: 172):

6) ***Some people** like pineapple on their pizza.*

7) ***Some candidates** misunderstood the question.*

In these examples, the main point is that only some but NOT ALL people like pineapple on their pizza. Some BUT NOT THE OTHER candidates misunderstood the issue. This reading has been called "basic proportional" (Payne & Huddleston 2002: 381; Langacker 2016: 6), "partitive" (Israel 2000: 173) or "selective" (Halliday & Matthiessen 2014: 366-369). Instances of this usage can be identified by reformulating the string as a partitive construction, e.g. *some of the people*, *some of the candidates*. Sahlin calls this usage "an indefinite assertive limiting quantifier" (1979: 14-16). In this function *some* is often pronounced as /sʌm/ with an unreduced vowel quality and it has been argued that it is primarily this stressed intonation which mainly codes the partitive meaning (Sahlin 1979). For this paper, what we consider to be relevant is that this function is clearly quantificational: it quantifies a proportion of a larger set.

In many examples, *some* must be interpreted in a non-partitive way. In examples 8) – 11) the NOT ALL implicature is clearly missing. This usage has been termed "non-proportional" or "non-selective":

8) *They try to use fewer electronic devices, so that they can save **some money**.*

---

[2] We will not discuss non-assertive *any* but note that this topic has been investigated extensively (see e.g. Bolinger 1977; Hirtle 1988; Israel 2000; Jacobsson 2002).

9)   *At the interview they asked **some questions***.

10)   *I need to buy **some cigarettes***.

11)   *We need **some milk**. Could you get it at the supermarket?*

In these examples *some* primarily introduces a new specific but unidentified discourse referent establishing the existence of N rather than quantifying N (Duffley & Larrivée 2012: 135). In this function, *some* is unstressed /səm/. These instances of *some* have been classified as indefinite 'light quantitative article' or 'near-article' selecting plural and non-count heads (Sahlin 1979; Chesterman 1991; Israel 2000).

Although we basically support the presented analysis, the problem with this two-fold categorization (partitive vs. non-partitive/article-like) is the following: Although it is true that none of the examples above are compatible with a partitive interpretation, they still show gradience regarding their expression of quantification. In examples 8) and 9), *some* feels less bleached than in 10) and 11), which is why it is debatable whether in the first two examples one should speak of 'article-like' usage. One could call this a type of "weaker" non-partitive quantification (e.g. Payne & Huddelston 2002: 358; Duffley & Larrivée 2012). In 8) *some* obviously entails the existence of money, but the context suggests that the speaker wants to save a certain (small) amount of money. Similarly, the main point of 9) seems to be that there was a certain number of questions. In contrast, *some* seems to have lost its quantificational semantics to a noticeably larger extent in 10) and 11).

This observable ambiguity relates to the following fact: an indefinite NP can often designate either a quantity OR a kind of a nominal type. For example, the sentence *There are some letters for you on the table* allows for two readings. The first option is that the speaker wants to stress that only a few letters and not a large pile of them are to be found on the table, whereas the example could also be read as a mere statement about the existence of letters (i.e. the kind reading; letters being present in contrast to not being present). Depending on the interpretation, one example is either 'weakly quantificational' or 'purely existential'.

A related question then is whether leaving out *some* would trigger the same semantic interpretation:

12)   a)   *I went to the post office to buy **some stamps***.

      b)   *I went to the post office to buy **Ø stamps***.

13)   a)   *We have received **some news** from Moscow this morning*

      b)   *We have received **Ø news** from Moscow this morning*.

From a traditional, functionalist perspective, which expects a difference in form to be indicative of a difference in meaning, one could argue that there is a subtle meaning difference between a) and b).[3] The speakers might have added *some* to the examples because they wanted to trigger the "quantity" reading rather than the existential "kind" reading which is mainly expressed by the bare alternative. At the same time, the examples above are clearly compatible with the non-

---

[3] See Quirk et al (1985: 274f.) or Chestermann (1991) for a detailed discussion of differences.

quantitative, existential kind reading as well. In this debate, we position our own research in the following way: we argue that in cases where *some* does not invite a quantity construal, it can be considered a near-article and is in competition with the zero option. Using the bleached version of *some* is often synonymous with leaving the head noun bare (e.g. 10 & 11). Although it is difficult to identify and distinguish the two interpretations, we have attempted to do so in our empirical corpus study (see categorization in section 3).

Another important function of *some* has been termed "vagueness marker" or "marker of non-specificity" (Payne & Huddleston 2002: 380). *Some* – in contrast to *a/an* – conveys the idea that the identity/specificity of the head noun is of little importance, cannot be recalled or is even completely unknown to the speaker: "For whatever reason the speaker either cannot or will not specify the identity of the referent" (Israel 2000: 175). In 14) the hope is to win in the lottery *some day* in the future on a non-specified date. In 15) the speaker's main message is that one (non-specified) way or any other way is acceptable. In 16) the people showing up in front of the office are specific but still unknown to the speaker.

14)   ***Some day*** *I will win the lottery.*

15)   *Find **some way** to destroy them.*

16)   ***Some idiots*** *showed up in front of my office.*

This 'vagueness' interpretation mainly emerges in combination with singular nouns but we also find it with plural nouns (as in 16).

Several additional functions of *some* are discussed in the literature, for instance a so-called approximator function, an augmentation/diminution function and an exclamative function (Warfel 1972; Israel 2000; Payne & Huddleston 2002; Duffley & Larrivée 2012). Many of these functions are more pragmatic, can be assigned on top of the basic functions, and heavily depend on the co-text. For our empirical investigation, we have decided to ignore these pragmatic functions because any interpretation is extremely context dependent and, in that sense, cannot really be considered an integral part of *some*'s semantic profile. Moreover, one of our unpublished pilot studies revealed that agreement between annotators is extremely low when assigning pragmatic functions, which shows that these meanings cannot be detected in an objective manner. This means that only the functions discussed above will form the basis of our classification scheme (see section 3).

3 Methodology

3.1 Data collection

The empirical investigation in this study is based on data from the British National Corpus (XML Edition). Examples of relevant noun phrases were identified in a syntactically annotated version of the corpus. The whole corpus was first automatically annotated for syntactic dependencies by means of the Stanford Neural Network Dependency Parser (Chen & Manning 2014). We then looked for all instances of the direct object dependency (DO); this resulted in a list of 3.4 million potential noun phrases used as the direct object of a verb in the corpus. We chose to restrict the dataset to only focus on direct objects because we wanted to keep the

syntactic position of our tokens constant to avoid introducing an additional factor that could potentially have a significant impact on NP determination. As is well-known, some sentence positions are associated with particular discourse properties, notably in terms of information structure, which may make it more likely for certain NP forms to occur in them, and which in turn may constrain the choice of determiner. For instance, subjects tend to correspond to discourse-given information, and for this reason they tend to be pronouns or definite noun phrases, and conversely, indefinite subjects are rather uncommon (cf. Du Bois 1987, and more generally the literature on Preferred Argument Structure for more observations of this kind). Restricting our dataset to a single sentence position eliminates this potential confounds on NP determination. We chose the direct object position because the dependency annotations make it easy to identify as a syntactically coherent category, and it is a priori not biased against indefinite NPs in the same way that some other grammatical relations (such as subject) are. We are aware, however, that this is a potential limitation of our present findings, and we do recognize the need for potential future studies expanding the scope to other sentence positions.

