

Article

The Emergence of Second Language Categorisation of the English Article Construction

Helen Zhao 

School of Languages and Linguistics, University of Melbourne, Melbourne 3010, Australia;
helen.zhao@unimelb.edu.au

Received: 7 September 2020; Accepted: 3 November 2020; Published: 7 November 2020



Abstract: This study examines the emergent cognitive categorisation of the English article construction among second language (L2) learners. One hundred and fourteen Mandarin-L1 learners of English, divided into two L2 proficiency levels (low-to-intermediate and advanced), were measured by a computer-based cloze test for the accuracy and response time of appropriate use of English articles in sentential contexts. Results showed that when learners acquired the polysemous English article construction they demonstrated stronger competence in differentiating individual form-function mappings in the article construction. L2 learners' patterns of article construction usage were shaped by semantic functions. Learners performed better on the definiteness category than on the non-definiteness categories, suggesting that learners were sensitive to the prototypicality of nominal grounding. Advanced learners demonstrated an increased sensitivity to semantic idiosyncrasy, but they lacked contextualised constructional knowledge. Competition among the functional categories and restructuring of functional categories are important ways of regularization that learners go through to acquire semantically complex systems such as articles.

Keywords: English article construction; categorization; prototypicality; schematization; cognitive grammar; second language

1. The Emergence of Second Language Categorisation of the English Article Construction

Within the framework of usage-based linguistic theories (Ellis 2008; Goldberg 1995, 2019; Langacker 1991, 2008; Tomasello 2003), the units that make up grammar are derived from language use. Language speakers comprehend messages given *forms*. They choose forms in order to convey intended messages (*meanings*) for production. Form-meaning mappings that are conventionalised become language *constructions* as a result of frequent usage in the community. The conventionalisation of form-function mappings is associated with processes of abstraction and schematisation. Abstract structures or schemata emerge as a result of generalisation of patterns across instances of language use. Grammar is a combination of such abstract structures that operate at levels of words, morphemes, phrases, and clauses. Such an approach to grammar shreds the traditional boundary between lexicon and syntax with an attempt to apply the generalisation commitment of cognitive linguistics to the study of language and language acquisition (Evans and Green 2006).

Constructions are polysemous (Lakoff 1987; Langacker 1991). There is seldom one-to-one mapping between form and meaning. Forms (*cues*) in language compete to be associated with the meaning to be expressed (MacWhinney 1987). Meanwhile, one form can be associated with multiple, context-dependent meanings (Croft and Cruse 2004). The schemata of a construction are instantiated by potentially many form-function mappings that vary in terms of centrality of category membership (Bybee 2010; Goldberg 1995). The central members (prototypes) of a construction, due to their high frequency of usage and high semantic and structural similarity with other members of the category, play a significant role in shaping the schemata of the construction.

Language learning is the acquisition of conventional form-function mappings. To acquire a construction (i.e., a system of form-function mappings) involves extracting structural and semantic regularities from large amounts of exemplars in the input. Drawing analogy based on input exemplars allows learners to formulate the formal schematic representation of the construction and come to understand the associated polysemous meanings. The majority of the empirical studies on usage-based acquisition of a second language (L2) have focused on verb-argument constructions (VACs) (Azazil 2020; Ellis 2008, 2016; Ellis and Ferreira-Junior 2009; Römer and Berger 2019; Sung and Kim 2020). The studies have adopted online processing methods such as recognition and naming tasks and acceptability judgment tasks, as well as free or structured production tasks. Their findings reveal that as L2 proficiency increases, learners gradually expand their VAC repertoire and productivity and approximate native usage. Learners develop implicit knowledge on the VACs that are influenced by factors such as VAC verb frequency, verb contingency (strength of form-meaning association), prototypicality of meaning, etc.

The current study shifts our focus to the nominals in English and investigates the usage-based acquisition of the English article construction. We present a contrastive review of traditional linguistics accounts versus usage-based accounts of English articles. The study collected empirical data from the L2 learners of two proficiency groups, and its findings shed light on the changes that learners went through in the process of forming more target-like schematic representation of the English article construction. The findings showed that when learners acquired the polysemous construction they were able to draw analogy among form-meaning mappings with semantic and structural similarities. This categorisation process was influenced by prototypicality and function competition.

2. Traditional Linguistic Analysis of English Articles

One of the earliest frameworks of article analysis is Bickerton's (1981) semantic wheel for noun phrase (NP) reference, marked by two semantic features, i.e., specificity (SR) and hearer knowledge (HK). SR means whether the speaker of the utterance has a specific referent in mind, whereas HK refers to whether the referent is assumed to be known to the hearer. The variations of the two features generate four distinct categories of meanings associated with the definite article (*the*), the indefinite article (*a, an*), and the zero article (\emptyset) (see Table 1). This type of categorisation has been widely adopted in L2 studies that attempted to examine the acquisition sequence of the articles (Butler 2002; Huebner 1983; Tarone and Parrish 1988). Huebner (1983), for instance, was a longitudinal case study of an L1 Cambodian learner of English whose definite article use was identified for three stages of development: the use was initially restricted to only [+SR, +HK] contexts at stage 1 and then was overgeneralised to all noun phrase environments at stage 2; the learner finally constrained its use to +HK contexts, achieving target-like form-meaning association.

