Dagmar Divjak*, Laurence Romain and Petar Milin

From their point of view: the article category as a hierarchically structured referent tracking system

https://doi.org/10.1515/ling-2022-0186
Received November 22, 2022; accepted March 8, 2023; published online June 9, 2023

Abstract: Full-fledged grammatical article systems as attested in Germanic and Romance languages are rather uncommon from a typological perspective. The frequency with which articles occur in these languages, together with the difficulty encountered in detecting them and the lack of a water-tight account of article use, make article errors one of the most frequent errors in language produced by L2 learners whose L1 does not feature an article system of similar complexity, all the while appearing unproblematic for L1 users. We present a conceptually and methodologically interdisciplinary approach to the grammatical category of articles in English and combine a usage-based, cognitive linguistic account of the function and use of articles that respects its discourse-based nature with a computational exploration of the challenges the system poses from the perspective of learning. Running a statistical classifier on a large sample of spoken and written discourse chunks extracted from the BNC and annotated for the five main determinants of article use reveals that Hearer Knowledge is the driver of a hierarchical system. Once Hearer Knowledge is acknowledged as the motivating principle of the category, article use becomes eminently predictable and restrictions are in line with the forms from which the articles have developed historically, with *the* and *a* acting as category defaults and zero acting as default override. Simulations with a computational model anchored in the psychology of learning shed light on whether and how human cognition would handle the proposed relations detected in the data. We find that different articles have different learnability profiles that, again, are in line with their historical development: while *the* can be learned from one strong indicator, the relationships for the zero article are less exclusive. On the basis of these findings,

***Corresponding author: Dagmar Divjak**, Department of Modern Languages, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK; and Department of English Language and Linguistics, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK, E-mail: divjak@bham.ac.uk. https://orcid.org/0000-0001-6825-8508

Laurence Romain and Petar Milin, Department of Modern Languages, University of Birmingham, Birmingham, UK, E-mail: l.m.y.romain@bham.ac.uk (L. Romain), p.milin@bham.ac.uk (P. Milin). https://orcid.org/0000-0001-9708-7031 (P. Milin)

we argue that the article category appears as a referent tracking system that grammaticalizes the principles of "audience design": it forces a speaker to track and mark reference from the vantage point of the memory of the hearer, thereby reducing the processing effort required from the hearer. This particular mindset inverses the typologically dominant situation in which this information is not explicitly marked by the speaker but implicitly retrieved from context by the hearer.

Keywords: articles; classification; English; learnability; referent tracking system simulation

1 Introduction

In Western European languages articles are the most frequently used words: they are obligatorily expressed whenever a noun is used. In English, Leech et al. (2001) report that *the* is the most frequently used word (frequency = 61,847), with *a/n* combined occupying fourth place (frequency = 25,056). In fact, based on British National Corpus data (Leech 1992), every fifth word is a noun, requiring an article. Irrespective of their frequency, articles are, quite notoriously, not the first words children acquire (Lidz et al. 2003: 151). Although children make few mistakes when using articles from a very young age (Brown 1973), their initial productions are limited in the sense that the set of articles that accompanies a particular noun is restricted, with certain articles reserved for certain nouns. Full productivity is achieved only gradually, once sufficient item-based knowledge has built up (Meylan et al. 2017).

Articles are relatively hard to detect in input, for L1 and L2 learners alike. L1 learners of French have been documented to segment articles and nouns incorrectly in the speech stream, considering the last consonant of the article as the first consonant of the noun, e.g., they segment *les arbres* 'the trees' as /le/+/zarbr/ (Chevrot et al. 2009). English articles come with their own challenges: they are either unstressed (*a/the*) or invisible (zero), which makes it difficult for L2 learners to learn them from exposure. The English article system emerges as one of the most difficult aspects of grammar and one of the latest to be fully acquired (Master 1990: 461). Even though we are dealing with a seemingly simple 3-way choice, there are no reliable rules to guide L2 learners: the use of articles has long defied accurate description. A cursory inspection of pedagogical grammars reveals disagreement in how the grammatical category is described, alongside a reliance on extensive lists of specific usage instances (Biber et al. 1999; Greenbaum 1996).

The frequency with which articles occur, in combination with the difficulty encountered in detecting them, and the lack of a water-tight account of article use that balances the requirements of broad coverage with sufficient detail (see also

Trenkic 2000: 23, 110), makes article errors one of the most frequent errors in L2 language production (Master 1990; Shin and Kim 2017; Thomas 1989; Trenkic 2008). Yet, despite the fact that the speech of many an L2 learner of English is rife with article errors, these errors are considered among the least irritating a non-native speaker can make (e.g., Magnan 1982; Vann et al. 1984; Master 1997: 216) and a 9-laboratories (N = 334) replication study of DeLong et al. (2005) did not find any evidence that native speakers of English respond to article errors (Nieuwland et al. 2018). Why are violations against an obligatorily marked category treated so leniently? What information do articles encode?

In this article we take a usage-based, cognitive linguistic approach to the function and use of articles, and explore the challenges the system poses from the perspective of learning. After a brief survey of the origins and properties of the grammatical category of articles (Section 2), we test the classification power of the five most frequently invoked morphological, lexical and pragmatic variables on a large sample of discourse chunks extracted from the (spoken and written) BNC (Sections 3 and 4). Computational simulations with an error-correction algorithm that mimics the way in which humans learn from data are used to explore the learnability of the article system based on these same five properties (Section 5). We use our findings to propose a hierarchy of experiential and grammatical dimensions that accurately captures usage of the grammatical category of articles in English in a large sample of naturally occurring data and reflects the historical origins of the article category. Our findings also stress the importance of embedding insights relating to human cognition methodologically: modeling a complex phenomenon in a way that reflects how we learn from data facilitates a radically different type of language description and linguistic theorizing (Section 6).

2 The article system

2.1 The geographical spread and historical emergence of article systems

The grammatical category of articles is rather uncommon from a typological perspective. Dryer (1989: 86) reported, based on a survey of 399 languages, that only 125 or about a third of the languages in his sample had articles. More strikingly, only 31 or 8 % of all languages surveyed, were found to have both definite and indefinite articles. Globally, the definite article is more widespread than the indefinite article: out of the 620 languages surveyed for the WALS, 377 had a definite article of some kind (Dryer 2013). A language that has a definite article is very likely to also have an

indefinite article, yet the converse does not hold (Heine 1997: 69): only 45 languages in the sample had an indefinite article but no definite article of any kind. Most of the languages that have both definite and indefinite articles are geographically clustered in Europe and in Western Europe in particular.

The fact that the definite article is more widespread than the indefinite article has been taken to indicate that definite articles emerged earlier than indefinite articles (De Mulder and Carlier 2011). Overall, the grammatical category of articles is a relatively "recent" phenomenon: its origins only date back to the 11th century. The emergence of an article category has been attributed to the loss of case and aspect systems in early Germanic (Leiss 2000), which offer alternative ways of expressing definiteness. Interestingly, a common grammaticalization process seems to underlie the formation of the article category: in most cases, the definite article traces back to a (weakened) demonstrative while the indefinite article developed from the numeral *one* (De Mulder and Carlier 2011). Sommerer (2018) sketches, for English, how the definite article *the* emerged in Late Old English from the dependent usage of the demonstrative *se* (*se* was also used independently as a complementizer or pronoun such as *that, which, who*); due to the length of its existence, it is more grammaticalized. Next, the indefinite article *a* developed from the numeral *one* in Middle English, and was applied to singular count nouns only, with non-counts and plurals not requiring an overtly marked article. Due to the fact that it is not overtly marked, the emergence of the zero article is more controversial (see Sommerer 2018: 59 for details). It can be argued that it appeared in Old English as a contrast between a marked definite and an unmarked indefinite or that it correlates with the emergence of the obligatory marking of indefinite *a* in Middle English. In present-day English, the indefinite remains unmarked for plural and non-count nouns.

2.2 Articles as a referential category: from world to discourse

In English, and many other European languages, articles are often considered part of a larger category of determiners which also contains demonstratives and various other words that occur in the same position within noun phrases (Dryer 2013): articles are typically not found when another determiner such as a possessive, demonstrative, distributive or quantifier is present (De Mulder and Carlier 2011). Determiners constitute a core part of the referential system of a language and are used to express the reference of a noun used in context. In English, demonstratives typically express the relation of the nouns they modify to the speaker. For example, demonstrative *this* indicates relative closeness to the speaker while *that* signals relative distance from the speaker. Possessives likewise take the speaker as source, with different determiners referring to the speaker (*I*), the addressee (*you*, singular

and plural), combinations including both speaker and addressee (*we*) or agents remaining outside of the conversation (*his/her/their*). It is unsurprising, then, that with few exceptions, articles have been interpreted from that same vantage point.

Within philosophy and the philosophy of language, much work on referring expressions sits within the discussion of denotation and revolves around the notion of presuppositions or the assumptions a speaker makes about a listener's knowledge. Definiteness, which includes the definite article *the*, has attracted the most attention (but see Burton-Roberts 1976). In philosophy, it was long thought that the notions of existence and/or uniqueness held the key to the appropriate use of definite descriptions and that they required reference to an entity in the real world (Russell 1905). Strawson (1950) moved away from real-world reference, thus reflecting the start of a pragmatic turn. This turn was further strengthened by work in psychology on audience design (Clark and Murphy 1982) which highlighted the importance of taking into account the knowledge of the hearer in encoding and resolving reference. Clark's (1996) notion of "common ground", or "the sum of [two people's] mutual, common or joint knowledge, beliefs, and suppositions" took hold in linguistics. Later, in philosophy, Roberts (2003: 288) argued that the presupposition alludes to the information available to the discourse participants rather than to the world. Hence, use of a definite Noun Phrase (NP) presupposes a familiar discourse referent that is unique among the discourse referents; familiar here does not require previously mentioned but merely entailed by the context of interpretation. The latter is illustrated in (1) from the BNC: the use of *the* before *local estate agency* indicates that the speaker assumes that the addressee is aware of the existence of said agency.

(1) *Had handsome Henry tried to foreclose? Or, seeing how pretty Mrs. Yardley was, had he suggested an alternative form of payment?*
Peggy thought she knew of at least another ten people who would not really be mourning the death of Henry Phipps.
There could be as many more? Dozens more? About whom she knew absolutely nothing at all.
*As manager of <**the**> local estate agency and building society, handsome Henry got around.*

Linguistics, long dominated by Structuralism,[1] likewise welcomed the focus on definiteness: *the* could be considered as the marked member of a privative

1 Generative syntacticians have also shown an interest in the topic of definiteness. They commonly assume that in a language with articles, the articles correspond to the heads of determiner projections. There are several possible treatments of languages that do not have articles and some of these treatments trigger interesting theory-internal inconsistencies (see, for example Despić 2019). These discussions do not bear on the topic investigated in this article, however, and will therefore not be fleshed out.

opposition. Seeing articles as a privative opposition between the active, marked *the* and the inactive, unmarked *a* would have encouraged the search for an invariant meaning of *the*, bypassing the larger discourse situation. Crosslinguistically, articles have been said to encode definiteness (for an overview see Kibort 2008). The linguistic interpretations of definiteness can be classified as based on either familiarity (Christopherson 1939, elaborated by Kamp 2013; Heim 1983) or (unique) identifiability (for a definition of *uniqueness* see Kadmon 1990; for *identifiability* see Chafe 1976 or DuBois 1980 or Lyons 1999; for *unique identifiability* see Givón 1984 or Hawkins 1978). Clark's (1996) notion of "common ground" operationalized Chafe's (1976) sense of the importance of the speaker's conception of what their addressee knows, and the idea of shared knowledge also runs as a red thread through Lambrecht (1994). Note that in linguistics, too, reference is to the discourse: real-world reference is not required (for discussion see Langacker 1991: 97–98).

(Unique) identifiability and familiarity are not equivalent (Birner and Ward 1994: 96–97): (unique) identifiability requires the referent of the Noun Phrase introduced by *the* to be the only entity of that type within the discourse and uniquely identifiable to the hearer, while familiarity requires only that the referent of the NP introduced by *the* has already been introduced into the discourse (Kadmon 1990: 274, 279; Birner and Ward 1994: 93; Epstein 2002: 336). Although most contemporary approaches to definiteness opt for either uniqueness (Abbott 1999; Hawkins 1978; Kadmon 1990) or familiarity (Chafe 1976; Green 2012), very often, identifiability and familiarity apply simultaneously: entities that are part of the discourse are identifiable, as in (2) below, where speaker and hearer both know there is a garage and that there is only one in that location, which makes it the only possible referent.