We used the dependency annotation to restrict the dataset to only relations between a verb and a common noun (i.e. the head of the direct object NP) and to automatically gather information about the structure of each noun phrase, in particular the determiners that they contain. As a result of this procedure, we were able to extract 20,915 NPs with *some* as their sole determiner. Of these, 12,753 are tagged as (NN), including count or non-count nouns, and 8,162 combine with a plural noun (NNS). The automatic extraction procedure also allows us to identify more than 700,000 NPs with a zero determiner, which correspond to cases in which no determiner was found in a modification relation with the head noun. These tokens will be used in the second part of our study (section 4.3).

3.2   Data annotation

Our research questions involve the semantics and function of the pre-head element *some*. Since these semantic features cannot be automatically retrieved from the corpus data, they require manual annotation. Annotating all examples manually would have been too time-consuming and impractical, hence we chose to work with small samples of randomly selected tokens. To answer RQ1 and RQ2, we created three samples of 350 examples of *some* each, corresponding to the three grammatical types of nouns combined with it, i.e. plural nouns (e.g. *crutches*), non-count nouns (e.g. *tea*), and singular count nouns (e.g. *tale*):

Sample 1: 350 ex. (*some* + N_{pl}), e.g. *They refused to lend him* ***some crutches***.

Sample 2: 350 ex. (*some* + $N_{non\text{-}count}$), e.g. *She brought him* ***some tea***.

Sample 3: 350 ex. (*some* + N_{sg})[4], e.g. *He told* ***some tale about his fathe*r**.

Note that the samples also include NPs which are modified by a pre-head adjective or include post-head modification or complementation. The 1050 examples were coded for the following variables: noun and verb lemma, grammatical type of head noun (singular count noun vs. non-count noun vs. plural noun), semantic type of head noun (concrete vs. abstract vs. measure

---

[4] In the process we excluded certain examples, like *some sort of X*, *some kind of X* because is these cases the (head) noun status of *sort* is often disputed.

noun), the presence/absence of modifiers before the head (pre-modification) or after the head (post-modification),[5] and the function of the determiner *some*, as detailed below. These variables were chosen as we hypothesize that they may have an influence on the function of *some*. For instance, we expect that quantificational use correlates with measure nouns. We also wanted to check if modification (e.g. by adjectives), which tends to have a specifying and set-shrinking effect with intrinsically partitive semantics, interacts with quantification and determination.[6]

In our analysis, we distinguish four main functions of *some*: partitive quantifier, weak quantifier, vagueness marker and near-article. To identify these functions more easily and to make our annotation less subjective, we use two kinds of tests on each token: a synonymous substitution test and a contrastive paradigmatic substitution test. The synonymous substitution tests consist in replacing *some* with another item with the same or a similar meaning. If the replacement is grammatical and does not cause a detectable change in meaning from the original sentence, this is taken to mean that *some* has the corresponding function. Contrastive paradigmatic substitution tests consist in replacing *some* with another item in the same functional paradigm. Each function of *some* can indeed be seen as a member of a larger set of determiners that code related meanings; for instance, the non-partitive quantifier use of *some* forms a paradigm with other non-partitive quantifiers, such as *several*, *many*, *a lot of*, etc. If the replacement of *some* by one of these items is grammatical, this is taken to mean that this use of *some* is part of the relevant paradigm.

The four main functions of *some* are described in turn below, with an explanation of the relevant tests and a few examples.

3.2.1 Partitive quantifier

The partitive quantifier use of *some* corresponds to NPs which refer to a small number of entities which are part of a larger set, or a small amount of something which is considered with respect to a full amount. This use of *some* is equivalent to an explicit partitive structure, i.e. *some of the N*.

17) *The crush was eased by moving **some fans** to an empty, condemned section of the ground.* (A9R, W:newsp: brdsht_nat:sports)

18) *Next, I set about finishing off **some chapters of a rheumatology book for GPs that my Bristol colleagues and I have been writing***. (BNK, W: misc)

19) *Explain how and why it improves **some files but not others***. (FPG, W:ac:tech_engin)

The larger set or full amount is usually implied by the sentence context, but it can also be directly referred to in a post-modifier, such as in 18) above. The existence of other entities in the set that are not involved in the predication can also be explicitly entailed by the use of

---

[5] For the sake of simplicity, we did not distinguish optional post-modification from obligatory complementation; both were lumped under the heading of post-modification.
[6] Additionally, a diachronic paper by Sommerer and Hofmann (2021) argues that in late Middle English and Early Modern English, it is modified NPs in which *some* starts to develop an article-like meaning first. These modified examples are interpreted as bridging contexts which invite grammaticalization and semantic bleaching of the quantifier *some*. This is another reason to test a potential correlation in Present Day English.

phrases such as "but not others", as in 19). This use of *some* is paradigmatically related to other determiners that also identify a certain number or amount within a larger set, e.g. *all*, *no*, etc.

Synonymous substitution test: *some* can be replaced with a partitive construction with the same meaning, e.g. *some of the*, *some but not all*, *some not others*, without changing the overall interpretation of the sentence. It can be redundantly preceded by *only* (*only some*) with no change in meaning.

Contrastive paradigmatic substitution test: *some* can be replaced with other partitive quantifiers with a different meaning, with or without an explicit partitive structure with *of*, e.g. *all (of the)*, *many/much (of the)*, *few/little (of the)*, *most (of the)*, *none of the*, *every*, *no*, etc.

3.2.2 Non-partitive quantifier

The non-partitive quantifier use also refers to a small number or small amount, but there is no larger set which the small number or amount is defined relative to.

20) *This afternoon we went to see a new workers' housing estate and visited **some homes***. (KAL, W:letters:personal)

21) *Finally, let us consider **some motion verbs that have built-in deictic components***. (J2K, W:ac:soc_science)

22) *Even though it may cause them **some embarrassment** it is in their interests that they commit themselves fully to telling the public the truth.* (HJ4, W:newsp:other:report)

Synonymous substitution test: *some* can be replaced with another quantifier with a similar meaning of 'small quantity', e.g. *a few/little*, *a (little) bit of*, *a small number/ amount of*, *a handful of*, etc. The replacement does not affect the overall interpretation of the sentence; in particular, the substitute quantifier does not contribute a meaning of quantification that was not in the original sentence, which would indicate an article use of *some* (see below).

Contrastive paradigmatic substitution test: *some* can be replaced with other non-partitive quantifiers, e.g. *many*, *much*, *several*, *a lot of*, etc.

3.2.3 Near-article/article-like function

This function covers cases in which some does not invite a quantifier reading, in that the context is one in which the number or amount referred to is irrelevant and/or does not necessarily have to be small. In other words, the speaker only tries to state the existence of the noun phrase referent, but not how much of it there is. Thus, *some* merely functions as an overt marker of indefiniteness and nothing more. Some examples are provided below.