The generative approach to articles (Hawkins et al. 2006; Ionin et al. 2004) adopts a similar analytical framework that is concerned with the semantics of *specificity* and *definiteness* that the authors described as two article choice parameters. Specificity is the corresponding term for SR and is concerned with whether a speaker has a referent in mind and intends to refer to it (referential) or not (quantificational; non-referential) (Fodor and Sag 1982). Ionin et al. (2004) added the additional property of "noteworthiness" to the definition of specificity, which means that the speaker has "something worthy to note about the referent in the present context" (Trenkic 2008). Definiteness is defined as the speaker and hearer's shared assumption of the unique identifiability of the referent in discourse (Ionin et al. 2004), which is essentially equivalent to the notion of HK (Bickerton 1981). Similar to Huebner's (1983) findings, Ionin et al. (2004) reported that L2-English learners from article-less L1s mistakenly associate *the* with specificity instead of definiteness and, as a result, overuse the definite article in specific indefinite contexts.

Table 1. Four Noun Phrase (NP) Categories and Semantic Functions (Tarone and Parrish 1988).
SR = Specific referent; HK = Hearer knowledge.

No.	Feature	Form	Semantic Function
1	[-SR] [+HK]	<i>the</i>	Generics (e.g., <i>The lion is a beautiful animal.</i>)
		<i>a/an</i>	Generics (e.g., <i>A lion is a beautiful animal.</i>)
		∅	Generics (e.g., <i>∅ Lions are beautiful animals.</i>)
2	[+SR] [+HK]	<i>the</i>	Unique referent or conventionally assumed unique referent (e.g., <i>The pope</i>)
			Referent physically present (e.g., <i>Ask the guy over there</i>)
			Referent previously mentioned in the discourse (e.g., <i>the man you have just said</i>)
3	[+SR] [-HK]	<i>a/an</i>	First mention in a discourse of [+SR] NP which is assumed to be not known to the hearer (e.g., <i>Dad gave me a car.</i>)
		∅	First mention of [+SR] NP following existential <i>have</i> and assumed to be not known to the hearer (e.g., <i>Our house has a garage.</i>)
			Equative noun phrases (e.g., <i>He is a nice man.</i>)
4	[-SR] [-HK]	<i>a/an</i>	Noun phrases in the scope of negation (e.g., <i>I don't see a pencil.</i>)
		∅	Noun phrases in the scope of interrogative (e.g., <i>Do you see a pencil?</i>)
			Noun phrases in irrealis scope (e.g., <i>If I had a million dollars, I'd buy a big yacht.</i>)

The above two-parameter framework to articles attempted to fit all the article functions to the four sub-categories [\pm definiteness, \pm specificity]. It has been a productive framework to conducting crosslinguistic comparisons and examining crosslinguistic influence in L2 article acquisition (García Mayo and Hawkins 2009; Hawkins et al. 2006) with the aim of looking for semantic universals in language acquisition. On the other hand, such a framework also has limitations. First, not all the article functions can fit into the four categories. A large amount of conventionalised and idiomatic usage of articles, such as *the Mississippi River*, *Lake Michigan*, *the Alps*, *Mount Fuji*, cannot easily fit into the four categories but need to be incorporated into the article learning model and be examined for L2 developmental stages together with the other general article functions. Second, some generative distinctions may not be that significant when we focus on the learner’s actual language use. For example, the distinction between the referential function (e.g., *She ate an apple*) and the quantificational function (e.g., *She wants an apple*) of the indefinite article is important in the generative framework because this semantic distinction is formally marked in some languages (but not in English). But it may not be necessary for L2-English learners to be aware of the referential-quantificational distinction when formulating the two example sentences that both use the indefinite article. Instead, it may be more important for them to be able to judge the countability of the referent (*apple*) and to determine whether the referent is uniquely identifiable in discourse.

3. The English Article Construction: Usage-Based Accounts

We first review Langacker’s (1991, 2008) cognitive grammar account of articles. In cognitive grammar, language structures are defined by highly schematic configurational concepts, e.g., contrast, boundary, change, contact, proximity, which are experientially-based and fundamental in our everyday life. Each such concept can be characterised semantically in terms of both a prototype (central instances in the category) and a schema (conception at a high level of abstraction extracted from usage events) instantiated by all instances of the category.

According to Langacker, articles are a category of nominal predications that defines the figure and ground relationship in discourse. Grounding is the primary schematic concept associated with article usage. Through nominal grounding devices such as articles and other determiners, the speaker directs

the hearer's attention to the figure (i.e., the intended discourse referent) in relation to a ground (i.e., the speech event and its participants). Grounding establishes a conceptualising relationship between the subject of conception (interlocutors with the speaker S and the hearer H) and the object of conception (the content evoked by a nominal expression).

Nominal grounding is prototypically overt. The intended discourse referent becomes overtly profiled in configuration, which is achieved through instantiation. A nominal type (e.g., *car* or *cars*) (represented by the top part of Figure 1a) involves an open-ended set of actual or imagined instances (represented by the bottom part of Figure 1a), while no instance is being profiled. Through the mental operation of instantiation, a type conception is transformed into an instance conception (e.g., *a car*) and as a result a salient instance in the type is being profiled in configuration (Figure 1b). The type/instance distinction lies in profiling (attention-directing) and in specificity. The instance conception is the default expectation for the indefinite article. In situations of unique instantiation (e.g., *the car*), i.e., when there is only one unique instance available of the specified type in the immediate scope of the discourse context (G) constructed by knowledge of the speaker (S) and the hearer (H), the definiteness conception applies (see Figure 1c). Because there is only one eligible candidate, it is unnecessary to distinguish it from other instances of the type. So an important distinction between the definite and the indefinite is that the indefinite fails to direct attention to any particular instance of the specified type. In addition, the generic¹ interpretation of definite reference reflects a special configuration of the definite, as it contains the construal of both type and unique instantiation. For example, in the sentence *the cat is a mammal*, the cat is construed as type—not a particular cat spatially instantiated, but meanwhile, it is construed as the unique instantiation of a type higher in hierarchy (mammal).