(2) [Speaker0084] *Is it that little row of shops just before you get to <**the**> garage then?*

It has also been argued that neither property alone can account for all acceptable uses of the definite article (Birner and Ward 1994: 101; Farkas 2002; Roberts 2003; Schwarz 2013): familiarity would be neither necessary nor sufficient (*the* can refer to an unfamiliar entity as long as it is uniquely identifiable) while unique identifiability is sufficient but not always necessary (*the* can be used when the entity is *believed* to be uniquely identifiable). In Example (3) below, the hearer does not need to know about these theme parks beforehand because the description around the head noun makes the entity identifiable.

(3) *Visitors to <**the**> Disney theme parks near Orlando, Florida, have a wealth of new-build state-of-the-art hotels to choose from […]*

2.3 The learnability of the English article system: the tandem of identifiability and familiarity

Different from work in the logical-philosophical and linguistic traditions, which has focused on referential *the*, studies exploring how mastery of the article system emerges in L1 and L2 have to account for the whole range of uses of the article system. For this purpose, acquisitional approaches have typically relied on a combination of (unique) identifiability and familiarity, and more specifically on a re-interpretation of these concepts as Specificity of the Referent (SR) and Hearer Knowledge (HK). Crucially, these two core concepts have been treated as existing on par. This is also the case in Huebner's (1983, 1985) so-called semantic wheel, which is one of the most widely used models for classifying noun phrase environments in English article acquisition studies. The two properties, HK and SR, can be present or absent, yielding the four basic types shown in Table 1 (cf. Thomas 1989). The types are not intended to give an exhaustive picture of article use in English but capture the major environments relevant to article use. Note that, except for Type 2, all types allow at least two articles; which of the allowed articles will be used is co-determined by the lexical properties of the noun, i.e., singular or plural, mass or count.

Importantly, these four basic types map neatly onto the major categories used in studies from the logical-philosophical and linguistic traditions. As discussed in Section 2.2, in linguistics, referentiality has been key. However, while most attention has gone to the study of definite reference and its interpretation, the categories of indefinite and generic reference likewise exist, as does non-referentiality. The focus of the acquisition literature on how the article system, in all its complexity, is mastered, has ensured a more balanced treatment of the four categories. This more balanced treatment of the article system, alongside the focus on acquisition or learning, makes the variables HK and SR excellent candidates for use in the current study in which learnability plays an important role.

The acquisition of referential systems, and of the article system in particular, has attracted quite some attention in the community of researchers working on first

Table 1: The four major article types, recognized in linguistics, mapped onto the concepts of HK and SR utilized in language acquisition, as summarized in Thomas (1989).

	SR+	**SR**–
HK+	**Type 2** ~ referential definites Possible article(s) *the*	**Type 1** ~ generics Possible article(s) *the, a/an, Ø*
HK–	**Type 3** ~ referential indefinites Possible article(s) *a/an, Ø*	**Type 4** ~ non-referentials Possible article(s) *a/an, Ø*

language acquisition, indeed. Brown (1973) was the first to document the development of article use (among other morphemes) in non-elicited spoken English longitudinally and showed that the number of errors children make is very low, and all involve specific referents: children erroneously used *a* for specific referents that were known to the hearer and *the* for specific referents that were unknown to the hearer. Early experimental studies were aimed at further assessing children's understanding of the presupposedness of referents, i.e., Hearer Knowledge. Work by Maratsos (1976) and Warden (1976) confirmed that the most typical error stems from using the definite article when the referent is specific but not known to the hearer. This data was later reused by Cziko (1986) to test and support Bickerton's (1981, 1984) language bioprogram hypothesis (LBH), which states that children are biologically programmed to make the specific/non-specific distinction (Referent Specificity). This would explain why children would initially only correctly use *the* in cases where the entity is specific and presupposed and *a* in cases where the entity is specific but not presupposed. Yet, data collected by Karmiloff-Smith (1979), who studied French-speaking children between the ages of 3–11, revealed a U-shaped pattern of development, where children start off with relatively accurate usage across the board, then move through a phase with many errors, before returning to accurate usage.[2] The notion of Hearer Knowledge appears to be the one that poses the most conceptual difficulties to learners and is the later of the two to develop, but it does start developing earlier than LBH would predict. For a recent discussion of the relation between Hearer Knowledge and Referent Specificity and how they impose bounds on variability, see Romain et al. (in press).

Yet, different from what we observe in L1, Hearer Knowledge and Referent Specificity do not seem to be sufficient to explain the use of articles to L2 learners whose L1 lacks articles (Ionin 2003; Trenkic 2008 and later work). For example, recent work by Shin and Kim (2017) summarizes that *the* is overgeneralized to contexts where the indefinite article is required; that L2 English learners also overuse *a/an* in definite contexts; and that those L2 English learners whose L1 does not have an article system tend to avoid using articles altogether. This comes as no surprise since few grammars rely on Hearer Knowledge and Referent Specificity (see also Trenkic 2000: 25) and those that do rely on these concepts do not clearly distinguish between the two. For example, the British Council (LearnEnglish) states that *the* should be used when the speaker and hearer know what they are talking about or can identify the referent and *a* when this is not the case. Authoritative grammars such as

2 This gradual development from initial imitation to full productivity was extensively discussed in the literature (Lieven et al. 1997; Pine et al. 2013; Pine and Lieven 1997; Pine and Martindale 1996; Valian 1986; Valian et al. 2009; Yang 2013) and recently supported by large computational simulation of article development based on a dense corpus (Meylan et al. 2017).

Greenbaum (1996) and Biber et al. (1999), instead, rely in the first instance on countability to differentiate between *the* and *a*, and HK takes second place: *the* is used when the noun phrase is assumed to be known to both speaker and hearer. Interestingly, Lyons (1999: 2–3) does make a distinction between *the*, used when the referent is assumed to be clear to both the hearer and speaker and *a*, used for a specific referent which is unknown to the hearer.

To remedy the problem, alternative accounts of article use have developed. Master's (1990) model of the English article system is one such account that has been subjected to classroom-based testing. Instead of relying on (assumed) Hearer Knowledge, known as Definiteness in his models, and Referent Specificity, Master proposed a binary, hierarchical system that collapses these features into one, and contrasts Identification (*the*) with Classification (*a*, zero). In a sense, this approach acknowledges the natural alignment and predominance of [Def+ Spec+] and [Def– Spec–] and promotes the primacy of Definiteness over Specificity for pedagogic purposes; we will return to this point in Section 6.3. While entities that are identified always take *the*, for entities that are classified, Countability and Number play an important role in distinguishing between *a* and zero: countable singular nouns take *a*, while countable plural nouns and non-countable nouns take zero. Within the subgroups, further semantic distinctions are considered that function as rules for article use, such as first mention (*a*) versus subsequent mention (*the*) and defining (*a*) versus limiting (*the*) post-modification.

2.4 From identifiability to accessibility

Cognitive-linguistic accounts of article use highlight one possible reason why explanations fail in L2: although most accounts do focus on Definiteness, they emphasize Reference and Identifiability at the expense of human cognition, and the dynamic relation between speaker and hearer in discourse in particular (Hinenoya and Lyster 2015: 398).

Chafe (1976: 54) insisted on the importance of what the speaker *assumes* the addressee to know or be aware of: the speaker chooses from different 'packaging statuses' "on the basis of his assessment of what the addressee's mind is capable of at the time". Continuing this idea, and expressing it in terms of the storage metaphors for memory which were prevalent at the time, has led to referential systems being likened to file-card systems (DuBois 1980: 210). "For every indefinite, start a new card. For every definite, update an old card." (Heim 1983: 168). It has, indeed, been argued that grammatical reference is not even about reference to an entity that exists in some world but about reference to a memory trace (Ariel 1988, 1994; Epstein 2002; Givón 1992). In this tradition, Ariel (1994) promoted "accessibility" instead of

identifiability as defining feature of a referential system: the definite article marks that the discourse referent is accessible and can be interpreted with the information available in the discourse context. It is the task of the speaker to choose an appropriate referential expression and the task of the hearer to identify the intended mental representation that corresponds to that referring expression (Ariel 1994: 38).

The reference to memory plays a crucial role in the historical development of the article category. König (2018: 169) describes the stages in the development of definite articles. In cataphoric use, identification is possible due to sufficient description. In anaphoric use, identification is supported by the preceding context. The change to recognitional reference or *emploi mémoriel* (4) and then associative reference (5) involves a major change in the type of context required for identification: there is a shift "from an external, situational or textual context to a more abstract context of association, of memorizing or of general availability in a universe of discourse". Associative use, typically the last to appear historically, is often regarded as the crucial step in the development of a definite article, since this use is not available for demonstratives (König 2018: 173–174).

(4) *You remember the restaurant we went to recently. That is where I found a wallet.*

(5) *We laid out the picnic. The coffee was still warm.*

Epstein (2002: 371) takes this idea one step further and, following Fauconnier (1994), casts the accessibility of nouns introduced by *the* in terms of Mental Spaces. Like Givón (1992), he considers the definite article as a mental processing instruction that triggers the processing procedure by which discourse referents are accessed, but the definite article is a marker of *low* accessibility: it signals "that the means for interpreting the NP in which it occurs is available somewhere in the configuration of mental spaces, as long as the appropriate spaces, elements and connections – i.e., the access path – can be constructed by the addressee." On this approach, listing the meanings of *the* becomes futile: the definite article indicates that the referent of the NP should be processed as accessible information. The exact interpretation is not specified by the grammar but determined by the context (Epstein 2002: 368).

As attractive as it may sound, this high-level specification of the function of *the* provides little guidance for the correct use of articles in context; as we saw in Section 2.3, grammatical restrictions associated with the historical origins of the articles still apply. For this reason, we examine a variety of variables proposed in the literature, with specific emphasis on variables used across studies of the development of article usage in L1 and L2: this ties in with our interest in the role learnability plays in the emergence and maintenance of linguistic systems. We aim to determine which dimensions of the linguistic and extra-linguistic experience are important for the

correct usage of articles, whether and how the relevant dimensions interact, and how their interaction affects learning.

3 Data and annotation

To achieve our aim, we take a usage-based approach and study all three articles as they occur in a random sample of discourse chunks extracted from a corpus. Working with a sufficiently large and representative sample of data that was not created for the purpose of linguistic analysis shifts the focus of the investigation, compared to so-called "armchair" approaches. A corpus-based approach provides insights into which usages are frequent and which are less frequent and allows us to capture those tendencies that apply to the bulk of article usage. At the same time, it can only look at what is contained in the sample, which rules out the use of established analytical techniques such as minimal pair analysis and substitution tests. Furthermore, the corpus-linguistic requirement to annotate all the data for the same variables and the statistical requirement to have a sufficient number of examples for each label limits the selection of variables to those that are widely applicable. We will not attempt a detailed semantic classification of any purported semantic differences the articles might convey, in the form of a polysemous radial network or otherwise (for a detailed discussion of the semantics associated with articles see e.g., Biber et al. [1999: 260–263] or Lyons [1999: 157–198]). Instead, we accept that the fine semantic details are determined by context. After all, "language does not carry meaning, it guides it" (Fauconnier 1994: xxii).

The dataset for this study was extracted from the British National Corpus (BNC, Leech 1992). We aimed to annotate samples of equal size (1,000 sentences each) from the Spoken and Written parts of the corpus. Information as to whether the chunk was from the spoken or written part of the corpus was retained and entered for the purpose of classification (see Section 4). Were deleted from the dataset: instances where the target article was wrongly annotated (and no substitute instance could be found in the paragraph), instances that constituted repetition (and if retained, the exact same chunk would be included twice), article errors (mostly due to odd transcriptions in the spoken section of the BNC), and instances where the context required to determine the variable values was missing. Because many more spoken examples than written examples had to be discarded, a further 500 spoken examples were extracted and annotated. This left us with 2,200 instances in total, 963 written and 1,237 spoken. These instances were annotated by the second author, according to the principles described below. Cases that triggered any hesitation were referred to the first author, who independently annotated these instances. Both annotations were then compared. Where annotators initially disagreed, the example was

discussed further until agreement was reached, and the reason for the agreed-upon justification for a particular variable value was recorded.