23) *Last monsoon when I needed **some musk to rub on a snake-bite**, I didn't have to look far.* (BNU, W:misc)

24) *He said he wanted **some company** because he was lonely and worried.* (G15, W:fict:prose)

25) *I was gonna buy **some pickled onions** today, till I saw the price!* (KST, spoken_conversation)

This use of *some* is semantically and grammatically equivalent to the zero-article, i.e. the covert coding of indefiniteness through bare uses of plural nouns and non-count nouns without a determiner. As a pure marker of indefiniteness, it is involved in a paradigmatic relation with other definiteness markers, such as the indefinite article *a/an*.

Synonymous substitution test: *some* can be deleted from the sentence (i.e. replaced by the zero-article) without loss in grammaticality and without any noticeable difference in meaning. Conversely, replacement by a quantifier (either partitive or non-partitive), such as *a bit of* or *a few/little*, alters the meaning of the sentence by adding a semantic feature (namely 'small amount') that could not be read from the original sentence.

Contrastive paradigmatic substitution test: *some* can be replaced by other non-quantifying determiners that also function purely as definiteness markers, e.g. *a/an* or zero-article.

3.2.4 Vagueness marker

The vagueness marker use of *some* is similar to the near-article use in that it also does not convey a quantifier meaning. However, it differs from it in that it does not provide a reference to a specific entity. The identity of the referent could be one of many but it is kept vague, and in many cases irrelevant or even unknown to the speaker. This semantic feature justifies that this use of *some* should be kept apart from other types of determiner uses.

26)   *I had to think up **some justification of my silliness**.* (FEE, W:fict:prose)

27)   *The woman and the android were walking through the dark corridors of the castle, trying to find **some means of escape**.* (HTY, W:fict:prose)

28)   *Adam had at the time concocted **some tale** to keep his father quiet, to keep him away even.* (CDB, W:fict:prose)

Synonymous substitution test: *some* can be replaced with items with a similar meaning of vagueness like *any*, *some kind of* or *a certain*.

Contrastive paradigmatic substitution test: Not applicable.

3.2.5 Other cases

As with most kinds of semantic annotation, it is very difficult in practice to reliably and objectively distinguish the main functions of *some* in corpus examples. While the tests mentioned above do improve the quality of our annotation, we are still left with ambiguous cases, i.e. tokens that pass both the meaning test and the distribution test of more than one category. This happens particularly often with the near-article and non-partitive quantifier categories. To distinguish these tokens from the unambiguous cases, they were coded as either "Likely near-article, possibly non-partitive quantifier" or "Likely non-partitive quantifier, possibly near-article", depending on which interpretation was considered primary or more likely. The latter decision was based on the outcome of the meaning test: specifically, on which of the two paraphrases constructed for the meaning tests sounded more natural, more likely and/or closer to the intended meaning of the original sentence. Examples (29a-c) below are cases which were judged more likely to be instances of the near-article function but could also

be interpreted as instances of the non-partitive quantifier. Examples (30a-c) show the opposite case.

29) a) *He spent **some time** in Holland in the company of an Indian guru with whom he had long walks and conversations.* (EVH, W: biography)

   b) *The basis of the calculations caused **some confusion among oil companies which were planning to run overnight computer programs on the impact of the changes on their operations**.* (K5H, W:newsp:other: commerce)

   c) *We should reintroduce **some extinct species** among our hills and glens.* (AS3, W:misc)

30) a) *I know that that caused **some concern**, but, broadly speaking, my hon. Friend is right to say that standards in our schools are very good.* (HHV, W:hansard)

   b) *As a result, LASMO's accounts for 1992 contain **some significant changes in disclosure** and the restatement of reported figures for 1991 and earlier years.* (GX8, institute_doc)

   c) *Still he and Child were able to give **some advice based on their experiences with African elephants**.* (B7G, W:non_ac:nat_science)

This careful approach to the identification of near-article vs. non-partitive quantifier tokens of *some* presents us with two different ways to classify the functions of *some*. Under the parsimonious coding scheme, the potential ambiguity between near-article vs. non-partitive is recognized, leading to a four-way distinction between two ambiguous and two non-ambiguous categories. Under the generous coding scheme, the unambiguous tokens of the near-article function and the ambiguous ones that are more likely to be near-article are accounted in a single near-article category, and the same is done for the non-partitive quantifier function. There are pros and cons to both approaches: the parsimonious coding scheme is more precise and leads to more reliable conclusions, but the higher number of categories leads to sparser data and less power for statistical analysis; the opposite is true for the generous coding scheme. In this paper, we will alternate between the two schemes depending on the requirements of the analysis, while recognizing the other scheme when appropriate.

Besides the four main functions of *some* listed above, the annotation also turned up one clear instance of the so-called approximation function.[7] As it is the only one of its kind in our dataset, we excluded it from the analysis. Altogether, 58 other tokens were removed from the dataset. Of these, 23 tokens could not be annotated, either because not enough context was provided in the sentence to reliably identify the meaning of *some*, or because the example could simply not be made sense of as a whole. 35 tokens were excluded because they were not actual instances of direct object noun phrases despite occurring next to a verb, due to tagging errors by the dependency parser; this included for instance the phrases *some time*, *some place*, and *some way* used as adjuncts, noun phrases that were actually the subject of a complement clause, and pre-modifiers to words such as *later* and *ago*, among others.

---

[7] Namely, *this would still leave **some 60m tonnes** travelling by road in 1993* (BMJ, W:misc)

After filtering and annotating the data in this way, a multinomial regression model was fitted to investigate factors which increase the likelihood for *some* to function as a near-article. We report its findings in section 4.2. For an investigation of the competition between zero marking and *some* (i.e. RQ3) we fitted a binomial regression model (cf. 4.3)

4 Empirical analysis

4.1 Distribution of functions

We first examined how the various functions of *some* are distributed between our three samples, i.e. in combination with singular count nouns, non-count nouns, and plural nouns. The frequencies of occurrence of each function in each sample are summarized in Table 2, and their distribution is visualized as barplots in Figure 1. For the two ambiguous categories ("Likely near-article / possibly non-partitive quantifier" and "Likely non-partitive quantifier / possibly near-article"), similar colors are purposely used (namely, two similar shades of green and orange respectively), so that these categories can be visually grouped together and at the same time distinguished if needed.