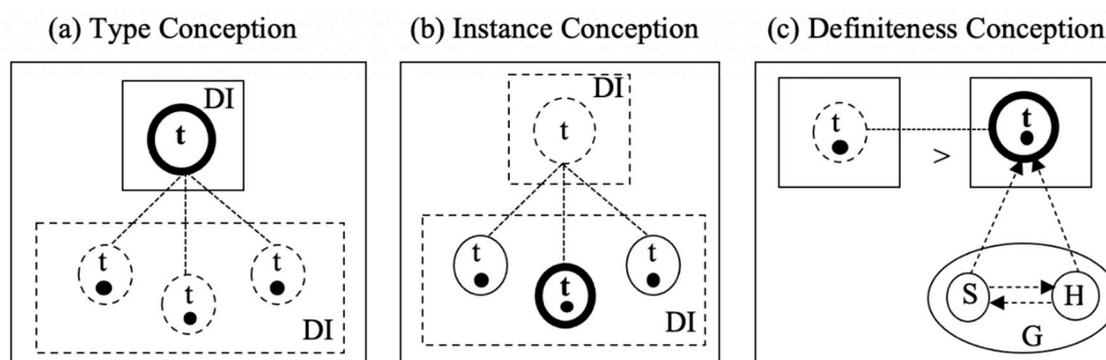


Figure 1. Schematic representations of conceptions of type, instance, and definiteness (Langacker 2008). S = Speaker; H = Hearer; G = Ground; DI = Domain of instantiation; t = type specification

Langacker also specified some non-prototypical types of grounding relationship. Grounding can be intrinsic, such as in proper names (e.g., *John, Australia*) “since the very meanings of such expressions imply the identifiability of their referents, they do not require a separate grounding element” (Langacker 2008, p. 272). Therefore, proper names are inherently definite. Abstract (mass) nouns such as *education* and *philosophy* also demonstrate inherent uniqueness (Radden and Dirven 2007). Grounding can be indirect, such as in possessives (e.g., *John's car*), in which the profiled instance of *car* is not related to the ground directly, but only indirectly, via the intrinsic grounding of *John*. Grounding can also be covert, while the noun is not grounded by a separate overt element. The zero article used with mass nouns (e.g., *∅ water*) encodes a kind of covert grounding. Such zero grounding indicates no profiled instance of the type because it is difficult to tease apart instances from type given the inherent property of mass nouns. Zero grounding leads to unboundedness in construal with no salient instance identified.

¹ The distinction between the specific and generic reference of the indefinite is played down in cognitive grammar (Verspoor and Huang 2008).

Some additional article usages such as with names of rivers, lakes, malls, and parks are highly conventionalised and display idiosyncrasy. Verspoor and Huong (2008) explained the idiosyncrasy in article use as an outcome of language evolution as a function of familiarity (Jespersen 1933). During the course of language evolution, phrases such as *the Oxford road* may experience changes and get evolved into *the Oxford Road* or *Oxford Road* and may eventually simply become *Oxford* as the name becomes more and more familiar to language users. Sometimes it is necessary to retain the definite article in proper names for semantic differentiation such as in names of states (*Ohio*) and names of rivers (*the Ohio*) (Radden and Dirven 2007).

The cognitive approach to grammar accommodates both general and idiosyncratic usages in principle, but focuses on accounting for general usages and largely ignores many idiosyncrasies of the less typical usages. To address this gap, Zhao and MacWhinney (2018) distinguished two categories of cues (signalling form-function mapping) in the English article construction: general cues and idiosyncratic cues. General cues (e.g., '*noncount* → \emptyset ', '*second mention* → *the*') represent regular usages and can be expressed by a transparent, general mapping from a general category, whereas idiosyncratic cues (e.g., '*singular mountain* → \emptyset ', '*plural mountains* → *the*', '*river* → *the*', '*lake* → \emptyset ') are based on small lexical fields and their usages can best be explained by aspects of phrasal structure and historical convention.

Zhao and MacWhinney (2018) extracted 86 article cues that they claimed to constitute the English article construction (excluding idiomatic usages such as *by \emptyset hand*). These cues are specific instantiations of the schemas for type, instance, and definiteness (Langacker 2008). The cues differ in frequency, reliability (MacWhinney 1987), and prototypicality. A prototypical instantiation of the type schema is the cue '*plural* → \emptyset ' (e.g., \emptyset *cars*); a prototype of the instance schema is the cue '*singular countable* → *a/an*' (e.g., *a Shakespearean drama*); and a prototype of the definiteness schema is the cue "*singular countable with post-modifiers* → *the*" (e.g., *the man she is dating*). Due to its comprehensiveness, Zhao and MacWhinney's (2018) usage-based cue system for the English article construction is adopted by the current study. Langacker's cognitive grammar lays the theoretical foundation for the study.