To respect the discourse nature of the phenomenon, chunks of text were extracted which include the sentence containing the target article and, where available, three sentences before and three sentences after for context. Working with discourse chunks rather than with individual sentences makes it possible to examine to what extent article choices are based on information that is available outside the sentence, as illustrated in Example (6).

(6) *My voice was strident and shrill.*
A startled passer-by gave me as wide a berth as possible.
I ran down the pavement to get as far away from the hotel as I could, then I sat down in a doorway and continued crying.
Slowly <***the***> *tears subsided and gradually I began to pull myself together.*
I had nearly got myself killed back there: what on ø earth did I think I was playing at?
The people back at the hotel didn't know me.
They didn't know I wouldn't steal ø money like that.
[BNC, ID 14]

After piloting, as described in SupMat_1, each target article (1 per discourse chunk) and the noun phrase it is part of were manually annotated for six variables that have played a prominent role in the literature (see Section 2): Hearer Knowledge (HK), Referent Specificity (SR), Number (singular, plural or neutral), Countability (countable vs. uncountable), Elaboration and Set Phrase. The values for Example (6) are given in (7).

(7) Known to Hearer, Specific Referent, Countable, Plural, No elaboration, Not a Set Phrase

The first two variables, HK and SR, capture elements that are part of the discourse situation in which speaker and hearer participate and have played a key role in the acquisition literature. As such they fit with our focus on learnability. The variables Number and Countability are morphological variables that have been used regardless of theoretical focus. Elaboration and Set Phrase are lexical in that they describe properties of the noun phrase. While these variables may appear straightforward, applying them to real data is not always straightforward. Therefore, some comments on annotation are in order. We also refer to SupMat_2 where we discuss a few more challenging examples of HK and SR combinations in more detail.

Hearer Knowledge is based on whether it appears that the speaker assumes that the hearer knows what they are referring to; this assumption is justified if there is reason to assume that the hearer has the required knowledge or if the required

information is available in the context. Note that knowledge, here, does not refer to propositional knowledge but to any knowledge accessible in discourse (cf. Karttunen 1976): hearer knowledge captures the extraction or profiling of presently activated sets of referents, which are created based on discursive, perceptual and conceptual information. The easiest case of HK+ is one where the referent has previously been explicitly mentioned; this can be in a verbatim fashion, but it is also possible to rephrase the initial mention (e.g., *Our neighbors have <a> cat. <The> animal catches mice in our garden.*). Furthermore, HK+ not only applies to a referent that has already been mentioned explicitly in one way or another but also to one that can be considered activated by something in the wider context that has been mentioned, such as the *tears* in Example (6) above which were activated by the mention of *crying*. The extralinguistic context must also be considered here as something that is referred to can be physically present, for example, and therefore HK+. Clues as to whether the hearer can be assumed to know about the referent can also be found in generally shared knowledge (e.g., *the sun*). A related group of referents that traditionally count as known to the hearer are nouns that refer to concepts or general representations of the entity (e.g., *<**ø**> Capitalism will brook no delay* or *I've never played <**ø**> golf*). Hearer Knowledge is considered absent if the hearer cannot reasonably be assumed to already know about or be able to use the context to retrieve the required information about the referent.

Specificity of the Referent is used to distinguish between nouns that refer to a specific referent and those that do not. A noun is marked as specific if a referent is singled out, i.e., if reference is made to one or several thing(s), as in *She is the chair of the interview panel* or *Somebody I know bought a bike* where the referent is a specific bike. A noun is marked as non-specific if its referent is to be considered generic, e.g., because it refers to a concept as a whole (e.g., *golf* in *I like to play golf),* or to all members of a category (e.g., *boys* in *I hate it when ø boys do that*). The label non-specific is also used for so-called non-referential expressions as *I want good shoes for winter* where there is no reference to a specific pair of shoes, nor to the category of *shoes* as a whole.

As will be clear from the explanation, for setting the values of HK, and to a lesser extent of SR, careful examination of the preceding context is of utmost importance: for every article/noun combination, it needs to be determined whether the referent of the noun is present in the preceding context, either directly or indirectly. The article itself cannot be used as an indicator of HK: contextual support of the type described above is required on our approach. This is particularly important when it comes to annotating instances of the zero article: our context-dependent approach enables us to annotate all instances in an identical fashion.

In annotating Number we used the label "singular" for singular nouns and "plural" for plural nouns. In cases where it seems difficult to decide between plural

and singular (e.g., collective nouns such as *staff* or *team*), the decision was made based on grammatical agreement when available in context, e.g., in the example *The police were baffled, police* was marked as plural because of the plural verb form *were*. Nominalized verbs were marked as neutral for number (e.g., *ø ground handling*).

Whether a noun is annotated as countable or not is often dependent on context. While it is true that some nouns are generally considered intrinsically countable (*cat [s]*) or uncountable (*cheese*), language can be modulated to accommodate for non-typical uses (see Michaelis [2004] for a discussion of coercion with respect to countability), e.g., *Can I have three beers please?*. Therefore, annotation was done based on the actual use of the noun in the specific context even if the same noun could be used differently in a different context, e.g., *the oils he uses for cooking* versus *he cooks with olive oil*.

We annotated Elaboration with four values: before (pre-modification), after (post-modification), both or neither. Elaboration includes any type of pre- or post-modifiers such as adjectives, complex determiners, prepositional phrases and relative clauses (among others), e.g., *a sort of ø crystal ball in which I could call up everything I had ever known* where the head noun is preceded by a modifier *sort of* and an adjective, which are both pre-modifiers and is followed by a relative clause, which is a post-modifier.[3]

The decision to annotate an expression as a set phrase was based on a number of factors such as whether an expression was deemed idiomatic (*in the clear*) or so frequently used that it might be considered idiomatic (*this time of ø year*), or simply due to its unusual form (*take ø advantage of*). These features tend to overlap, too, as many idioms have unusual syntax e.g., *many a time*. This variable is potentially the most subjective as it is not an easy task to define what counts as an idiom (see Croft and Cruse 2004: 230). Therefore, we annotated as set phrases expressions that are conventionalized and at least partially unpredictable. We also conducted further searches, both in the BNC and on the Internet, to ascertain whether any (other) articles could be used (e.g., *this time of the year*) without changing the meaning of the chunk (compare here *at a time* vs. *at the time*).

Cross-tabulating the raw data reveals an interesting pattern that does not correspond to the expectations raised by previous accounts of the English article system that put HK and SR on a par. As shown in Table 2, Referent Specificity does not distinguish particularly well between the articles *a* and *the* but Hearer Knowledge

3 Of course, even four options simplify the situation: a pre-modifier could be a pre-determiner, as *local* is in Example (1), but could also be an adjective that does not function as pre-determiner; only a pre-determiner licenses the use of *the*. Zhao and MacWhinney (2018) mention that intervening adjectives make it more difficult for L2 learners to select the right article, and they attribute this to the increased distance between article and noun; it may, however, well be due to the intricate role the adjective plays in article selection.

Table 2: Distribution of Hearer Knowledge and Referent Specificity over articles in the sample.

HK	HK+	HK–	SR	SR+	SR–	Total
zero	151	449		54	449	503
a	3	722		153	572	725
the	938	34		710	262	972
Total	1,092	1,108		917	1,283	2,200

achieves an excellent split: there are only 3 instances in the raw data that allow *a* when the referent is (presented as) known to the hearer, and 34 instances of *the* when the referent is (presented as) unknown to the hearer. Hearer Knowledge, however, does not give a clear picture for the zero article: although the zero article is predominantly used in contexts where the referent is (presented as) unknown to the hearer (in 74.8 % of all instances), it is Referent Specificity that more accurately predicts the use of the zero article with 89.2 % of zero cases being SR. In other words, the raw data already reveals that different articles may not be learned from the same variables; we will come back to this observation in Section 5.

Before proceeding to the analysis of the data, recall that Table 1 above summarizes how the different types of referentiality, typically relied upon in philosophical-linguistic approaches to definiteness, map onto the concepts of Hearer Knowledge and Specificity of the Referent that are prevalent in acquisition research and will be used in this study, too, due to its focus on learnability. Taken together, Hearer Knowledge and Specificity of Referent distinguish four main types of reference: the referential definites (HK+, SR+), the generics (HK+ SR–), the referential indefinites (HK–, SR+) and non-referential uses (HK–, SR–). Our data was additionally annotated for these variables using the mapping outlined, and all analyses were run using both annotations; the Type-based results are presented in SupMat.

4 Classifying authentic usage data

Our dataset was manually annotated for information on the five different variables discussed above: Hearer Knowledge, Referent Specificity, Number, Countability, Elaboration and Corpus. A sixth variable, Corpus (written vs. spoken) was included to allow for different functioning of the article system in written versus spoken language. Different combinations of the first five variables have been proposed to account for article usage, many of which treat HK and SR as equally important and applicable to the entire system, with Count and Number accounting for subparts of the system only (see Section 2). In order to determine whether and how these

variables work together to decide whether the article used in a given context will be zero, *a* or *the* and which variable(s), if any, should be promoted to occupy the dominant position in a potential hierarchy, we run a tree and forest model.

Tree and forest models are a type of statistical classifier that relies on recursive partitioning to achieve optimal classification accuracy. A tree model discards non-significant predictors automatically and naturally allows for interactions. Its final result provides a procedure for deciding which article will be used in a sentence. Because a single tree is likely to overfit the data, a classification forest is constructed as a means to validate the proposed tree model. A forest relies on bootstrap samples, that is, samples of size *N* drawn with replacement from the original dataset with *N* observations. Using the R randomForest package (Liaw and Wiener 2018) and the party package (Hothorn et al. 2006; Zeileis et al. 2008), both a classification forest and a classification tree were constructed, with the forest grown from 500 random samples and the number of variables to consider at each split set to 2. A classification tree is presented in Section 4.1, while the random forest and the train/test validation results are provided in SupMat_3. The models presented here are run on a dataset that excludes the 23 instances that were tagged as not being marked for number, and the 419 instances that were tagged as a set phrase, leaving 1,768 instances; for completeness, classification models for the full data set and for the set phrases only are presented in SupMat_4.

4.1 A classification tree

The classification tree for the article data, based on all variables presented in Section 2 above,[4] is presented in Figure 1. Each split in the tree is labeled with a decision rule. Before determining the rule at each node, the algorithm inspects several predictors and selects the one that is most useful; in this case, the best prediction results were obtained if two variables were considered at each split. The algorithm does not look ahead, however, and cannot consider decisions that would yield a slightly worse split locally but would do significantly better globally. This is known as a greedy approach.

The first node (in the oval) in this tree is based on Hearer Knowledge. The accompanying *p*-value indicates that articles are well separable if we know whether the item referred to was or was not (presented as) known to the hearer. Hearer Knowledge thus plays a crucial role in distinguishing between the articles: Hearer Knowledge is found highest up in the tree, executing the first split.

4 We also provide a tree & forest model for the same data, classified according to the cells of Table 1 which represent referential definites, referential indefinites, generics and non-referentials. For details we refer to SupMat_5.