	Singular count nouns	Noncount nouns	Plural nouns
Vagueness marker	257	16	8
Near-article	71	102	50
Likely near-article / possibly non-partitive quantifier	0	92	70
Likely non-partitive quantifier / possibly near-article	0	47	54
Non-partitive quantifier	0	53	68
Partitive quantifier	0	23	80
Approximator	0	0	1
Unclear	11	7	5
Not an object noun phrase	11	10	14
Total	350	350	350

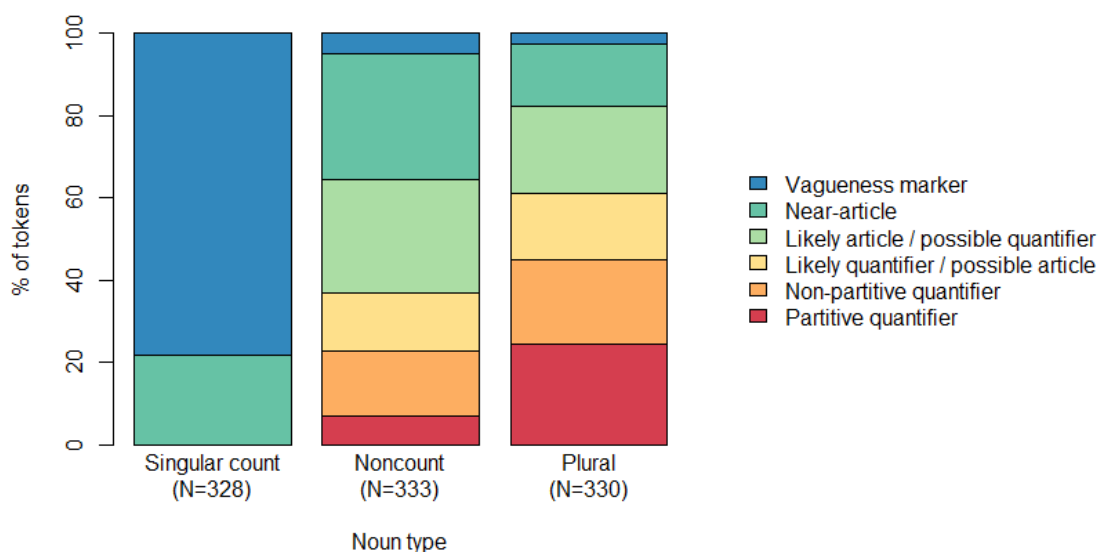Table 2: Frequency of the functions of *some* in each sample.

Figure 1: Distribution of the functions of *some* for each noun type.

As can be seen from Figure 1, the functions of *some* are distributed differently between the noun types. For example, with singular nouns, the vagueness function clearly dominates, but it is very uncommon with the other types of nouns, while plural nouns show the highest proportion of partitive readings. Non-count nouns and plural nouns are roughly equally likely to be used with the non-partitive quantifier function.

Importantly for our research questions, we also find the near-article function to be quite prominent with all noun types. It is the most common use of *some* with non-count nouns and the second-most common use with singular nouns (after being a vagueness marker). Under what we call the 'parsimonious coding scheme', the non-ambiguous near-article tokens make up 31% of the non-count noun data, and 15% of the plural noun data; under what we call the 'generous coding scheme' (i.e. when the ambiguous article tokens are added to the unambiguous tokens), these percentages increase to 58% and 36% respectively. All in all, these figures indicate that treating the near-article use of *some* as a marginal, barely emergent phenomenon could not be further from the truth. Rather, it is a well-established use that represents a significant share of the usage of *some* in a corpus of contemporary English. This gives credence to our view of *some* as a serious candidate to complete the paradigm of overt (in)definiteness marking with English NPs.

4.2   Grammatical behavior of functions

In this section, we turn to how the various functions of *some* behave grammatically with regard to the properties of the NP which we coded in our dataset, i.e. the presence of pre-modifiers and post-modifiers and the semantic class of the head noun. In particular, we focus on whether the near-article function of *some* differs in any way from the other functions in its grammatical behavior in our corpus data, especially from the quantifier functions, and how these differences relate to our hypothesis. Because the quantifier functions do not by nature occur with singular count nouns, it makes no sense to contrast the near-article function with the quantifier functions

with these nouns, since the latter functions are simply not available. As a consequence, we remove singular nouns from the dataset and consider only non-count nouns and plural nouns, which are attested with all four functions of *some* in our data. There is also a methodological reason to do so: since singular nouns never occur with the quantifier functions, they make a perfect prediction regarding the occurrence of these functions, i.e. whenever a noun is singular, there is a 0% chance for the quantifier functions to occur. This property is incompatible with the assumptions of logistic regression, which we use to analyze these data, but removing singular nouns fixes the issue.

To explore how the grammatical features coded in our dataset are distributed across the functions of *some*, and whether there are any significant differences between functions, we fitted our data to a multinomial logistic regression model (cf. Train 2009). Multinomial logistic regression aims to capture the statistical relation between a set of predictors and a categorical outcome (the dependent variable). It is an extension of the more well-known binomial logistic regression, but where the latter only supports binary dependent variables, multinomial logistic regression is designed to handle categorical dependent variables with more than two possible values. In our case, the dependent variable is the function of *some*, and the predictors are the grammatical properties coded for each token, i.e.: the presence/absence of pre-modification in the NP (yes vs. no), the presence/absence of post-modification/complementation in the NP (yes vs. no), the type of head noun (non-count vs. plural), and the semantic class of the head noun (concrete vs. abstract vs. measure). We used the generous coding scheme for the annotation of the functions of *some*, i.e. the two potentially ambiguous categories are counted with their most likely reading, and therefore we end up with only four categories: vagueness marker, near-article (both unambiguous and ambiguous), non-partitive quantifier (both unambiguous and ambiguous), and partitive quantifier.[8]

We used the R package mlogit to fit our data to this model (Croissant 2020). In this implementation of multinomial logistic regression, one level of the dependent variable is chosen as the reference level, and the model measures the effect of predictors on the selection of each other level compared to the reference level (cf. Train 2009, Levshina 2015: 275-283). We choose the near-article function as the reference level, and thus the model measures to what extent the near-article function differs from each of the other functions in terms of the various predictors. The fitted model is reported in Table below, with the coefficients of all effects and their levels of statistical significance. Significant effects are marked in bold. For the predictor "noun type", the reference level is non-count nouns, and for "semantic class" it is concrete nouns. This means that the effects reported in Table for these predictors are measured relative to the reference level, i.e. it is the effect of the variable when its value is changed from the reference level to another level. Pre-modification is involved in a significant interaction with

---

[8] This was done mostly to simplify the analysis and the interpretation of results, and to provide more statistical power for the model, as six levels is quite a high number for multinomial regression in a dataset of this size. In any case, we did not find substantial differences between the following analysis and one based on a six-level dependent variable, at least in terms of the claims that they support. On the whole, the unambiguous near-article category in the parsimonious coding scheme tends to differ from the other categories in the same ways as in the analysis presented below; however, the splitting of categories seems to introduce a measure of noise and random variation.

noun type, which we also report in Table 3. No other significant interactions between the other variables were found.

Near-article vs.	Vagueness marker	Non-partitive quantifier	Partitive quantifier
(Intercept)	-3.145 ***	-0.63 ***	-1.851 ***
Pre-modification (yes)	0.498	-0.835 *	-2.028 .
Post-modification (yes)	-0.542	0.092	-0.046
Noun type (plural)	0.471	0.404 .	1.761 ***
Semantic class (abstract)	0.918 .	0.145	-0.035
Semantic class (measure)	3.089 *	2.744 **	2.496 *
Pre-modification (yes) x Noun type (plural)	-1.511	0.945 *	0.934

Table 3: Coefficients of the multinomial regression model fitted by mlogit on the data for non-count nouns and plural nouns. The coefficients measure the effect of predictors on the likelihood of each function vs. the near-article function (reference level). Statistically significant effects are marked in bold, and their significance level is captured by the following symbols: $p < 0.05$: *, $p < 0.01$: **, $p < 0.001$: ***. The dot '.' indicates marginal significance ($p < 0.1$). McFadden's $R^2 = 0.073$.