4. A Usage-Based Account of Language Development

Usage-based researchers argue that language acquisition is a bottom-up process of item-based learning via generalisation, categorisation, and schematization (Bybee 2008; Ellis 2003). Experience of tokens such as *the man at the door*, *the man who came*, *the man she is dating* form item-based local generalisations, i.e., '*the-man-Post-modifier*'. Learners' semantic encoding of a local generalisation is more contextualised, restricted, and incomplete. Once the category is formed, the speaker does not entirely discard the memory of the exemplars upon which the generalisation is based (Bybee 2010; Langacker 1987). Novel tokens are compared for structural and semantic similarity against the already existing tokens in the linguistic memory. Increased exposure and usage create opportunities for the emergence of more abstract categories and generalisations such as '*singular countable with post-modifiers* → *the*' as a schematic representation of a definite article usage.

Meanwhile, learners are exposed to many exemplars such as *a man who came*, *a book he wrote*, *an actor on stage*. The experience of these items competes with the linguistic memory of the existing schematic structure and 'forces' learners to refine its semantic encoding until they realise that singular countable nouns with post-modifiers only take the definite article when the referent is a unique instantiation of a specified type in the discourse, i.e., '*singular countable with post-modifiers* → *the*'. This semantic refinement is not easy to acquire for L2 learners. Learners may simply fail to register the notion of unique instantiation in various speech contexts, especially because articles are not a morphologically salient form and most article errors may not affect communication and get corrected. If there are more cases in English where an indefinite article is used for singular countable nouns with post-modifiers, the more frequent generalisation wins the competition and may create an overshadowing or even blocking effect on the learning of the less frequent generalisation (Ellis 2006).

Empirical research on the development of constructional knowledge has focused on first language acquisition (e.g., [Ambridge and Lieven 2015](#); [Goldberg 1999](#); [Lieven et al. 1997](#)). There are only a very limited number of L2 studies that have examined how L2 learners acquire the constructional knowledge of verb-argument constructions ([Azazil 2020](#); [Roehr-Brackin 2014](#); [Römer and Berger 2019](#)), motion constructions ([Li et al. 2014](#)), negations and interrogatives ([Eskildsen 2012, 2015](#)), and lexical constructions ([Crossley and Salsbury 2011](#); [Eskildsen 2008](#)). The overall findings of these studies aligned with the usage-based prediction of constructional emergence from lexical items to more abstract patterns. With the increase of L2 competence, learners expanded their constructional repertoire and productivity. Yet, most of these studies have “relied almost exclusively on small sets of data, often collected from only one learner or a small number of learners” and there is lack of data collected from learners with higher L2 proficiency levels ([Römer and Berger 2019](#), pp. 1090–91). Furthermore, almost all the studies have exclusively focused on the examination of the verb systems in language and there is a severe lack of attention to the acquisition of NP-related constructions.

5. The Present Study

The current study addresses the gaps identified in the existing literature on usage-based L2 acquisition of constructional knowledge by analysing changes in learners’ use of the English article construction in a lab-based experimental task at two proficiency levels from low-to-intermediate to advanced. It adopts quantitative methods to investigate the emergence of L2 categories as learners develop constructional competence with regard to the use of English articles. The study aims to investigate the following research questions:

1. What categories emerge from second language usage of article cues?
2. Does L2 learners’ emergent categorisation of article cues change as a function of increased proficiency? If yes, how?

6. Methodology

6.1. Participants

One hundred and fourteen Mandarin-L1 learners of English (38 males, 76 females) participated in this experiment. They were students studying in a public university in Beijing that specialises in foreign language studies. All participants gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the protocol approved by the Institutional Review Board (IRB00000603) of HS11-402.

The pool of participants included two proficiency levels: low-to-intermediate and advanced. The low-to-intermediate group consisted of 50 participants who were enrolled in the English distance learning program in the university. They had online evening classes on weekdays and face-to-face campus classes on the weekend. Their average age was 25 (range between 19 to 30 years old). Their average years spent learning English was 11 (range from 8 to 14 years). Due to lack of feasibility, an English proficiency test could not be administered in this group. Their English class instructor described their English proficiency to be low-to-intermediate with a fair amount of individual variations within the group. The study was administered to this group in a weekend face-to-face class in a computer lab on campus. Only 39 participants in this group finished the session.

The advanced group were 64 participants who were first-year and second-year full-time English majors at the same university. Their average age was 19.5 (range between 19 to 20 years old). Their average years spent learning English was 9 (range between 7 to 11 years). The Michigan Test of English Language Proficiency (MTELP) was administered to measure their general English proficiency. Their MTELP mean score was 83.5/100 (SD = 10.2/100), which placed most of them between the “very good” (upper intermediate) and “high command” (advanced) categories. The study was administered to this group in a regular English class in a computer lab. All the participants in this group completed the session.

6.2. Measurement

A computer-based cloze test was administered to measure participants' performance of English article usage in sentential contexts, for example, "Jennifer bought ____ earrings she found in a fashion magazine". Participants used a drop-down menu bar to make a choice between three options: *the*, *a/an*, and \emptyset . Four items were placed on one computer screen to avoid having too many items per screen and in a random order. The computer only logged the first response given by the participants for each item. Participants were allowed to move to the next screen only after they completed the responses to all four items on the screen. Once moving on, they were not allowed to go back to a previous screen. At the end of the test, each participant was shown their percentage of accuracy in the test. The study was programmed in JavaScript by a professional research programmer.