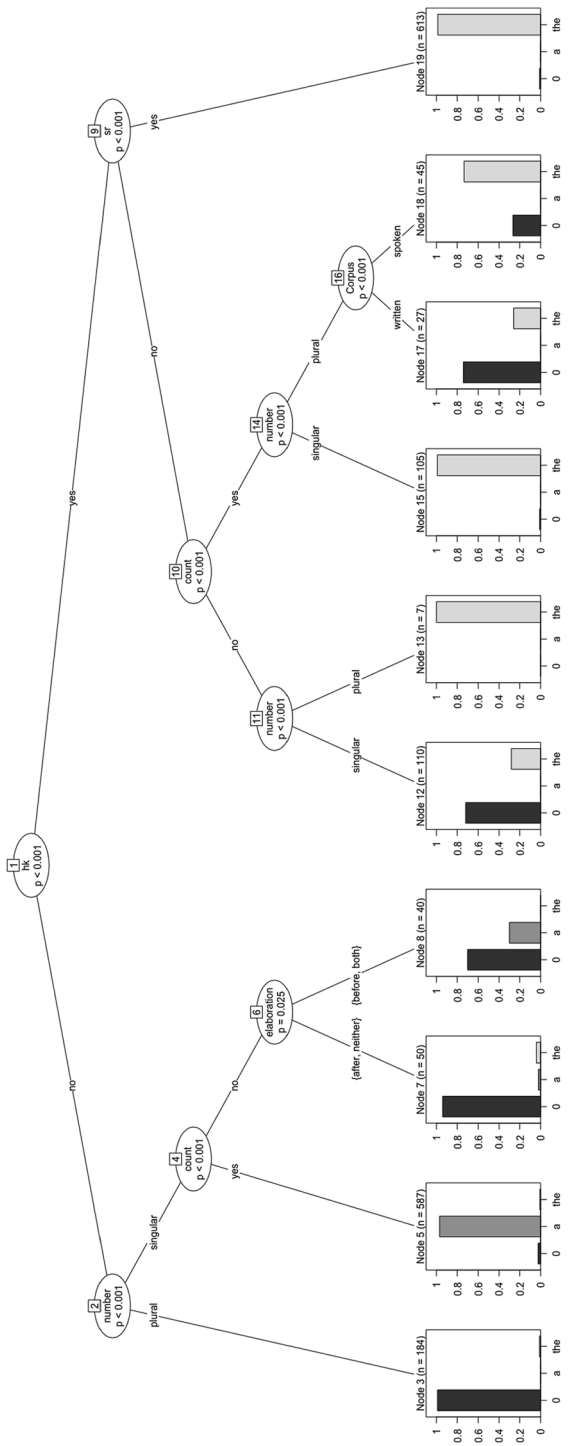

Figure 1: Classification tree based on manually annotated discourse chunks extracted from the BNC, excluding set phrases and nouns that have no number.

Table 3: Classification accuracy using one tree.

	zero	a	the	Total
zero	**356**	13	42	411
a	14	**569**	4	587
the	19	2	**749**	770
Total	389	584	795	**1768**

Highest values in boldface.

The search for the locally best-performing splitting criterion is now repeated for the remainder of the data. At each next branch, a new decision rule is presented that directs us further down the tree past increasingly purer nodes. In cases where Hearer Knowledge is not assumed, the property that comes into play next for splitting the remainder of the data as best as possible is Number. If the noun is plural, we are led to a terminal leaf node of 184 items where *zero* is the article of choice, with few exceptions. But if the noun is singular, further properties come into play. If the singular noun is a count noun, we are led to another terminal node of 587 items where *a* is the article of choice, with few exceptions. If the singular noun is uncountable, elaboration plays a role (albeit with $p = 0.025$ this split is the least reliable of all splits): in the 50 cases where the elaboration follows the noun or is absent, zero is the preferred choice; in the 40 cases where the elaboration precedes or both precedes and follows the noun, the choice is divided between zero and *a*, in a 70/30 split.

The other side of the tree captures the situation in which Hearer Knowledge is assumed. If a noun is (presented as) known to the hearer, Referent Specificity plays an important role in further categorizing the data. If the noun has a specific referent, we are immediately led to a terminal leaf node (613 instances) for which the article of choice is *the*. If the referent of the noun is not specific, however, we are led down a more complicated path via Countability and Number: if the noun is uncountable and plural, we arrive at a terminal node (7 instances) with *the* as the preferred choice, but if the noun is uncountable and singular, only 30 % of cases get *the*, and zero is preferred. If the noun is countable and singular (105 instances), *the* is strongly preferred, whereas if the noun is plural (72 instances), Corpus plays a role: in the Written corpus, about 80 % of cases are marked with zero, while the other 20 % gets *the*, while in the Spoken corpus, the proportions are inverted. We will come back to this observation in Section 4.2.

All in all, the tree correctly classifies 94.6 % of all instances, scoring 356/411 for zero articles, 569/587 for *a* articles and 749/770 for *the* articles (Table 3). This is three times as good as randomly choosing, which would yield 1/3 correct.

Moreover, in our tree-based model all three articles are most often predicted correctly, i.e., the highest values are on the diagonal in the table, signaling that the classification accuracy is good for all three articles. That being said, the zero article is slightly less well predicted than the two other articles. Yet, incorrect classifications are also revealing. Instances of zero that are incorrectly predicted are most often predicted as *the*, while incorrect predictions of *the* and *a* are typically predicted as zero.

A single tree is likely to overfit the data, however; growing a forest based on resampling mitigates against this risk. On our data, a random forest of 500 trees reaches a very similar correct prediction rate of 94.57 %. These results are outstanding, with only 13.4 % of all zero instances (55/411), 2.7 % of all *the* instances (21/770) and 3.1 % of all *a* instances incorrectly classified (18/587). Details are provided in SupMat_3.

On the basis of the forest, the importance of each variable can be calculated. Using random permutation of the labels of each variable, the relative importance of the different predictors for the classification accuracy of the model is assessed. The result is shown in Figure 2 where the left panel reveals that Hearer Knowledge is the strongest predictor by far:[5] if removed, there is an 81.8 mean decrease in prediction accuracy (i.e., the mean difference of error rates in standard units). HK is followed by Number, the removal of which would trigger a 65.9 decrease in prediction accuracy, and by Count which would trigger a 51.5 decrease. Removing Referent Specificity reduces prediction accuracy by 26 only, while removing Elaboration and Corpus reduces the prediction accuracy by 10.6 and 11.2 respectively. In other words, Elaboration and Corpus are "fine-tuning variables" (Divjak 2015) and taken on their own do not contribute much to the correct classification of articles in use. For a discussion of the Gini values, displayed in the right panel, see SupMat_3.

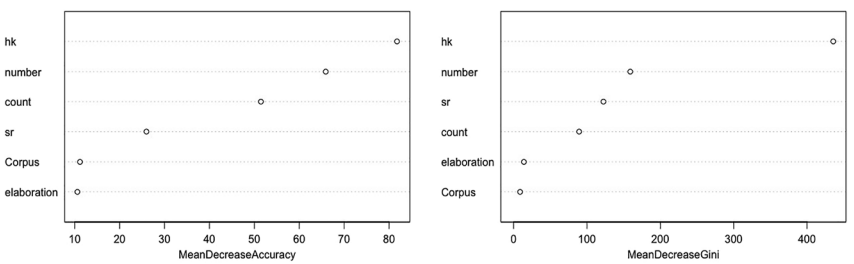

Figure 2: Mean decrease in classification accuracy per removed variable.

5 Note that the values will change insignificantly with every run, due to sampling.

4.2 Interpretation

Overall, the picture that emerges from the classification exercise reveals that the situation becomes simple when we move away from referentiality and let Hearer Knowledge occupy the dominant position. The fact that Hearer Knowledge comes out as the single strongest predictor by far and that Referent Specificity turns out to be a rather unreliable predictor is unexpected given the literature on article usage, surveyed in Section 1. Although accounts of articles tend to mention the hearer, the hearer is not typically their primary concern. Rather, work across philosophy and linguistics has attempted to pin down the meaning of the concept "definiteness" within the framework of referentiality; this has led "definiteness" to be interpreted as (uniquely) identifiable, which is easily re-interpreted in terms of specificity.

If the referent is (presented as) known to the hearer a choice needs to be made between *the* (787 cases) and zero (118 cases), with very little option for *a* (2 cases). In other words, *the* is the default in the case of HK+, as in (9a). For zero to be used if the referent is (presented as) known to the hearer, the referent cannot be specific (i.e., it needs to be SR–), and the noun needs to be either uncountable singular (9b) or countable plural (9c).

(9) a. *The tour finished in the kitchen and she started to busy herself around* <***the***> *cooker.*
b. *Moreover,* <***ø***> *death is a time of great stress to those you love most.*
c. *You shouldn't hit* <ø> *boys.*

If the referent is (presented as) unknown to the hearer a choice needs to be made between *a* (582 cases) and zero (271 cases), with no option for *the* (8 cases). In this case, *a* can be considered the default (10a). For zero to be used if the referent is (presented as) unknown to the hearer, the noun needs to be either uncountable singular (10b) or plural (10c). This restriction follows straightforwardly from *a* having developed from the numeral *one*: *one* can only be used with singular, countable nouns.

(10) a. *Tulip wafer cups can hold* <***a***> *single scoop of ice cream or desserts such as fruit and cream.*
b. *The final stages of training amount to the handing over of all the responsibility for* <***ø***> *safety to the student.*
c. *I started to read* <***ø***> *newspapers and magazines more, and, for just about the first time in my life, I began to take an interest in current affairs.*

In other words, Hearer Knowledge is a very powerful predictor of article use and the default for either of its values (presence or absence of hearer knowledge) is overridden in the same way, i.e., by the noun being singular uncountable or plural.

On the HK+ side, within SR–, countable plural nouns prefer *zero* in written language, as in (11a) but *the* in spoken language, as in (11b). These could be considered cases of weak definites (cf. Aguilar-Guevara and Zwarts [2010] for a discussion of plural weak definites), where the definite article is used even though the referents are not uniquely identifiable. In spoken language, referents may be more grounded in discourse and therefore more likely to be paired with the definite article.

(11) a. *As today, <**ø**> drinking establishments held an important social and cultural position in Medieval and Early Modern England.*
b. *Generally, if you 're at home for <**the**> bins and all and I fucking Ah but there's no bins now I mean with the wheelie bins they're the g best thing ever, stops rats, dog everything.*

Also in the case of SR–, *the* is occasionally retained with uncountable singular nouns, e.g., (12) where the noun phrase does not signal a specific referent but *the* is licensed by the assumption that mention of wetland plants implies the presence of moisture. In these cases, the referent is not specific per se, but because the referent's existence can be inferred from the context or because the noun phrase contains some form of elaboration, as in (13), it is considered as known to the hearer and thus may take *the*.

(12) *Wetland plants will be in their element, so long as they are given generous mulches to keep <**the**> moisture in.*

(13) *One puzzling decision for a critic to make is to decide whether to isolate any one artist as <**the**> leading figure of the group.*

In the case of HK–, the presence of elaboration can make *a* acceptable in the case of singular, non-count nouns, as in (14), (15) and (16).

(14) *He found in Byers <**a**> bracing skepticism like his own.*

(15) *Amanda filled the flat with so much energy it hurt: humming, calling out snatches of news from the radio, crunching toast at <**a**> painful volume.*

(16) *I know members will be sorry to hear that Mr. Robert and Mr. Ken are both unwell and will wish me to convey to them the council's best wishes for <**a**> speedy recovery.*

In addition to showing which variables contribute more or less (see Figure 2), the classification also reveals that not all articles are governed by the same variables and that not all variables occupy the same place in the hierarchy for all articles. If the entity is presented as known to the hearer, Referent Specificity is more important than Countability, which in turn is more important than Number. If the entity is presented as unknown to the hearer, however, we see the variables reversed in

terms of importance: Number is more important than Countability, while Referent Specificity plays no role.

In sum, our classification approach has revealed major differences in the importance of various available indicators for achieving success in selecting an article for use given the context. It has enabled us to turn that information into a decision tree which helps select the most likely outcome, given properties of the context. Arguably, this approach, which is rooted in probability calculations run post-hoc, once the body of observations is complete, is not how L1 users navigate language. In Section 5, we introduce a different kind of algorithm, rooted in human cognition, to gain an understanding of how dimensions of contexts of use might become associated with an article incrementally, as experience accumulates.

5 Learning a complex system from authentic data

Crucial for a cognitive account of any grammatical category is the answer to the question of whether a proposed model offers a plausible explanation from a use(r) perspective (cf. Master 1990: 466). This entails or invokes a psychological perspective. The question of how the observed complexity can emerge and be maintained is ultimately a question about learnability: the more complex and the less systematic the relationships are between input cue(s) and outcome(s), the harder the learning challenge. In this section, we will model whether human cognition, cast in terms of a simple, incremental process of error-driven learning, becomes sensitive to the relationships described.

5.1 Error-correction learning

Classification trees and forests may well reveal the complexity and (un)systematicity inherent in the system, but they do not represent a viable account of the hypothesized *process* by which articles are selected for use. Classifiers do not learn from data the way humans do. Hence the question remains: can real language users learn to master the English article system using the variables we tested in Section 4, and do they assign the same relative importance to each of these variables as the classifier suggested? Learning English articles from a set of simple binary opposites (e.g., known/unknown to hearer, specific/non-specific referent, singular/plural, etc.), generated from a sample of sentences attested in the BNC, represents an interesting language modeling challenge. First and foremost, the statistical classifier suggested that the relationship between cues (variables) and outcomes (articles) is only partly systematic at best. In other words, the set of cues does not map straightforwardly

onto the three article outcomes, as the cues differ in both their general importance and in their specificity – in being informative about one or more possible article outcomes. Our computational model will give us the opportunity to observe what gets learned as the dynamics of error correction and cue competition unfold.