Except for post-modification, all variables are involved in some kind of statistically significant difference between the near-article function and one or more other functions. Positive effects of a variable indicate that the relevant function of *some* is more likely to occur in our dataset than the near-article function when the corresponding level of the variable is found, and conversely, negative effects indicate that the relevant function is less likely to occur than the near-article function. The value of the coefficient captures the strength of the effect; higher absolute values correspond to stronger effects.

The effects found for noun type line up with the descriptive statistics reported in the last section. Compared to the near-article, the partitive quantifier function is significantly more commonly found when the noun is plural than when it is non-count, a similar effect is found with the non-partitive quantifier function, although it is weaker and only marginally significant. These findings mirror what we observed earlier with the distribution of functions in the three samples of noun types.

Turning to modification in the NP, there is no significant effect of post-modification found in the model. However, there is a significant negative main effect of pre-modification on the non-partitive quantifier function, and a much stronger and almost significant one ($p = 0.0502$) on the partitive quantifier function (also negative). In other words, the quantifier functions are both less typical with pre-modified head nouns, while the near-article function is more common with them. However, pre-modification is also involved in a significant interaction with plural nouns, with which there is a positive effect on the use of the non-partitive quantifier. This interaction counters the negative main effect just mentioned. This means that the effect of pre-modification on the use of the non-partitive quantifier function is largely restricted to non-count nouns, but the one on the partitive quantifier function is in evidence with both non-count and plural nouns.

We argue that the effect of pre-modification can be explained in terms of information density with respect to the functional differences between quantifier *some* and near-article *some*: quantifying a noun is relatively costly in terms of information density, and pragmatically it

tends to be the main thing that the speaker asserts about this noun. This leaves less room for pre-modification, which would make the NP semantically heavier. Near-article *some*, on the other hand, seems to leave the NP more open to pre-modification, as it has lost its partitive value and has become more semantically neutral, this being in line with our claim that it has become a marker of indefiniteness. At the same time, when a speaker uses an adjectival pre-head modifier, the semantics of this modifier often have a set-shrinking effect itself. In a way, the modifier indirectly quantifies and partitions by selecting a subset from a whole set (i.e. *red cars* vs. *cars*). If the modifier indirectly takes up a partitive function, coding it a second time by the quantifier *some* is somehow redundant. We hypothesize that this may invite semantic bleaching.

As to semantic class, we find only one effect of abstract nouns vs. concrete nouns on the use of the vagueness marker function, but it is only marginally significant; overall there is not much evidence for any of the functions showing a preference for any of these kinds over the other, compared to the near-article function. Effects are found with measure nouns, with which all three functions occur significantly more commonly than with concrete nouns (the reference level) compared with the near-article function, which, conversely, is seen to disprefer these nouns. Similar effects of measure nouns are found when the reference level is changed to abstract nouns, except for the vagueness marker,[9] confirming that this difference is not specific to the choice of reference level and is indeed due to measure nouns standing out against both the other semantic classes. This too can be explained in terms of functional differences between the functions of *some*: measure nouns (e.g. hours, kilometers,…) are meant to be quantified, hence they are less commonly found with a function of *some* that has lost its quantification value.

4.3   Competition between near-article *some* and zero marking

As noted earlier, with non-count and plural nouns, indefiniteness can be marked either by leaving the determiner slot empty (i.e. zero article) or through the use of the near-article function of *some*. In this section, we examine this competition and address to what extent, if at all, the two coding strategies are interchangeable. Specifically, we investigate what contextual factors, if any, might make it more likely for near-article *some* to be used over zero-marking, or vice versa. To do so, we use binomial logistic regression to model the relation between the use of *some* vs. zero-marking and a range of predictors. Binomial logistic regression is a simpler version of multinomial logistic regression used in the previous section, which is restricted to binary outcomes. This method is commonly used in variationist studies, where it has been typically applied to predicting what determines the choice between grammatical variants that are deemed to be somewhat equivalent.

To implement logistic regression, we need a sample containing tokens of both near-article *some* and zero-marking. We obtained the former kind of tokens from the *some* dataset; we chose the generous coding scheme and selected all near-article tokens (ambiguous and non-ambiguous) in order to give our analysis more statistical power, This provided us with 193 tokens with non-

---

[9] The coefficients and *p*-values of measure nouns in this alternate model are as follows: for the vagueness marker function, $\beta = 2.171$, $p = 0.147$; for the non-partitive quantifier function, $\beta = 2.6$, $p = 0.014$; for the partitive quantifier function, $\beta = 2.531$, $p = 0.02$.

count nouns, and 120 tokens with plural nouns. Although zero-marking is much more common than near-article *some* according to our data, we decided to match the near-article samples to the same number of tokens of zero-marking, instead of fitting the models with a dataset that mirrors the frequency imbalance of the two variants. This means that the baseline preference for each variant is artificially corrected to chance level, when a representative sample would contain a statistical bias towards zero-marking. On the other hand, matching the two samples for size allows us to observe the same degree of variation with each variant, and to give an equal amount of statistical power to each alternative.

To facilitate the selection of zero-marking tokens, we used the syntactically tagged database of direct object noun phrases mentioned in Section **Error! Reference source not found.**. Candidates for zero-marking were identified as those NPs that did not contain any words qualifying as a determiner in a dependency relation with the head noun. Our aim was to select tokens of zero-marking that at least *could* also be phrased with *some*, if only marginally. In order to do so, we first restricted the dataset to only head nouns that were also found in the relevant sample of near-article *some*. This ensured that we were dealing with nouns that are actually attested with both kinds of indefiniteness coding. Second, we manually selected a random sample of this dataset that contained only tokens that actually have an indefinite referent, to the exclusion of those that make a generic reference to an entire class of entities (e.g. *I hate paper straws*), or those involved in a fixed idiomatic expression (e.g. *part company*).

Following these criteria, we extracted 193 randomly selected tokens of zero-marking with noncount nouns, and 120 with plural nouns. These tokens were then annotated for the same grammatical features as the tokens of *some* analyzed above, namely the presence of absence of pre-modifiers and post-modifiers and the semantic class of the head noun (concrete vs. abstract vs. measure noun). Additionally, we also annotated for polarity in both the zero-marking and near-article *some* data. As noted in section 2, *some* is considered a positive polarity item, i.e. a word that is repelled by negative contexts, in that it is either unacceptable in a negative clause (cf. 31), or if acceptable takes wide scope over negation; hence 32) below cannot be interpreted as the negative counterpart of *I read some books*, as its meaning entails the reading of all the books except some of them (both examples are from Israel 2000: 170).

31)   **You don't have some peanut butter on your chin.*

32)   *I didn't read some books.*

Hence in such a context, alternatives such as *any* or the zero article would be used instead of *some* for the intended narrow scope negative meaning to be expressed or for the sentence to be grammatical altogether.

In light of these facts, we have good reason to suspect that the use of near-article *some* might be affected by polarity. What we mean here by polarity is not restricted to the contrast between negative vs. affirmative sentences, but is understood more broadly as a set of non-affirmative contexts, which we identify as those sentence contexts which negative polarity items such as *ever* or *any* are restricted to. Lists of contexts licensing negative polarity items can be found in the vast literature on the topic, notably Buyssens (1959), Ladusaw (1980), and Zwarts (1995).