The current study examined 23 article cues² selected from the cue system by Zhao and MacWhinney (2018). Appendix A presents the cue list with metalinguistic explanations and exemplars. In terms of token frequency, the 23 article cues accounted for more than 80% in total of the article tokens in the L1-English written corpus analysis by Zhao and MacWhinney (2018). In the cloze test, participants were measured for 8 items per cue with a total number of 184 sentences (23 cues \times 8 sentences).

The sentences had an average word count of 10.4 words (ranging from 9 to 12 words) and were selected from the Corpus of Contemporary American English (COCA) (Davies 2008) and the Article Book (Cole 2000), but were modified to include lexical items that matched the learners' proficiency level. The sentences were selected and modified to provide an obligatory context for the choice of the target article without additional referential or contextual information. The participants were explicitly instructed not to assume any prior or later discourse for each sentence item. Two native English speaking research assistants validated the sentences and the obligatory contexts for article choices. The cloze test reliability was verified by means of the internal consistency of responses to the items that made up the test. Cronbach's alpha coefficient was 0.584 for the low-to-intermediate group and 0.781 for the advanced group, which was computed on the basis of percentage of accuracy.

The dependent measures of the cloze test were mean accuracy and response time. Only the response time of *correct* responses were analysed (Jiang 2013). To explore L2 usage-based categories of article cues, scatterplots were generated for a descriptive graphic illustration of the cue categorisation in the two proficiency groups. The scatterplots' X-axis was set as the logarithmic group mean accuracy of article cues and the Y-axis was the logarithmic group mean response time. Then two hierarchical clustering analyses were performed in R (R Core Team 2013) to explore the emerging categories in learners' interlanguage article usage in the two proficiency groups, respectively. The clustering criterion was set as accuracy, because accuracy in the current offline cloze task was a more reliable measure of learner performance. To address the second research question, accuracy and response time data were analysed using SPSS (version 21.0, IBM Corp.). The specific inferential statistics tests are described in the following Results sections.

7. Results

The low-to-intermediate proficiency learners showed a mean accuracy of 0.48 (SD = 0.07) and a mean response time of 6.64 second (SD = 1.36). The advanced proficiency group had a mean accuracy of 0.68 (SD = 0.07) and a mean response time of 6.18 s (SD = 0.84). Univariate analyses of variance (ANOVAs) showed significant group differences in mean accuracy ($F(1, 101) = 207.729, p < 0.0001, \eta^2 p = 0.673$) and mean response time ($F(1, 101) = 4.569, p = 0.035, \eta^2 p = 0.043$).

² Generic use of the definite (e.g., *The cat is a mammal*) and of the indefinite (e.g., *A cat is a mammal*) are not included in the current measurement.

7.1. Emergent Categories in the Low-to-Intermediate Proficiency Group

Figure 2 presents the scatterplot of article cues (accuracy by response time) in the low-to-intermediate group. The majority of the target cues are concentrated in the centre of the scatterplot due to relatively similar mean accuracies (in between 0.40 and 0.50). Only a few cues scored either much higher or much lower in accuracy. Low-to-intermediate learners performed well on two general cues (cue 21, *second mention with variation* → *the* and cue 22 *part of* → *the*) and one proper name cue (cue 14 *the Street/Road/Avenue of XX* → *the*), all of which are definite cues. They showed poor and random performance on three idiosyncratic zero article cues, i.e., cue 10, *exceptions of construction names* → \emptyset , cue 18, *political/military institution as adjective* → \emptyset , and cue 23, *disease names* → \emptyset .

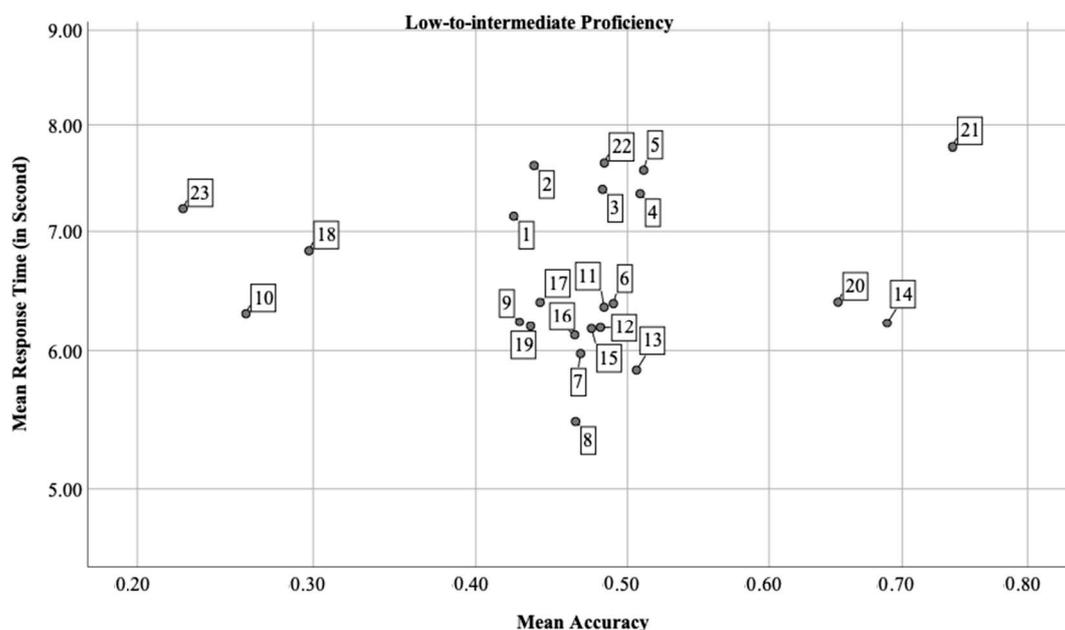


Figure 2. Scatterplot of cues (accuracy by response time) in the low-to-intermediate group.