To shed light on the mechanisms that drive the learning and maintenance of the English article system, we ran a computational simulation study using an error-driven learning algorithm, the Widrow-Hoff rule (WH: Widrow and Hoff 1960). This rule is, in essence, identical to the one introduced in psychology by Rescorla and Wagner (RW: Rescorla and Wagner 1972; Rescorla 2008). In a nutshell, the rule defines how an organism learns from its own errors in order to adapt to the task at hand. More specifically, the rule learns incrementally, i.e., on an event-by-event basis, to associate the presence or absence of an outcome (here, one of the three articles) with that of the presence of a cue (in this case, one of the variable values, for example, HK+). The rule is incremental (i.e., event-by-event) as it re-estimates connection strengths or weights from each cue to each outcome on each trial, which for us is after each (annotated) sentence. If, over learning trials, a given cue is consistently present when an outcome is present, their connection is strengthened, but if a given cue is repeatedly present when the outcome is absent, the weight on the connection between them is weakened. Since the weights are updated as experience accumulates, over time, some cues become indicative of an outcome, while many become irrelevant; as experience accrues, any systematicity in cue-outcome (re) appearances becomes apparent and it is those systematicities that are learned. The overall support that an outcome gets from the cues, its activation, is the sum of the weights on the connection between those cues and the outcome.

The dynamic re-estimation ensures minimal error against the backdrop of all prior experience, including so-called positive and negative evidence, i.e., evidence of the presence of a particular outcome given a specific cue, but also evidence of the absence of an outcome that could have occurred or would have been expected given a specific cue. In other words, learning continually evolves and is driven by the discrepancy between the current "best guess" and the true outcome. A mismatch between the best guess and the true outcome causes weights to weaken, while matches between the best guess and the true outcome encourage weights to strengthen. What drives the process of learning is the competition between cues to carry weight for matching an outcome. Crucially, cue competition happens organically as cue-outcome contingencies are typically imperfect (see Tolman and Brunswik [1935] for the original proposal, and Romain et al. [2022] for a recent application), especially in language which appears to be the perfect opposite of simple one-to-one, univocal mappings. This constitutes a rich basis for error-driven learning (cf. Baayen et al. 2011; Chen et al. 2008; Divjak et al. 2021; Ellis 2006a, 2006b; Haykin 1999; Milin et al. 2017a, 2017b; Ramscar et al. 2010).

5.2 Computational simulation

For our learning simulation we used the same data as for the construction of the classification tree and forest, described above in Section 3. There were 1,768 events that could take one of three articles, *a, the* or zero. Each event consisted of specific values for the five variables Hearer Knowledge (yes/no), Referent Specificity (yes/no), Number (singular/plural), Countability (yes/no) and Elaboration (before/after/both/none). For example, the boldfaced sentence in (17) was encoded as indicated in (18), where the article represents the outcome, and the variable values are the cues.

(17) *However, ø party guidelines circulating in East Berlin forbade any attacks on the socialist state or the leading role of the Communist party. They also made it clear that the party fears that New Forum and ø other opposition groups could turn into ø mass movements. New Forum tried to take the Politburo up on its offer to talk, and again demanded ø legal status so that it would not constantly be accused of ø subversion.*
Erich Honecker, the party leader, spoke about the country's problems for the first time in <an> address to ø party leaders but simply echoed the Politburo statement.
He flatly rejected ø talks with the opposition, saying: 'We don't need ø suggestions for the improvement of ø socialism that are really intended to cause its demise.' It is not known how many of the people arrested in last Saturday's protests on the night of East Germany's 40th anniversary have been freed. About 150 are estimated to be still in ø jail.
[BNC 1932]

(18) unknown_specific_countable_singular_elabAfter_A

For the current learning setup there are some specifics that need to be mentioned. First, in addition to the 12 cues we have assumed that the type of language, spoken versus written, constitutes a context cue or constant learning background; this essentially means that everything else in the context of learning is considered an undifferentiated whole (i.e., equally informative or uninformative). A constant background is informative about the relative frequency of an outcome (e.g., how many times did an outcome occur overall) as well as about the number of trials on which cue and outcome are not coupled (see background rate: Rescorla 1968). In the present study, the background was implemented as an additional cue that was present on all trials.[6] Second, the training assumed the learning rate parameter

6 Inclusion of a background appears more feasible in a small-scale computational simulation such as the one presented below, based on carefully annotated data, where it is rather straightforward to

γ = 0.01 which is set in advance and kept constant; this is a commonly used value, small enough to guarantee incremental learning (cf. Enquist et al. 2016; Rescorla and Wagner 1972). Slow or incremental learning is expected to be rather typical for animals and humans, although rapid and/or dramatic changes are known to be possible too. Somewhat simplified, smaller learning trusts accumulated experience that is summarized in the connection weights, while high (er) rates of learning would allow more pronounced changes in those weights (this is tightly related to the Principle of Minimal Disturbance; cf. Widrow and Lehr 1990; Chen et al. 2008; Milin et al. 2020). Third, the order in which examples are presented is important in WH learning. In the current simulation, the order of events was randomized: although the model is eminently sensitive to the order of presentation, we do not assume that anyone would see these examples (which constitute a random sample drawn from the BNC) in the order in which they occur in our dataset.

5.3 Interpretation

The learning process is visualized in Figure 3 which displays how the grammatical category of English articles would be learned from exposure to naturalistic data, enriched with those properties that have been proposed in the literature as governing article usage. Each of the panels represents an outcome, zero (left panel), *a* (middle panel) or *the* (right panel). The variables are contrastively color-coded, with all values per variable encoded in different shades of the same color, e.g., Known to Hearer is shown in light blue while Unknown to Hearer is pictured in dark blue. Cues above the zero line are positively associated with the outcome, while cues below the zero line are negatively associated (dissociated from) with the outcome.

Across the three articles, we observe different types of learnability dynamics. The exact values of the weights, registered at each quartile of training, are given in SupMat_6.[7]

The article *the* shows a specific learning profile: one cue stands out as strongly associated (Known to Hearer) at 0.44 and another one as strongly dissociated (Unknown to Hearer) at −0.33. The second strongest positively associated cue, Countability, stands at 0.15 having started off at 0.12, followed by Specificity of the

draw inferences about how informative 'everything else' (i.e., the background) is for an outcome (compare here, for example, the use of a constant background in Spellman 1996 and the simulation of this study in Danks 2003). The inclusion of a background in large-scale computational training is less straightforward. In principle, one could argue for a variety of backgrounds, which can be more or less general and more or less stable.

7 Here too, the values will change insignificantly with every run, due to randomisation.

Figure 3: Computational simulation of learning the English article system from a sample of authentic sentences with an incremental, error-correction learning algorithm. Left panel: learning of the outcome zero; middle panel: learning of the outcome *a*; right panel: learning of the outcome *the*.

Referent, which starts off strong at 0.19 but ends up at 0.13; the opposite values, Non-Specific and Uncountable, show the same pattern at the negative end. The trend lines for most other variable values are parallel and remain flat throughout the training.

The article *a* shows a different profile: three cues are relatively strongly associated with *a* and their opposites are relatively strongly dissociated from *a*. These are, in order: Unknown to Hearer (0.34, and −0.28 for Known to Hearer), Singular (0.25, and −0.19 for Plural) and Countable (0.19, and −0.13 for Uncountable). The remaining cues show some association or dissociation initially but return to zero halfway through the training, i.e., they become irrelevant.

The learning profile for the zero article resembles the learning profile for *a* in that more than one cue stands out from the rest; these are Uncountability (0.35) and Plurality (0.32), followed by Non-Specificity of Referent and Unknown to Hearer (both around 0.17). Here too, some of their opposites are negatively associated with *a*; these are Countable (−0.15) and Singular (−0.12). Different from *a*, many cues develop a weakly positive association with the zero outcome (values ranging from 0 to 0.10). Interestingly, the two properties zero is best learned from, Plurality and Uncountability, are not exclusive to zero: they are strongly negatively associated with *a* and weakly positively or negatively associated with *the*.

In other words, each article has a different learning profile. While *the* is characterized by one strong cue for and against, *a* relies on three cues for and against, while zero adds a large number of weakly positively associated cues into the mix. Moreover, the learning simulation confirms that the relevant dimensions of experience intersect: not all articles are learned from the same properties, and not all properties are equally important for all articles. For *the*, learning is dominated by Hearer Knowledge and being known to the hearer is the strongest cue for *the*. This variable is also important for the article *a*, but here being unknown to the hearer is the requirement. In addition, Number (singular) and Countability (countable) play important roles for *a* too. The tree & forest model suggested that the zero article competes with both *a* and *the*. It shares some core properties with *a*, but requires different values: Number (plural) and Countability (uncountable). The same is the case for its relation with *the* with which it shares its dependence on Referential Specificity but zero requires non-specificity rather than specificity.

In drawing parallels between the results from a well-known statistical classifier such as the tree and forest model and a simple error-correction model it is interesting to note that they yield highly comparable results. The fact that the hierarchy of variables, reflecting their importance, is virtually identical across both models speaks to the reliability of our learning algorithm. However, in our learning

simulations, the hierarchy emerges organically as iterative error correction proceeds: quite rapidly, Widrow-Hoff (Rescorla-Wagner) develops a strong taste for associating hearer knowledge (HK+) with the definite article; less rapidly, hearer knowledge (HK–) gets associated with the indefinite article and number (singular) and countability (countable) with the zero article, while some cue-outcome associations appear promising initially but later on lose their appeal (such as SR for *the*), and others never really develop. These insights into the process are added by incremental learning algorithms, while classification models highlight the end result. Furthermore, and different from a statistical classifier, our error-correction simulations reveal what happens for each individual article: rather than averaging the importance of a variable for all outcomes, our learning algorithm lets us observe how individual variable values relate to individual articles. This resolves the challenges that (multi)-collinearity so often poses in linguistics since language is redundant by design. Widrow-Hoff and related algorithms are particularly resilient to these issues (for details see Widrow 1959; Widrow and Hoff 1960). In addition, the learning algorithm shows how the relationship between a variable value and an article develops over time, as more examples are encountered. The opportunity to observe the development of a cue-outcome association and the speed or rate with which it develops is a distinct advantage.

6 General discussion

In this study, we took a usage-based approach and tracked articles as used in a representative random sample of written and spoken language, extracted from the BNC. This sample of real data, which excluded fixed expressions, included preceding and following context that was used to annotate each instance for the five key properties that had previously been identified in the literature as capturing the core dimensions of article use, i.e., Hearer Knowledge, Referent Specificity, Number, Countability and Elaboration. Annotation was carried out by taking a cognitive perspective on reference that moves away from "the strictly referential assumption" and allows articles to play a role in the dynamic relationship between the minds of speaker and hearer, as it is shaped by and in discourse (Epstein 2002: 335). This step change in approach provides the conditions for the idea we propose, i.e., that of articles functioning as a hierarchically structured linguistic indexing system, centered on the memory of the hearer, all the while obeying the grammatical restrictions imposed by their historical origins.

6.1 Articles as a hierarchically structured, learnable grammatical system

Based on the five dimensions, Hearer Knowledge, Referent Specificity, Number, Countability and Elaboration, a statistical classifier achieved a correct prediction rate of over 94 %. Crucially, to achieve this result, the core dimensions need to be seen as hierarchically ordered rather than being of equal importance and Hearer Knowledge needs to be promoted to first position in that hierarchy: Hearer Knowledge is the variable that, on its own, achieves the highest classification accuracy. It distinguishes reliably between *the* which is the default in the case of HK+, and *a* which is the default in the case of HK–. To understand how these defaults are overridden to make way for the zero article, other variables need to be brought into play, i.e., Referent Specificity, Countability and Number. These variables hold different degrees of importance depending on whether hearer knowledge is assumed or not: in the case of HK+, the next important variable is Referent Specificity, followed by Countability and Number. But in the case of HK–, the order is reversed. Interestingly, the historical emergence of the article system remains visible: the default for HK– is *a*, but *a* requires singular countables; other cases are handled by zero. Elaboration plays only a minor, fine-tuning role in our model.