In our dataset, the following types of non-affirmative contexts could be identified and were classified as negative polarity:

- Negative clause, either by means of a negative adverb like *not* or *never*, or through another form of negation like *none*, *nothing*, *not that* etc., e.g. *I am afraid I don't keep <u>brandy</u>*, *None of Stotler's 65,000 customers had lost <u>money</u>*.
- Subordinate clause with a non-affirmative predicate, either lexically (e.g. verbs like *doubt* or *refuse*) or though grammatical negation of the main clause, e.g. *Some schools hesitated to implement <u>new ideas</u>*.
- Interrogative sentences, e.g. *But do the follicles actually contain <u>eggs</u>?*
- Conditional clauses with *if*, e.g. *Simply, if we donate <u>money</u>, we should expect it to be spent on British goods*.
- Predication of excess with *too* or *so*, e.g. *It is too indeterminate in pitch to give <u>satisfactory support</u>*.
- Restrictor of a universal quantifier like *everything*, *anyone*, etc., e.g. *Anyone who wishes <u>information about bus services</u> goes to a bus*.

The combined annotated datasets of non-count nouns and plural nouns, with both near-article *some* tokens and zero-marking tokens in each case, were each fitted to a logistic regression model with the occurrence of *some* as dependent variable, and pre-modification, post-modification, and polarity as predictors. We used the 'lrm' function from the 'rms' package in R. A frequency predictor related to the head noun of each token was also included, as from a usage-based perspective we expect frequency to play a role. Three types of frequencies were extracted from the corpus and tested: (i) the frequency of the noun lemma in the entire BNC, (ii) the frequency of the noun lemma used as the head of a direct object of any verb in the entire corpus, and (iii) the frequency of the noun lemma used the head of a direct object with the same verb as the one found in each token; the latter two frequencies were calculated from the syntactically annotated dataset. All frequencies were log-transformed in order to reduce the impact of outliers. Unsurprisingly, all three frequencies are highly correlated, hence only one of them should be included in the model so as not to violate the statistical assumption of regression modelling. After trying all three frequency predictors separately, we found the first one (overall frequency of the noun lemma) to have the strongest effect, hence we only kept this one in the models discussed below. Finally, the semantic class of the head noun did not add significant effects and did not improve the fit of the model, and since it was also shown not to be of any relevance to the near-article function of *some* in the previous analysis, we decided to drop this factor from the model to gain statistical power.

The logistic regression models for non-count nouns and plural nouns are summarized in Table and Table respectively. The statistics of all predictors are listed, along with their statistical significance values. Remember that the dependent variable is the use of near-article *some* as opposed to zero-marking; this means that positive effects correspond to predictors that make the use of *some* more likely in the model, while negative effects are predictors that make the use of zero-marking more likely (and vice versa). The concordance index *C* returned by the 'lrm' function is 0.73 for the first model and 0.81 for the second one, which correspond to acceptable discrimination and excellent discrimination respectively according to Hosmer &

Lemeshow (2000: 162). In other words, these models discriminate well and are a good fit to our data.

	Estimate	Std. Error	z value	*p*-value
(Intercept)	**5.668**	**1.005**	**5.642**	**< 0.001**
Pre-modification (yes)	**-0.466**	**0.245**	**-1.904**	**0.057**
Post-modification (yes)	0.020	0.249	0.079	0.937
Polarity (positive)	**0.765**	**0.453**	**1.689**	**0.091**
logFreqHeadNoun	**-0.671**	**0.096**	**-6.971**	**< 0.001**

Table 4: Summary of the logistic regression model for noncount nouns. Significant and marginally significant effects are marked in bold. logFreqHeadNoun = log-transformed frequency of the head noun lemma in the BNC. $C = 0.73$, $R^2 = 0.252$.

	Estimate	Std. Error	z value	*p*-value
(Intercept)	**5.836**	**1.455**	**4.01**	**< 0.001**
Pre-modification (yes)	**0.577**	**0.311**	**1.855**	**0.064**
Post-modification (yes)	-0.280	0.315	-0.89	0.373
Polarity (positive)	**1.940**	**1.090**	**1.779**	**0.075**
logFreqHeadNoun	**-0.828**	**0.122**	**-6.804**	**< 0.001**

Table 5: Summary of the logistic regression model for plural nouns. Significant and marginally significant effects are marked in bold. logFreqHeadNoun = log-transformed frequency of the head noun lemma in the BNC. $C = 0.810$, $R^2 = 0.379$.

The two models are quite similar and show that the same factors seem to be at play for both non-count nouns and plural nouns. In both models, there is quite a strong positive effect of positive polarity, especially for plural nouns, although it is only marginally significant ($p < 0.1$). This means that *some* is more frequently used with positive polarity, which lines up with our intuition that *some* is avoided in negative polarity contexts, and that zero-marking is preferred in such contexts. There is also in both models a strong and very significant negative effect of the frequency of the head noun. This means that high-frequency nouns tend to be used more commonly with zero-marking, and lower-frequency nouns with near-article *some*. This effect can receive a straightforward interpretation in a usage-based model: following our hypothesis that the usage of *some* as a marker of indefiniteness is a relatively recent development that may only slowly be gaining ground, a usage-based approach to grammar predicts that it may be subject to conservation effects due to high token frequency, i.e. high frequency items are more entrenched in the established construction and thus resist the spread of the newer construction. Similar effects are attested in many areas of the grammar of various languages (cf. Bybee 1995, Bybee & Thompson 1997, Ogura 1993, *inter alia*), and this seems to be what we are observing with near-article *some* when compared to the more established zero-marking construction.

Turning to modification, there is no significant effect of post-modification in either model. Pre-modification does appear to have an effect in both models, albeit an only marginally significant one ($p < 0.1$). However, the effect is negative for non-count nouns, and positive for plural nouns. In other words, with pre-modified non-count nouns speakers tend to avoid *some* and prefer zero-marking, but with pre-modified plural nouns *some* is preferred. It is not clear to us how to interpret this result.

5   Interpretation

The analyses in the previous section have revealed the following facts: *Some* clearly functions as an article-like element exclusively coding existence/ indefiniteness in many examples. This is especially the case for NPs with non-count and plural noun heads. At the same time, *some* is even a possible alternative for the indefinite singular article *a/an*. In many singular NP examples, the context does not suggest that speakers have chosen *some* to express non-specificity or vagueness. In many cases, *some* does nothing but introduce the existence of a referent, synonymous to the indefinite singular article *a/an*. That does not deny the fact that the indefinite article *a/an* is clearly the default marker of indefiniteness in singular contexts in Present Day English. It is worth mentioning, that already in Old English *some* and its predecessors competed with the numeral (*ane* 'one') to code indefinite existence. Obviously, the numeral won the competition and became the default marker of definiteness. It seems that the PDE usage of *some* with singular nouns is a leftover from these times (Sommerer & Hofmann 2021).