The hierarchical clustering analysis revealed two macro-level clusters with three categories (see Figure 3). At the bottom of the horizontal dendrogram is a large cluster with Category 1 and at the top half is another cluster with Categories 2 and 3. Table 2 illustrates the specific article cues that fall into the emergent categories. Category 1 is essentially a definite article category (except for cue 20, which was identified as a separate cue from all the sub-groupings in the category), whereas Categories 2 and 3 are non-definite article categories that include zero article and indefinite article cues. Category 2 is the category for general non-definite article cues (except cue 23), whereas Category 3 has all the idiosyncratic zero article cues.

Within Category 1, there are some sub-groupings such as (3, 5) (7, 9) (12, 14), (21, 22) which are distributed and not strong enough to form a separate category but bear some functional and structural similarities. Cue 3, *plural with post-modifiers* → *the*, and Cue 5, *singular countable with post-modifiers* → *the*, are two general definite cues in which unique identification is achieved via the structure of post-modification. Cue 21, *second mention with variation* → *the*, and cue 22, *part of* → *the*, are two general cues in which uniqueness is defined or inferred in the discourse frame. Cue 7, *geographical feature names* → *the*, and Cue 9, *construction names* → *the*, are two definite cues of proper names. Cue 12, *the University/College of XX* → *the*, and Cue 14, *the Street/Road/Avenue of XX* → *the*, are proper name cues that adopt the same *of*-structure. The sub-clustering in Category 1 reveals the low-to-intermediate learners’ effort to make sense of the polysemous use of the definite article through drawing local analogies between cues with structural and functional similarities and making differentiations.

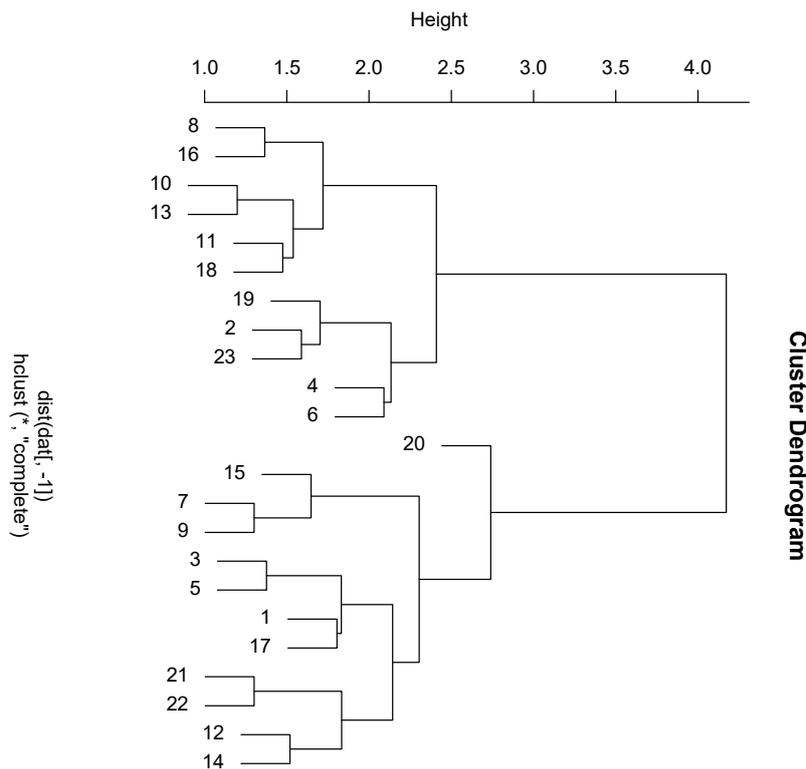


Figure 3. Article cue cluster dendrogram in the low-to-intermediate proficiency group.

Table 2. Usage-based categories of article cues in low-to-intermediate proficiency group.

Category	Cue No.	Cue Name
Category 1	1	non-countable with post-modifiers → the
	3	plural with post-modifiers → the
	5	singular countable with post-modifiers → the
	7	geographical feature names → the
	9	construction names → the
	12	the University/College of XX → the
	14	the Street/Road/Avenue of XX → the
	15	the architecture of XX → the
	17	political/military institution → the
	20	'go to' recreation activity → a/an
Category 2	2	non-countable → ∅
	4	plural → ∅
	6	singular countable → a/an
	19	'go to' habitual location → ∅
	23	disease names → ∅
Category 3	8	exceptions of geographical feature names → ∅
	10	exceptions of construction names → ∅
	11	XX University/College → ∅
	13	XX Street/Road/Avenue → ∅
	16	XX's architecture → ∅
	18	political/military institution as adjective → ∅