Simulations with a computational algorithm that implements how humans learn from data show how the English article system would be learned using these key variables or cues. The simulations reveal the intricate relationships between the dimensions that govern usage of the different articles and point in the direction of different profiles: different subsets of variables and different variables values act as positive or negative cues for different articles. While the use of *the* is largely determined by Hearer Knowledge, the use of *a* hinges additionally on referent non-specificity and entity countability. The zero article, however, requires a different subset of variables, i.e., plurality, uncountability and non-specificity of reference. Crucially, not all outcomes are learned from equally good cues: while *the* is characterized by one strong cue for and against, *a* relies on three cues for and against, while zero additionally adds a large number of weakly associated cues into the mix. Interestingly, both the classification and simulation analyses show that usage of the zero article is the hardest to capture. This is unsurprising given its historical origins: the zero article does not stem from an older, specific form but instead appears to have collected any contexts *a* and *the* could not lay claim to.

Based on these findings, we argue for the grammatical category of articles to be considered a hierarchically structured, hearer-focused communicative device: the category presents itself as a grammaticalized referent tracking system that catalogs knowledge for the benefit of the hearer. This is not all surprising since, as argued by

Hopper and Traugott (2003: 231), grammaticalization is motivated by the interaction between speaker and hearer where the speaker aims to make communication as informative as possible for the hearer (cf. also Sommerer 2018: 258). Fascinatingly, the historical origins of the English articles shine through in this hierarchy: the grammatical constraints appear lower down the hierarchy, supporting the now dominant discourse functions, the further the articles have grammaticalized, while the unmarked article soaks up what does not fit with the historical origins of *a/the*. We will unpack this proposal in the following sections.

6.2 A hearer-focused referent tracking system

It has been suggested that, in Germanic, articles started to emerge in response to the erosion of the case system (Leiss 2000), which offers different, and more limited, ways of expressing definiteness. Yet, the article system in English, as we currently know it, differs from the case systems we still observe in the Slavonic languages: while cases encode relations between participants in an event, articles (now) appear to encode the relation between the information shared by participants in a discourse situation. The system that falls out from our data is guided by Hearer Knowledge: each value of the HK variable has a default value, with HK+ characterized by the use of *the* and HK– typified by the presence of *a*. Each default value competes with zero, and any overrides of the default rely on specific values of Referent Specificity, Countability and Number.

Rather than seeing the use of the definite article as speaker centered (Epstein 2002: 371), we see the article category in its entirety as a hearer-focused system, powered by the speaker. Being a referential grammatical category, the article is obligatorily expressed with every noun and allows the referents of nouns to be tracked in discourse. According to our data, it is the speaker's task to mark the status of the referent of each noun they use for the hearer: for every noun, the speaker reminds the hearer of whether or not the speaker assumes the referent of that noun to be known to the hearer.

Reference tracking, a procedure whereby languages allow participants to be tracked from clause to clause, is well-attested across languages (Comrie 1999; Foley and Van Valin 1984, 1985, 1985; Haiman and Munro 1983; Stirling 1993), yet has mostly been studied for languages with predominantly oral traditions. Often, the reference marker is a free or bound morpheme. An example is possessive suffixes that are used as subject trackers in some Siberian dialects: in spontaneous narratives, the referents can have been mentioned 6–7 speech turns earlier, which translates into 10–15 s

of memory time (Florian Siegl, personal communication, January 2017). These numbers are reminiscent of known limits on working memory. Despite the controversy, there is agreement that working memory is a short time memory system with limited capacity: information is held for somewhere between 15 and 30 s unless it is continually refreshed (rehearsed). Information is lost due to interference, i.e., new information that enters working memory displaces the information that was present. By renewing any relevant short-term memory representations, the speaker reduces the processing burden posed on the hearer.

A grammatical article category that marks referent status obligatorily would also facilitate a more efficient memory search. Recall that DuBois (1980: 168) and Heim (1983: 210) talked about "filing cards" with indefinites requiring "a new card", and definites requiring updating of "an old card". Givón (1992) proposed that the grammar of referential coherence is about retracing information from episodic memory. The article system could indeed be seen as a signal to the hearer to retrieve information from memory and more specifically, to specify whether they should retrieve the entity from semantic or episodic memory. Semantic and episodic memory are two types of long-term memory systems that differ according to the information type they hold: semantic memory stores general facts while episodic memory contains memory for personal facts. If the entity is marked as known to the hearer, the hearer knows they have experience with that particular entity and can rely on episodic memory for more detailed information. If the entity is marked as unknown to the hearer, the hearer knows they can only retrieve a generic placeholder from semantic memory, that is, a noun phrase that refers to a concept or a general representation of the entity.

Our proposal shares some similarities with earlier proposals that have put discourse central. Epstein (2002: 371) sees the use of the definite article as speaker centered in that the speaker determines how the hearer should interpret the upcoming NP: *the* is used to "guide addressees in establishing mental spaces and appropriate connections between the elements in those spaces". However, our data suggest that this account overestimates the power the speaker has. In our sample, HK (and SR) were annotated with reference to the context, i.e., the values of these variables were checked against the preceding discourse. We found very few "mismatched" occurrences, where *the* is used although the entity is clearly unknown to the hearer (8 cases) or where assumed hearer knowledge gives rise to the usage of *a* (2 cases). It appears rather difficult to override the expectations raised by the context, i.e., to choose an article to signal that a nominal referent needs to be interpreted as known or unknown to the hearer if the context does not support this interpretation.

This redundancy of encoding also provides an explanation for the finding that article errors rarely cause communication problems (Master 1997: 216). A series of

studies conducted between the 1970s and 1990s explored the gravity (or severity) of various types of grammatical errors (e.g., Ensz 1982; Johnson and Jenks 1994; Magnan 1982; Vann et al. 1984). Strikingly, article errors were perceived as among the least irritating, in both French and English, and across spoken and written forms of the language. During reading, too, the article is frequently skipped, arguably because it is short, frequent and predictable (Angele and Rayner 2013: 649). In fact, articles are skipped 16 % more often than other equally short words (Drieghe et al. 2008) even when they are not predictable from the context, i.e., when they are syntactically unlicensed (Gautier et al. 2000) and regardless of whether they are used felicitously or infelicitously (Angele and Rayner 2013: 654, 656).

It should be noted that even though Hearer Knowledge is crucial in the choice of article and typically encoded in the context, in some cases the context allows both options and the choice of the article indicates the assumptions the speaker makes. Recall Example (1) and compare the use of *the* in (19) with the use of *a* in (20). In (19), *the* indicates that the speaker assumes that the addressee will assume or know both that there is a local estate agency and which one is being referred to (HK+, SR+). If we were to swap *the* for *a*, as in (20) the assumption is different, with the agency introduced as a new referent, potentially one of many, in the discourse.

(19) *As manager of <**the**> local estate agency and building society, handsome Henry got around.*

(20) *As manager of <a> local estate agency and building society, handsome Henry got around.*

Romain et al. (in press) present results from a 3-alternative forced choice task with 181 L1 speakers of English to determine the bounds on article variability. Their analysis of the properties of the contexts in which alternative construals are allowed versus inhibited reveals that only contexts classified as SR+ restrict construal, with most respondents selecting the same article for usage if the referent is specific.

6.3 Learning to master a hearer-focused referent tracking system

The rebranding of the article category as a hearer-focused system of referent tracking highlights that the challenge of mastering the system not only lies in detecting the relevant dimensions for the use of articles in input; it also requires a specific mindset on the part of the speaker. In essence, using articles correctly requires the speaker to learn to take the hearer's perspective, to track a referent for their hearer from that perspective, and, at every mention, to express whether that

referent has previously been mentioned in or can be inferred from discourse. Which cognitive demands does such a system pose on the speaker?

L1 children have been found to overgeneralize the definite article for use in contexts that cannot be assumed known to the hearer (see Cziko 1986 for an overview). This tendency has been interpreted as evidence for the assumption that children would initially associate *the* with Specificity of Referent, and stronger even, for the hypothesis that Specificity would be the primary concept, with children long unaware that the article system has anything to do with the hearer (Cziko 1986: 896). In fact, generative linguists have taken the issues children experience with HK, or Presupposedness as they label it, as evidence for the innateness of the category of specificity. It has been hypothesized that the ability to take an interlocutor's perspective would need cognitive maturation, out of the Piagetian egocentric phase, a phase of "non-differentiation between one's own and other possible points of view" (Piaget 1951; for an overview see Kesselring and Müller 2011); this starts to happen around the age of 2. Children appear very sensitive to the information they have shared with their interlocutors in discourse from that age onwards, indeed. Matthews et al. (2006) found that children respond to prior mention in discourse as an indicator of referent accessibility from age 2, but it takes them until they are 3 to respond to perceptual availability as an indicator of referent accessibility.

The requirement to track and mark referential status for the hearer may also go a long way toward explaining the challenge the article system poses for L2 learners whose L1 does not include an obligatory article category. As with spatial referential systems (Levinson 1996), the learner needs to figure out that the system is about a speaker marking information for their hearer from the hearer's perspective. In a sense, the article category represents a grammaticalized version of "audience design" (Bell 1984). Yet rather than it being about how speakers voluntarily tailor utterances for addressees from a stylistic point of view, it is about how the grammatical system forces speakers to tailor reference to their hearer. To do this right, speakers need to inhabit the same discourse reality or knowledge space as (they assume exists on the part of) their hearer and tailor reference from the hearer's vantage point.

Butler (2002: 472–473) confirmed that adult Japanese learners of English found the concept of Hearer Knowledge the most difficult to grasp. Yet, mastering this concept improves article usage considerably. Hinenoya and Lyster (2015) tested Epstein's (2002) processing instruction hypothesis in the classroom and found that students who had been trained to engage themselves as speakers in a situation outperformed students who had received explicit instruction of article use in terms of identifiability, familiarity and uniqueness. The typical introduction of HK as definiteness does not sufficiently highlight the importance of the hearer. Master's (1990) model is the only one that promotes the primacy of hearer knowledge over

specificity for pedagogic purposes, but in his terminology too, hearer knowledge is called definiteness, which is typically explained in terms of identifiability, familiarity and uniqueness: *the* is there to identify, pick out, or individuate the referent so that the hearer can identify what is being discussed (Lyons 1999). Unfortunately, so explained, hearer knowledge is not very different from referent specificity; it comes as no surprise then that L2 learners, too, consider referent specificity as the conceptual driver of the article category and overgeneralize the definite article to first mention contexts because they initially associate *the* with the specificity of the referent (Huebner 1985; Thomas 1989: 351).

Our corpus study also showed that the information articles convey is retrievable from the context. L2 learners whose L1 lacks articles are used to tracking the referential status of an entity as hearers by relying on contextual information. Moving from an L1 that lacks articles to an L2 that expresses them overtly thus requires shifting perspective from being the hearer who needs to track this information implicitly to the speaker who needs to mark this information explicitly for their hearer. Work on learning helps us understand why this might be difficult: in situations in which a cue is already associated with an outcome, learning to associate a new (and informative) cue with that same outcome is difficult. Overshadowing and blocking explain this phenomenon (originally: Pavlov 1927 and Kamin 1969; for more recent discussions see, for example, Mackintosh [1971], Kruschke [2001]) and Ellis (2006b) has argued compellingly that similar "forces" are at work in L2 learning.

Of course, to some extent, learning to track reference from the perspective of the hearer may require learning the implicit cultural view of what can be assumed as known. This, again, links to the argument that learning, in general, is always contextualized (see, for example, Boddez et al. [2011] regarding the context-dependency of the blocking effect). For language learning, more specifically, the context seems to hold particular importance (for initial discussion see Anđel et al. 2015). Ultimately, proper (cultural) contextualization might turn out to be crucial for the efficient adaptation to a new system and for grasping its fixed and variable parts.

6.4 Reconciling discourse concerns with historical restrictions

So far, we have focused on the article system as a speaker-driven, hearer-focused system of referent tracking. But, of course, Hearer Knowledge is not the only variable at play. The hierarchical system is multivariate, and at least three further variables are needed to guarantee accurate usage, i.e., Referent Specificity, Countability and Number. Yet, the relative importance of these variables depends on whether Hearer Knowledge is presupposed or not. And interestingly, some of these variables encode

grammatical properties that link back to the origins of the articles in the demonstrative pronoun or the numeral *one*.