At the same time, in plural and non-count contexts, we argue that *some* has started to compete with the historically 'older' bare option in interesting ways. Here, the regression model in 4.3 is full of relevant findings. First, high frequency nouns are less open to being marked by *some*. Second, *some* prefers positive polarity whereas zero marking is more often found in negative statements. It is likely that the existence and functions of *any* currently still slows down *some*'s expansion into the negative domain. Although some of our findings are hard to interpret (e.g. influence of modification), we conclude that our data empirically confirms that *some* is a versatile word which – among other meanings – has clearly taken up the function of an indefiniteness marker. Even if the bare option is still much more frequent overall, *some* is an acceptable alternative.

We now would like to discuss the theoretical implications these (synchronic) findings have for a diachronic model of language change. We believe that synchronic distributions can — albeit indirectly — shed some light on diachronic developments. We argue that the functional versatility of *some* is due to it undergoing further grammaticalization (including processes like semantic bleaching, obligatorification and paradigmatization). According to Lehmann "in many cases of paradigmatization, grammaticalized elements join preexistent paradigms and assimilate to the other members" (1995[1982]: 120). These members then tend to be tightly integrated into the paradigm. Moreover, when elements join a grammatical paradigm they tend to be used more obligatorily. Lehmann speaks of a decrease in "transparadigmatic variability" (124). At the same time, *some*'s trajectory constitutes a case of constructional competition (Zehnenter 2019). *Some* and the bare marking option have entered a competitive situation in which *some* as the newcomer tries to establish itself as a new coding option. The question is why *some* has entered this competition? The grammaticalization and competition seems to be triggered by analogical thinking (Fischer 2007; de Smet 2012) and what could be called 'paradigm pressure' (Sommerer 2018).

To fully understand this line of argumentation the following diachronic facts need to be discussed briefly: Originally completely lacking an article category, English developed one as a result of a complete systemic reorganization of (in)definiteness marking in OE and ME from

covert to overt marking of discourse status (Christophersen 1939; Rissanen 1967; Traugott 1992; Sommerer 2018, Sommerer & Hofmann 2021). First, it became obligatory to mark definiteness overtly. As a consequence, the OE demonstrative pronoun se 'that' was recruited and grammaticalized into the definite default article *the*. This was followed by the numeral *ane* 'one' to develop into the indefinite article *a(n)* for singular nouns in ME.

Interestingly, the diachronic process of recruiting overt markers of (in)definiteness has somehow stalled with indefinite NPs that have a plural or non-count noun as their head. Similar to other languages (e.g. German or Italian), it is not necessary to use an overt determiner to indicate discourse status in indefinite plural and noun-count contexts in PDE. A detailed discussion of why these contexts seem to be successfully resisting overt marking in many languages far exceeds the scope of this paper. Although we acknowledge that covert zero-marking of indefiniteness is a typological strategy which seems evolutionarily stable and works fine in many languages[10], we still present the following hypothesis: in English, this covert coding option might decline in the future as it constitutes the 'odd-man-out' in an article system where in all other cases (see Table 5) definiteness status has become marked overtly and obligatorily in referential NPs.[11]

	referential NPs			
	overt obligatory definiteness marking		overt obligatory indefiniteness marking	
	singular	plural/mass	singular	plural/mass
early OE	-	-	-	-
middle OE	(+)	(+)	-	-
late OE	+	+	-	-
early ME	+	+	(+)	-
middle ME	+	+	+	-
late ME/EME	+	+	+	(+)/-

Table 5. Development of overt (in)definiteness marking in English (Sommerer & Hofmann 2021:148)

Recruiting *some* as a near-article for plural and non-count contexts systemically 'fills the last empty cell' which corresponds to the speakers' behavior of coding indefinite status overtly in all cases (sg/ nc/ pl). In other words, the majority strategy gets analogically extended. A linguistic analogical action is performed when the language user understands common similarities between two strings (generalization), abstracts a more abstract pattern (analogical reasoning) and applies this to a third instance (analogical extension). In our case, the speaker recognizes that (in)definiteness in many cases is marked overtly by a determiner, abstracts this strategy of obligatory overt marking and recruits *some* as a new member of the paradigm to

---

[10] Admittedly, article paradigms are 'incomplete' in many languages in the sense that they only have a definite but not an indefinite article (e.g. Icelandic, Bulgarian, Rumanian). In other words, many languages have (article) paradigms which use zero marking as a functional choice as well.

[11] It is clear that one will always find exceptions to the rule of overt marking (e.g. some idiomatic expressions or light verb constructions lack an overt article), but we are discussing the default coding strategy for the vast majority of NP cases.

function as a new article in plural and mass noun contexts. This, at the same time, is a case of paradigm completion.

Adding this overt determiner in these previously unmarked contexts has certain cognitive and processing advantages as well: first the overt determiner is a quick signal for nominality and referentiality (Hawkins 2004: 87-89). Obligatory overt definiteness marking increases communicative speaker-listener interaction by reducing difficult inferencing from context. An overt definiteness marker as a functional element to code anaphoric intertextual relations improves informativeness by distinguishing between referential and non-referential contexts or definite and indefinite contexts (Sommerer & Hofmann 2021). Moreover, it is more straightforward to mark referentiality overtly and obligatorily in all cases than to constantly remember in which particular cases (combinations) it is required and in which it is not. Admittedly, 'paradigm pressure' (in this case its completion) is one of many driving forces of language change and can be counteracted by other forces. We definitely do not claim that *some* is currently on its way to oust the zero marking option. At the moment, zero marking is still the predominant strategy. We have merely showed that *some* has entered the competition and identified variables which favor its use. Ultimately only time will tell in which direction English is heading regarding this paradigm's completion.

6 Conclusion & Outlook

In this article we have shown that in Present Day English *some* functions in an article-like manner quite often. In many examples, the quantificational aspect no longer plays any role and *some* has turned into a semantically bleached marker of indefiniteness. That being said, *some* still fulfills its original function of a partitive quantifier and it is also often used in a non-partitive quantificational way. This empirical, corpus-based study, which is the first of its kind, confirms that the functional distribution of *some* with singular count nouns deviates significantly from the one with non-count and plural nouns. The vagueness function is especially high with singular nouns, whereas the near-article function is most prominent with non-count nouns. At the same time, the second study on the competition between *zero* and *some* has revealed that *some* is dispreferred with high frequency nouns and in negative polarity statements.

We have interpreted this synchronic distribution as evidence that *some* has been undergoing grammaticalization and has bleached its quantificational semantics when it gets recruited as the new and most recent member of the article paradigm. It currently is an infrequent alternative marker of indefiniteness next to *a/an* for singular nouns and an possible alternative to the zero article for plural and non-count contexts. We have argued that the current observable situation is a case of ongoing constructional competition with *some* in the role of David and zero marking as Goliath. It is still acceptable and the default option not to overly code indefiniteness in plural and non-count context, but *some* has already become a serious competitor especially in positive statements and might be gaining ground in the future. This process was interpreted as a case paradigm completion triggered by paradigm pressure. At the same time, the observable phenomenon could also be conceptualized as a case of 'overtification' that might be driven by

analogical thinking and communicative efficiency. The majority pattern of overt marking gets extended to the cases where covert coding is still acceptable.