7.2. Emergent Categories in the Advanced Proficiency Group

In contrast to the low-to-intermediate group’s scatterplot (Figure 2), the scatterplot of article cues for the advanced learners (Figure 4) was much more distributed. The big concentrated cluster

at the centre disappeared and was replaced by several small clusters that can largely be visually segmented into three accuracy levels, i.e., 0.40–0.60, 0.60–0.80, 0.80–1.00. The two general cues (cue 21 *second mention with variation* → *the* and cue 22 *part of* → *the*) that were performed well among the low-intermediates have been mastered by the advanced learners. This group also scored high on cue 1 *non-countable with post-modifiers* → *the*, cue 3 *plural with post-modifiers* → *the*, and cue 5 *singular countable with post-modifiers* → *the*, which indicates that advanced learners have learned to interpret definiteness from relative clauses and preposition phrases as post-modifier structures. Yet, they continued to perform poorly and almost random on some idiosyncratic zero article cues, i.e., cue 8 *exceptions of geographical feature names* → \emptyset , cue 10 *exceptions of construction names* → \emptyset , and cue 18 *political/military institution as adjective* → \emptyset . Idiosyncratic cues that require the zero article are clearly the most difficult type of article cues for both low-to-intermediate and advanced Chinese EFL learners.

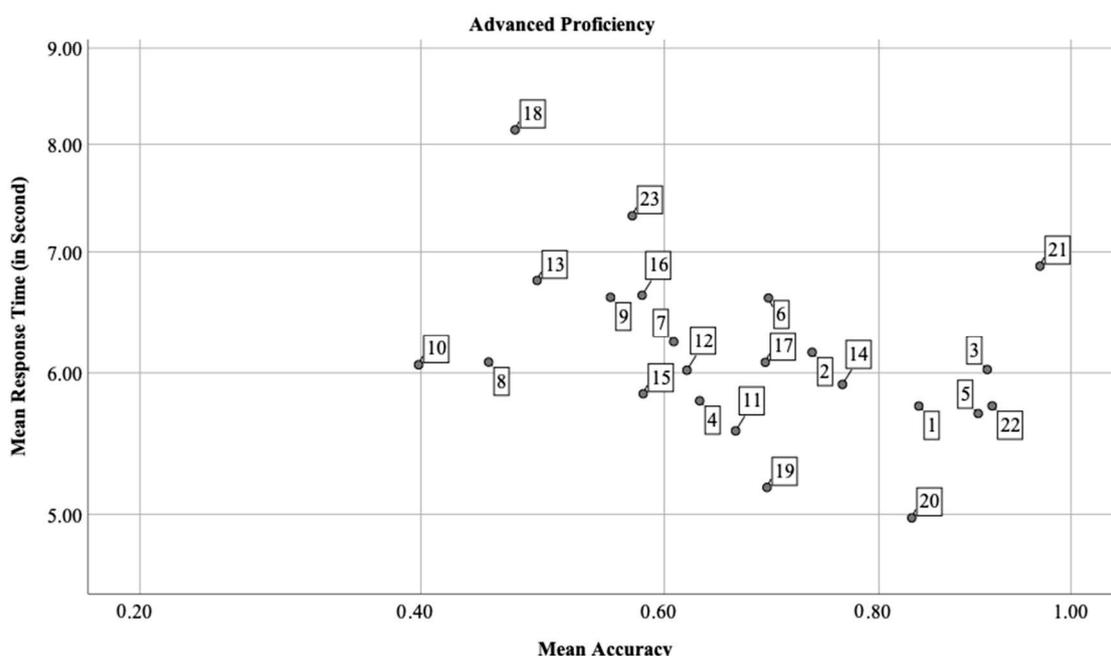


Figure 4. Scatterplot of cues (accuracy by response time) in the advanced group.

The hierarchical clustering analysis in the advanced proficiency group yielded two macro-level clusters with four categories (see Figure 5). The bottom of the horizontal dendrogram is a large cluster with Categories 1, 2, and 4, and at the top is a smaller cluster with Category 3. Table 3 presents the article cues clustered into the four emergent categories.

There was a striking similarity between the L2 usage-based categorizations in the two proficiency groups. The article cues categorised into Categories 2 and 3 in the two groups were identical. The only group difference was that the large cluster of Category 1 in the low-to-intermediate group was restructured to two separate categories (Categories 1 and 4) in the advanced group’s usage-based categorisation, leaving Category 1 with all the general definite article cues (except cue 20) and Category 4 with all the idiosyncratic definite article cues.

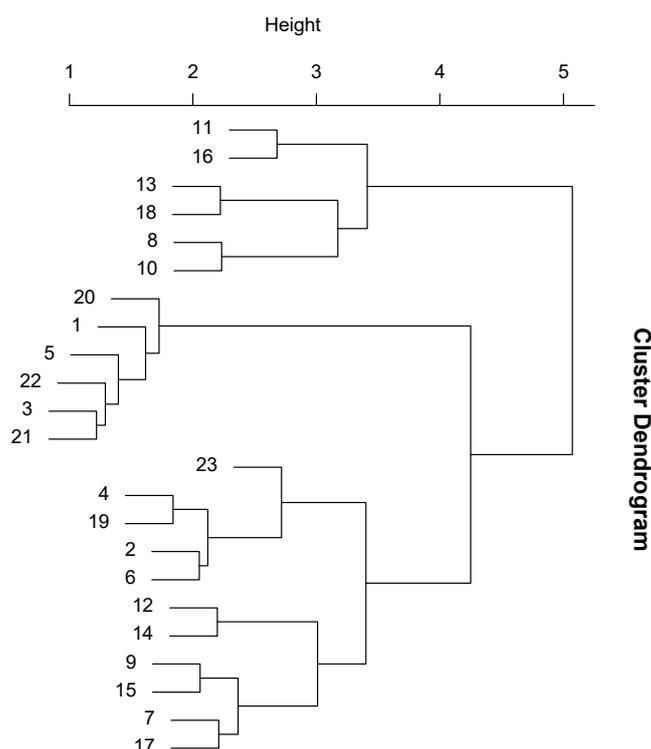


Figure 5. Article cue cluster dendrogram in the advanced proficiency group.