Once the value of Hearer Knowledge has been established there are two other variables that play a major role. In the case of HK+ this is Referent Specificity, where most instances are *the*. This is not entirely surprising, as the combination of HK+ SR+ conspires towards an easily retrievable referent, which is close to the origin of *the* as the demonstrative *se* (cf. also Langacker [1991:103] for a comparison of *the* and demonstratives). Specific referents, by definition, can be identified specifically and they could potentially be pointed at: they are one specific entity or a set of entities that are uniquely recognizable.

In the case of HK–, the second most important variable in line is Number. At this end of the spectrum, we typically have something that is unknown to the hearer and plural (or uncountable). This aligns with the potential vagueness of the zero article, which is reserved for plurals or uncountable nouns. When these conditions are not met, and we are dealing with a singular countable item, then *a* becomes the article of choice. As noted above (Section 2.1), restrictions on using *a* follow straightforwardly from *a* having developed from the numeral *one*: *one* can only be used with singular, countable nouns. The only instances where an uncountable noun takes *a* is when the noun is modified (e.g., *a willing and prompt obedience*), thus singling out a particular type of the concept denoted by the noun.

A note is in order concerning Countability and Number. Both Master (1997) and Butler (2002: 472–473) found that L2 speakers have problems deciding on the countability status of a referent (which, in their view, in turn, affects their judgment of HK). Countability is heavily context dependent, indeed, and many uncountable nouns are sometimes used countably (e.g., *That little boy's having a long sleep, isn't he*?) while mostly countable nouns can also be used uncountably (e.g., *It just tastes of ø banana*). Number is not necessarily straightforward either. Even for uncountable nouns we find both singular and plural nouns, e.g., *clothes* is a plural uncountable noun. These may be difficult to identify for learners as they are rather rare, but usually plural uncountable nouns take the plural marking *–s*. They do require a quantifier to be used as individual referents, e.g., *a pair of scissors.*

In sum, our findings highlight the need to present the article system as a hierarchical system, governed by Hearer Knowledge. As is typical of many grammaticalized categories, usage is governed by a range of variables, in this case a combination of newer discourse concerns and older morphological restrictions. Correct use of the article system as it currently exists requires the cognitive ability to present information from the interlocutor's point of view; this discourse demand needs to be reconciled with grammatical restrictions which are found lower down the hierarchy. A comprehensive view of a system exhibiting this level of complexity was made possible by the application of machine learning techniques: while

standard statistical classifiers reveal the major trends in the system, algorithms based on insights from learning unveil that not all outcomes are learned from the same cues and that some outcomes need to be learned from amalgams of less stable and interdependent individual cues.

7 Conclusions

In this study we presented a conceptually and methodologically interdisciplinary approach to the grammatical category of articles in English. We took a usage-based, cognitive linguistic approach to the function/meaning and use of articles and explored the challenge the system poses from the perspective of learning. Testing the main dimensions of experience that had been proposed in the literature on a large sample of spoken and written discourse data extracted from the BNC revealed that Hearer Knowledge is the driver of the system. Once Hearer Knowledge is acknowledged as the motivating principle of the category, article usage, which has long defied a succinct but encompassing description, becomes eminently predictable: the presence and absence of Hearer Knowledge each have their default articles, and these defaults are overridden in the same way, in line with restrictions on the forms from which the articles have developed. Computational simulations with an error-correction algorithm confirm that different subsets of variables act as cues for the different articles and that some articles need to be learned from more complex sets of less reliable cues.

We used our findings to argue for articles as a hierarchically structured, hearer-focused communicative device: the category presents itself as a grammaticalized indexing system that catalogs speaker knowledge for the benefit of the hearer. Fascinatingly, its historical origins shine through in this hierarchy: the further the articles have grammaticalized and support the now dominant discourse functions, the lower down the hierarchy the grammatical constraints are, and the unmarked zero article soaks up what does not fit with the historical origins of *a/the*. In a sense, the article category represents a grammaticalized version of "audience design". Yet, rather than it being about how speakers voluntarily tailor utterances for addressees from a stylistic point of view, it is about how the grammatical system forces speakers to express reference from the vantage point of the hearer.

This article presents new ways to approach complex grammatical phenomena. Simulating learning computationally on a carefully annotated sample of authentic discourse data offers an exciting methodological opportunity to observe how potential cues compete and acquire their strength as usage accumulates. This makes visible how usage and learnability interact in shaping the system, which opens up new avenues for thinking about languages as dynamic systems.

Acknowledgments.: Oghenetekevwe Kwakpovwe tagged the BNC with instances of the zero article. Daisy Collins completed the first round of annotation, included in SupMat. Both students were paid through a stipend from grant RL-2016-001.
Research funding: This research for this article was supported by a Leverhulme Trust Leadership Award RL-2016-001 to Dagmar Divjak, which funded all authors.
Data availability statement: The dataset supporting the conclusions of this article, together with the SupMat, can be downloaded from https://doi.org/10.25500/edata.bham.00000943. The R code necessary to reproduce the statistical models is available at https://github.com/ooominds/The-article-category-as-a-referent-tracking-system.

References

Abbott, Barbara. 1999. Support for a unique theory of definiteness. *Proceedings of Semantics and Linguistic Theory (SALT)* 9. 1–15.

Aguilar-Guevara, Ana & Joost Zwarts. 2010. Weak definites and reference to kinds. *Proceedings of Semantics and Linguistic Theory (SALT)* 20. 179–196.

Anđel, Maja, Jelena Radanović, Laurie B. Feldman & Petar Milin. 2015. Processing of cognates in Croatian as L1 and German as L2. *NetWordS 2015-Word Knowledge and Word Usage: Proceedings of the NetWordS Final Conference on Word Knowledge and Word Usage: Representations and Processes in the Mental Lexicon*.

Angele, Bernhard & Keith Rayner. 2013. Processing the in the Parafovea: Are articles skipped automatically? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39(2). 649–662.

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

Ariel, Mira. 1994. Interpreting anaphoric expressions: A cognitive versus a pragmatic approach. *Journal of Linguistics* 30. 3–42.

Baayen, Harald R., Petar Milin, Dusica F. Đurđević, Peter Hendrix & Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118(3). 438.

Bell, Allan. 1984. Language style as audience design. *Language in Society* 13(2). 145–204.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan & Randolph Quirk. 1999. *Longman grammar of spoken and written English*. London: Longman.

Bickerton, Derek. 1981. *Roots of language*. Ann Arbor: Karoma Press.

Bickerton, Derek. 1984. The language bioprogram hypothesis. *Behavioral and Brain Sciences* 7(2). 173–188.

Birner, Betty & Gregory Ward. 1994. Uniqueness, familiarity, and the definite article in English. In *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society: General session dedicated to the contributions of Charles J. Fillmore*, 93–102. Berkeley, CA: Berkeley Linguistics Society.

Boddez, Yannick, Frank Baeyens, Dirk Hermans & Tom Beckers. 2011. The hide-and-seek of retrospective revaluation: Recovery from blocking is context dependent in human causal learning. *Journal of Experimental Psychology: Animal Behavior Processes* 37(2). 230–240.

Brown, Roger. 1973. *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Burton-Roberts, Noel. 1976. On the generic indefinite article. *Language* 52(2). 427–448.

Butler, Yoko G. 2002. Second language learners' theories on the use of English articles: An analysis of the metalinguistic knowledge used by Japanese students in acquiring the English article system. *Studies in Second Language Acquisition* 24(3). 451–480.

Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*, 25–55. New York: Academic Press.

Chen, Zhe, Simon Haykin, Jos J. Eggermont & Suzanna Becker. 2008. *Correlative learning: A basis for brain and adaptive systems*. John Wiley & Sons.

Chevrot, Jean-Pierre, Céline Dugua & Michel Fayol. 2009. Liaison acquisition, word segmentation and construction in French: A usage-based account. *Journal of Child Language* 36. 557–596.

Christopherson, Paul. 1939. *The articles: A study of their theory and use in English*. Oxford: Oxford University Press.

Clark, Herbert H. 1996. *Using language*. Cambridge: Cambridge University Press.

Clark, Herbert & Gregory L. Murphy. 1982. Audience design in meaning and reference. *Advances in Psychology* 9. 287–299.

Comrie, Bernard. 1999. Reference-tracking: Description and explanation. *Sprachtypologie und Universalienforschung* 52. 335–346.

Croft, William & Alan D. Cruse. 2004. In *Cognitive linguistics*. Cambridge: Cambridge University Press.

Cziko, Gary A. 1986. Testing the language bioprogram hypothesis: A review of children's acquisition of articles. *Language* 62(4). 878–898.

Danks, David. 2003. Equilibria of the Rescorla–Wagner model. *Journal of Mathematical Psychology* 47(2). 109–121.

De Mulder, Walter & Anne Carlier. 2011. The grammaticalization of definite articles. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of grammaticalization*. Oxford: Oxford University Press.

DeLong, Katherine A., Thomas P. Urbach & Marta Kutas. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 8. 1117–1121.

Despić, Miloje. 2019. On kinds and anaphoricity in languages without definite articles. In Ana Aguilar-Guevara, Julia Pozas Loyo & Violeta Vázquez-Rojas Maldonado (eds.), *Definiteness across languages*, 259–291. Berlin: Language Science Press.

Divjak, Dagmar. 2015. Exploring the grammar of perception. A case study using data from Russian. *Functions of Language* 22(1). 44–68.

Divjak, Dagmar, Petar Milin, Adnane Ez-Zizi, Jarosław Józefowski & Christian Adam. 2021. What is learned from exposure: An error-driven approach to productivity in language. *Language, Cognition and Neuroscience* 36(1). 60–83.

Drieghe, Denis, Alexander Pollatsek, Adrian Staub & Keith Rayner. 2008. The word grouping hypothesis and eye movements during reading. *Journal of Experimental Psychology: Learning, Memory and Cognition* 34(6). 1552–1560.

Dryer, Matthew S. 1989. Article-Noun order. *Papers of the 25th Annual Regional Meeting of the Chicago Linguistic Society, Part One: The General Session, 83–97*.

Dryer, Matthew S. 2013. Definite articles. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

DuBois, John W. 1980. Beyond definiteness: The trace of identity in discourse. In Wallace L. Chafe (ed.), *The pear stories* (Advances in Discourse Processes), 203–274. Norwood, NJ: Ablex.

Ellis, Nick C. 2006a. Language acquisition as rational contingency learning. *Applied Linguistics* 27(1). 1–24.

Ellis, Nick C. 2006b. Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics* 27(2). 164–194.

Enquist, Magnus, Johan Lind & Stefano Ghirlanda. 2016. The power of associative learning and the ontogeny of optimal behaviour. *Royal Society Open Science* 3(11). 160734.
Ensz, Kathleen Y. 1982. French attitudes toward typical speech errors of American speakers of French. *The Modern Language Journal* 66(2). 133–139.
Epstein, Richard. 2002. The definite article, accessibility, and the construction of discourse referents. *Cognitive Linguistics* 12(4). 333–378.
Farkas, Donka F. 2002. Specificity distinctions. *Journal of Semantics* 19(3). 213–243.
Fauconnier, Gilles. 1994. *Mental spaces: Aspects of meaning construction in natural language*. Cambridge: Cambridge University Press.
Foley, William A. & Robert D. Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
Foley, William A. & Robert D. Van Valin. 1985. Information packaging in the clause. In Timothy Shopen (ed.), *Language typology and syntactic description*, 282–364. Cambridge: Cambridge University Press.
Gautier, Vincent, Kevin J. O'Regan & Jean F. Le Gargasson. 2000. The-skipping' revisited in French: Programming saccades to skip the article 'les'. *Vision Research* 40. 2517–2531.
Givón, Talmy. 1984. *Syntax: A functional-typological introduction I*. Amsterdam & Philadelphia: John Benjamins.
Givón, Talmy. 1992. The grammar of referential coherence as mental processing instructions. *Linguistics* 30(1). 5–55.
Green, Georgia M. 2012. *Pragmatics and natural language understanding*. London: Routledge.
Greenbaum, Sidney. 1996. *The Oxford English grammar*. Oxford: Oxford University Press.
Haiman, John & Pamela Munro (eds.). 1983. *Switch reference and universal grammar: Proceedings of a symposium on switch reference and universal grammar*. Amsterdam & Philadelphia: John Benjamins.
Hawkins, John A. 1978. *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. London: Croom Helm.
Haykin, Simon S. 1999. *Neural networks: A comprehensive foundation*, 2 edn. London: Prentice Hall.
Heim, Irene. 1983. File change semantics and the familiarity theory of definiteness. In Rainer Bäuerle, Christoph Schwarze & Arnim von Stechow (eds.), *Meaning, use, and interpretation of language*, 164–189. Berlin & New York: Walter de Gruyter.
Heine, Bernd. 1997. *Cognitive foundations of grammar*. Oxford: Oxford University Press.
Hinenoya, Kimiko & Roy Lyster. 2015. Identifiability and accessibility in learning definite article usages: A quasi-experimental study with Japanese learners of English. *Language Teaching Research* 19(4). 397–415.
Hopper, Paul J. & Elizabeth Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics* 15(3). 651–674.
Huebner, Thomas. 1983. *A longitudinal analysis of the acquisition of English*. Michigan: Karoma.
Huebner, Thomas. 1985. System and variability in interlanguage syntax. *Language learning* 35(2). 141–163.
Ionin, Tania. 2003. *Article semantics in second language acquisition*. Boston: Massachusetts Institute of Technology Dissertation.
Johnson, Ruth & Frederick L. Jenks. 1994. Native speakers' perceptions of nonnative speakers: Related to phonetic errors and spoken grammatical errors. *Paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages, Baltimore, MD, 8–12 March 1994*.
Kadmon, Nirit. 1990. Uniqueness. *Linguistics and Philosophy* 13(3). 273–324.
Kamin, Leon J. 1969. Predictability, surprise, attention and conditioning. In Byron A. Campbell & Russell M. Church (eds.), *Punishment and aversive behavior*, 279–296. New York: Appleton-Century-Crofts.