Many questions need to be answered in future research. The presented corpus study has exclusively focused on object position in Present Day English and it will be necessary to extend the analysis to other argument positions (subject position etc.). It is likely that the functional distribution might be different in other argument positions. As a next step, it will also be necessary to conduct a diachronic study by using historical corpora and some periodization (e.g. in the EEBO, ARCHER, COHA) to see if the article-like function of *some* has increased in the last 300 years (see Sommerer & Hofmann 2021 for an increase in ME and EModE). At the same time, psycholinguist experiments could shed more light on English speakers' preferences when it comes to choosing *some* instead of zero marking. That being said, this paper has successfully shown that *some* has entered the article paradigm in English and should not be overlooked as a marker of indefiniteness.

7   References

Barðdal, Jóhanna, Elena Smirnova, Lotte Sommerer & Spike Gildea (eds.). 2015. *Diachronic Construction Grammar*. Amsterdam: John Benjamins.

Bolinger, D. 1977. *Meaning and Form*. London: Longman.

Buyssens, Eric. 1959. Negative Contexts. *English Studies* 40: 163-169.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* 10(5): 425–455.

Bybee, J., & Thompson, S. 1997. Three frequency effects in syntax. *Berkeley Linguistics Society* 23, 65–85.

Chen, D. and Manning, C. 2014. A Fast and Accurate Dependency Parser Using Neural Networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 740-750. Doha, Qatar: ACL.

Chesterman, Andrew. 1991. *On Definiteness. A Study with Special Reference to English and Finnish*. Cambridge: Cambridge University Press.

Chesterman, Andrew. 1993. Articles and no articles. In Andreas Jucker (ed.). *The Noun Phrase in English. Its Structure and Variability*, 13-24. Heidelberg: Universitätsverlag C.Winter.

Christophersen, Paul. 1939. *The Articles: A Study of their Theory and Use in English*. Oxford: Oxford University Press.

Croissant, Yves. 2020. Estimation of Random Utility Models in R: The mlogit Package. *Journal of Statistical Software*. 95(11): 1–41.

De Smet, Hendrik. 2012. The course of actualization. *Language*, 88, 601–633.

Diessel, Holger. 2019. *The Grammar Network. How Linguistic Structure is shaped by Use*. Cambridge. CUP.

Downing, Angela. 2015. *English Grammar. A University Course*. New York: Routledge.

Du Bois, John W. 1987. The discourse basis of ergativity. *Language* 64: 805-855.

Duffley, Patrick & Pierre Larrivee. 2012. Exploring the relation between the qualitative and quantitative uses of the determiner some. *English Language and Linguistics*. 16(1):131-149.

Farkas, Donka. 2002. Varieties of indefinites. *Proceedings of SALT XII*, 1-25.

Fischer, Olga. 2007. *Morphosyntactic Change. Functional and Formal Perspectives*. Oxford: Oxford University Press.

Fischer, Olga. 2010. An analogical approach to grammaticalization. In K. Stathi, E. Gehweiler, & E. König (Eds.), Grammaticalization. Current Views and Issues (pp. 181–219). Amsterdam: John Benjamins.

Goldberg, Adele E. 2006. *Constructions at Work. The Nature of Generalization in Language*. Oxford: Oxford University Press.

Halliday, Michael A.K. & Christian M.I.M. Matthiessen. 2014. *An Introduction to Functional Grammar*, 4th edn., revised by Christian M.I.M. Matthiessen. London: Routledge.

Hawkins, John A. 1978. *Definiteness and Indefiniteness. A Study in Reference and Grammaticality Prediction*. London: Croom Helm/Routledge.

Hawkins, John A. 1991. On (in)definite articles: implicatures and (un)grammaticality prediction. *Journal of Linguistics* 27. 405–442.

Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Hilpert, Martin. 2021. *Ten Lectures on Diachronic Construction Grammar*. [Distinguished Lectures in Cognitive Linguistics]. Brill.

Hirtle, W.H. 1988. Some and any: exploring the system. *Linguistics* 26: 443-477

Hosmer, D.W. & Lemeshow, S. 2000. *Applied Logistic Regression*. New York: Riley.

Huddleston, Rodney & Geoffrey K. Pullum (eds.). 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Israel, Michael. 2000. Some and the pragmatics of indefinite construal. *Annual Meeting of the Berkley Linguistics Society* 25(1): 169-182.

Jacobsson, Bengt. 2002. The Indefinites Some and Any in Linguistic Theory and Actual Usage. *Studia Neophilologica* 74(1): 1–14.

Ladusaw, W.A. 1980. *Polarity Sensitivity as Inherent Scope Relations*. PhD dissertation, University of Texas at Austin, IULC: Bloomington, Indiana.

Langacker, Ronald W. 2016. Nominal grounding and English quantifiers. *Cognitive Linguistic Studies* 3(1) ,1-31.

Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar, Vol. 2, Descriptive Application*. Stanford, CA: Stanford University Press.

Lehmann, Christian. 1995 [1982]. *Thoughts on Grammaticalization*. [LINCOM Studies in Theoretical Linguistics, 1]. Munich: Lincom Europa.

Levshina, Natalia. 2015. *How to do Linguistics with R: Data exploration and statistical analysis.* Amsterdam: John Benjamins.

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

Narrog, Heiko & Bernd Heine. (2021). Grammaticalization. Oxford: Oxford University Press.

Ogura, M. (1993). The development of periphrastic *do* in English: A case of lexical diffusion in syntax. *Diachronica*, 10(1), 51–85.

Partee, Barbara, Alice ter Meulen & Robert E. Wall. 1990. Mathematical metods in linguistics. Dordrecht: Kluwer.

Payne & Huddleston. 2002. Nouns and Noun Phrases. In Rodney Huddleston and Geoffrey Pullum (eds), *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. Harlow: Longman

Rissanen, Matti. 1967. *The Uses of One in Old and Early Middle English*. Helsinki: Modern Language Society.

Sahlin, Elisabeth. 1979. Some *and* Any *in Spoken and Written English*. Uppsala: Acta Universitatis Upsaliensis.

Smirnova, Elena & Sommerer, Lotte. 2020. The nature of the node and the network: open questions in Diachronic Construction Grammar. In Lotte Sommerer & Elena Smirnova. (Eds.), *Nodes and Networks in Diachronic Construction Grammar*. [CAL, Vol. 27]. Amsterdam: John Benjamins, 1-42.

Sommerer, Lotte. 2018. *Article Emergence in Old English. A constructionalist perspective.* [TIEL series, Vol. 99]. Berlin: De Gruyter Mouton.

Sommerer, Lotte & Hofmann, Klaus. 2021. 'Constructional competition and network reconfiguration: investigating *sum(e)* in Old, Middle and Early Modern English'. *English Language and Linguistics,* 25(1), 1-33.

Train. Kenneth. 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press.

Traugott, Elizabeth C. 1992. Syntax. In Richard M. Hogg (ed.), *The Cambridge History of the English Language. Volume 1. Old English. From the beginnings to 1066*, 168-289. Cambridge: Cambridge University Press.

Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.

Zehentner, Eva. 2019. *Competition in language change: The rise of the English dative alternation*. Berlin: De Gruyter Mouton.

Zwarts, Frans. 1995. Nonveridical contexts. *Linguistic Analysis* 25: 286-312.