Kamp, Hans. 2013. A theory of truth and semantic representation. In Klaus von Heusinger & Alice ter Meulen (eds.), *Meaning and the dynamics of interpretation: Selected papers of Hans Kamp*, 329–369. Leiden: Brill.

Karmiloff-Smith, Annette. 1979. *A functional approach to child language: A study of determiners and reference*. Cambridge: Cambridge University Press.

Karttunen, Lauri. 1976. Discourse referents. In James D. McCawley (ed.), *Syntax and semantics 7: Notes from the linguistic underground*, 363–386. New York: Academic Press.

Kesselring, Thomas & Ulrich Müller. 2011. The concept of egocentrism in the context of Piaget's theory. *New Ideas in Psychology* 29(3). 327–345.

Kibort, Anna. 2008. *Grammatical features inventory: Definiteness*. SMG, University of Surrey. https://www.smg.surrey.ac.uk/features/.

König, Ekkehard. 2018. Definite articles and their uses: Diversity and patterns of variation. In Daniël Van Olmen, Tanja Mortelmans & Frank Brisard (eds.), *Aspects of linguistic variation*, 165–184. Berlin & Boston: De Gruyter Mouton.

Kruschke, John K. 2001. Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology* 45(6). 812–863.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus and mental representations of discourse referents*. Cambridge: Cambridge University Press.

Langacker, Ronald W. 1991. *Foundations of cognitive grammar*. Stanford, CA: Stanford University Press.

Learn English: English Grammar: English grammar reference: Determiners and quantifiers: the indefinite article 'a' and 'an'. British Council. https://learnenglish.britishcouncil.org/grammar/english-grammar-reference/indefinite-article?page=2 (accessed 16 May 2023).

Learn English: English Grammar: English grammar reference: Determiners and quantifiers: the definite article 'the'. British Council. https://learnenglish.britishcouncil.org/grammar/english-grammar-reference/definite-article (accessed 16 May 2023).

Leech, Geoffrey. 1992. 100 million words of English: The British National Corpus (BNC). *Language Research* 28(1). 1–13.

Leech, Geoffrey, Paul Rayson & Andrew Wilson. 2001. *Word frequencies in written and spoken English: Based on the British National Corpus*. London: Routledge.

Leiss, Elisabeth. 2000. *Artikel und Aspekt: Die grammatischen Muster von Definitheit*. Berlin & New York: De Gruyter.

Levinson, Stephen C. 1996. Frames of reference and Molyneux's question: Crosslinguistic evidence. In Paul Bloom, Merrill F. Garrett, Lynn Nadel & Mary A. Peterson (eds.), *Language and space*, 109–169. Cambridge, MA: MIT Press.

Liaw, Andy & Matthew Wiener. 2018. Breiman and cutler's random forests for classification and regression, version 4.6–14.

Lidz, Jeffrey, Henry Gleitman & Lila Gleitman. 2003. Understanding how input matters: Verb learning and the footprint of universal grammar. *Cognition* 87. 151–178.

Lieven, Elena, Julian M. Pine & Gillian Baldwin. 1997. Lexically-based learning and early grammatical development. *Journal of Child Language* 24. 187–219.

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

Mackintosh, Nicholas J. 1971. An analysis of overshadowing and blocking. *Quarterly Journal of Experimental Psychology* 23(1). 118–125.

Magnan, Sally S. 1982. Native speaker reaction as a criterion for error correction. In Alan Garfinkel (ed.), *Report of Central States Conference on the Teaching of Foreign Language*, 30–46. Skokie, IL: National Textbook.

Maratsos, Michael P. 1976. *The use of definite and indefinite reference in young children*. Cambridge: Cambridge University Press.
Master, Peter. 1990. Teaching the English articles as a binary system. *TESOL Quarterly* 24(3). 461–478.
Master, Peter. 1997. The English article system: Acquisition, function, and pedagogy. *System* 25(2). 215–232.
Matthews, Danielle, Elena Lieven, Anna Theakston & Michael Tomasello. 2006. The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics* 27. 403–422.
Meylan, Stephan C., Michael F. Frank, Brandon C. Roy & Roger Levy. 2017. The emergence of an abstract grammatical category in children's early speech. *Psychological Science* 28(2). 181–192.
Michaelis, Laura. 2004. Type shifting in construction grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics* 15(1). 1–67.
Milin, Petar, Dagmar Divjak & Harald R. Baayen. 2017a. A learning perspective on individual differences in skilled reading: Exploring and exploiting orthographic and semantic discrimination cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43(11). 1730–1751.
Milin, Petar, Laurie Beth Feldman, Michael Ramscar, Peter Hendrix & Harald R. Baayen. 2017b. Discrimination in lexical decision. *PLoS One* 12(2). e0171935.
Milin, Petar, Harish Tayyar Madabushi, Michael Croucher & Dagmar Divjak. 2020. Keeping it simple: Implementation and performance of the proto-principle of adaptation and learning in the language sciences. *arXiv preprint arXiv:2003.03813*.
Nieuwland, Mante S., Stephen Politzer-Ahles, Evelien Heyselaar, Katrien Segaert, Emily Darley, Nina Kazanina, Sarah Von Grebmer Zu Wolfsthurn, Federica Bartolozzi, Vita Kogan, Aine Ito, Diane Meziere, Dale J. Barr, Guillaume A. Rousselet, Heather J. Ferguson, Simon Busch-Moreno, Xiao Fu, Jyrki Tuomainen, Eugenia Kulakova, Matthew E. Husband, David I. Donaldson, Zdenko Kohut, Shirley-Ann Rueschemeyer & Falk Huettig. 2018. Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *eLIFE* 7. e33468.
Pavlov, Ivan P. 1927. *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. New York: Dover Publications.
Piaget, Jean. 1951. Pensée égocentrique et pensée sociocentrique. *Cahiers Internationaux de Sociologie* 10. 34–49.
Pine, Julian M., Daniel Freudenthal, Grzegorz Krajewski & Fernand Gobet. 2013. Do young children have adult-like syntactic categories? Zipf's law and the case of the determiner. *Cognition* 127. 345–360.
Pine, Julian M. & Elena Lieven. 1997. Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics* 18. 123–138.
Pine, Julian M. & Helen Martindale. 1996. Syntactic categories in the speech of young children: The case of the determiner. *Journal of Child Language* 23(2). 369–395.
Ramscar, Michael, Daniel Yarlett, Melody Dye, Katie Denny & Kirsten Thorpe. 2010. The effects of Feature–Label–Order and their implications for symbolic learning. *Cognitive Science* 34(6). 909–957.
Rescorla, Robert A. 1968. Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology* 66(1). 1–5.
Rescorla, Robert A. 2008. Rescorla-Wagner model. *Scholarpedia* 3(3). 2237.
Rescorla, Robert A. & Allan R. Wagner. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Abraham H. Black & William F. Prokasy (eds.), *Classical conditioning II: Current research and theory*, 64–99. New York: Appleton-Century-Crofts.
Roberts, Craige. 2003. Uniqueness in definite noun phrases. *Linguistics and Philosophy* 26(3). 287–350.

Romain, Laurence, Adnane Ez-zizi, Petar Milin & Dagmar Divjak. 2022. What makes the past perfect and the future progressive? Experiential coordinates for a learnable, context-based model of tense and aspect. *Cognitive Linguistics* 33(2). 251–289.

Romain, Laurence, Dagmar Hanzlíková, Petar Milin & Dagmar Divjak. (In press). Ruled by Construal? Framing article choice in English. *Constructions and Frames*.

Russell, Bertrand. 1905. On denoting. *Mind* 14. 479–493.

Schwarz, Florian. 2013. Two kinds of definites cross-linguistically. *Language and Linguistics Compass* 7(10). 534–559.

Shin, Yu K. & YouJin Kim. 2017. Using lexical bundles to teach articles to L2 English learners of different proficiencies. *System* 69. 79–91.

Sommerer, Lotte. 2018. *Article emergence in Old English. A constructionalist perspective*. Berlin & Boston: De Gruyter Mouton.

Spellman, Barbara A. 1996. Conditionalizing causality. In David R. Shanks, Keith James Holyoak & Douglas L. Medin (eds.), *The psychology of learning and motivation: Advances in research and theory*, 167–206. New York: Academic Press.

Stirling, Lesley. 1993. *Switch-reference and discourse representation*. Cambridge: Cambridge University Press.

Strawson, Peter F. 1950. On referring. *Mind* 59(235). 320–344.

Thomas, Margaret. 1989. The acquisition of English articles by first-and second-language learners. *Applied Psycholinguistics* 10(3). 335–355.

Tolman, Edward C. & Egon Brunswik. 1935. The organism and the causal texture of the environment. *Psychological Review* 42(1). 43–77.

Trenkic, Danijela. 2000. *The acquisition of English articles by Serbian speakers*. Cambridge: Cambridge University Dissertation.

Trenkic, Danijela. 2008. The representation of English articles in second language grammars: Determiners or adjectives. *Bilingualism: Language and Cognition* 11(1). 1–18.

Valian, Virginia. 1986. Syntactic categories in the speech of young children. *Developmental Psychology* 22(4). 562–579.

Valian, Virginia, Stephanie Solt & John Stewart. 2009. Abstract categories or limited-scope formulae? The case of children's determiners. *Journal of Child Language* 36(4). 743–778.

Vann, Roberta J., Daisy E. Meyer & Frederick O. Lorenz. 1984. Error gravity: A study of faculty opinion of ESL errors. *TESOL Quarterly* 18(3). 427–440.

Warden, David A. 1976. The influence of context on children's use of identifying expressions and reference. *British Journal of Psychology* 67(1). 101–112.

Widrow, Bernard. 1959. Adaptive sampled-data systems: A statistical theory of adaptation. *IRE Wescon Convention Record* 4. 74–85.

Widrow, Bernard & Marcian E. Hoff. 1960. Adaptive switching circuits. *WESCON Convention Record* 96–104. https://www-isl.stanford.edu/~widrow/papers/c1960adaptiveswitching.pdf.

Widrow, Bernard & Michael A. Lehr. 1990. 30 years of adaptive neural networks: Perceptron, Madaline, and backpropagation. *Proceedings of the IEEE* 78(9). 1415–1442.

Yang, Charles. 2013. Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences of the United States of America* 110(16). 6324–6327.

Zeileis, Achim, Torsten Hothorn & Kurt Hornik. 2008. Model-based recursive partitioning. *Journal of Computational and Graphical Statistics* 17(2). 492–514.

Zhao, Helen & Brian MacWhinney. 2018. The instructed learning of form-function mappings in the English article system. *The Modern Language Journal* 102(1). 99–119.