Table 3. Usage-based categories of article cues in advanced proficiency group.

Category	Cue No.	Cue Name
Category 1	1	non-countable with post-modifiers → the
	3	plural with post-modifiers → the
	5	singular countable with post-modifiers → the
	20	'go to' recreation activity → a/an
	21	second mention with variation → the
	22	part of → the
Category 2	2	non-countable → ∅
	4	plural → ∅
	6	singular countable → a/an
	19	'go to' habitual location → ∅
	23	disease names → ∅
Category 3	8	exceptions of geographical feature names → ∅
	10	exceptions of construction names → ∅
	11	XX University/College → ∅
	13	XX Street/Road/Avenue → ∅
	16	XX's architecture → ∅
	18	political/military institution as adjective → ∅
Category 4	7	geographical feature names → the
	9	construction names → the
	12	the University/College of XX → the
	14	the Street/Road/Avenue of XX → the
	15	the architecture of XX → the
	17	political/military institution → the

7.3. Learner Performance on Categories in Proficiency Groups

The descriptive statistics of mean accuracy and response time in the proficiency groups are presented in Table 4. For the low-to-intermediate learners, repeated measures ANOVAs revealed a

main effect of category in mean accuracy ($F(2, 76) = 41.586, p < 0.0001, \eta^2 p = 0.523$) and response time ($F(2, 76) = 9.922, p < 0.0001, \eta^2 p = 0.207$). Bonferroni pairwise comparisons suggested the categorical difference in accuracy was due to the significantly higher accuracy in Category 1 than Categories 2 and 3 ($ps < 0.0001$) and no accuracy difference between Categories 2 and 3 ($p = 1.000$). Their accuracies of performance on the Categories 2 and 3 cues were below chance level, given only three options for them to choose, indicating that the low-to-intermediate learners genuinely struggled with the usage of the zero article and indefinite article cues. The categorical difference in response time was yielded by the significantly faster response time in Category 3 than Categories 1 and 2 ($ps = 0.001$) and no time difference between Categories 1 and 2 ($p = 0.878$).

Table 4. Mean accuracy and response time of usage-based categories.

Proficiency Group	Article Category	Mean Accuracy	SD	Mean Response Time	SD
Low-to-intermediate	1	0.64	0.13	6.76	1.49
	2	0.32	0.17	6.95	1.44
	3	0.29	0.23	6.15	1.51
Advanced	1	0.89	0.06	5.85	0.85
	2	0.67	0.16	6.21	0.98
	3	0.51	0.23	6.54	1.31
	4	0.64	0.21	6.12	1.30

In the advanced group, repeated measures ANOVAs also revealed a main effect of category in mean accuracy ($F(3, 95) = 48.245, p < 0.0001, \eta^2 p = 0.434$) and response time ($F(3, 157) = 6.870, p = 0.001, \eta^2 p = 0.098$). Bonferroni pairwise comparisons suggested several significant categorical differences in accuracy: Category 1 vs. Categories 2, 3, 4 ($ps < 0.0001$), and Category 2 vs. Category 3 ($p < 0.0001$). The categorical difference in response time was due to the significantly faster response time in Category 1 than in Category 2 ($p = 0.020$) and Category 3 ($p < 0.0001$).

8. Discussion

The current study was, to the best of our knowledge, the first empirical research that investigated L2 learners' development of constructional knowledge in the acquisition of the English article construction. Rather than observing the emergent item-based or more abstract patterns in more naturalistically collected learner data, this study adopted quantitative methods to statistically cluster 23 article cues in learners' interlanguage based on learner performance at two proficiency levels in a controlled experimental task. The findings revealed that when learners acquire the polysemous English article construction they demonstrated stronger competence in differentiating individual form-function mappings in the article construction. Learners' performance space of article cues (scatterplots in Figures 2 and 4) moved from a densely concentrated cluster with the majority of cues glued together to a much more scattered space with clearer cue distinctions and local clusters of cues. The change reflected learners' enhanced abilities to analyse cue distinctions and to draw analogy among form-meaning mappings with semantic and structural similarities while analysing cue features. This development of constructional knowledge is shown as a process of doing analysis, differentiation, analogy, and categorisation.

The findings of the hierarchical clustering analyses revealed that L2 learners' patterns of article construction usage were shaped by semantic functions. The emerged clusters did not seem to be based on input frequency, as some clusters contained cues of high frequency and low frequency. For instance, the Category 2 cues observed in both proficiency groups includes cues of very high token frequency (*plural* → \emptyset , *non-countable* → \emptyset , *singular countable* → *a/an*) and of very low frequency (*'go to' habitual location* → \emptyset , *disease names* → \emptyset) based on the English written corpus analysis by Zhao and MacWhinney (2018).

Instead, semantic functions motivated a taxonomic generalisation across related cues. The categories that emerged in both proficiency groups were largely based on semantic function.

Category 1 is the definiteness category, whereas Categories 2 and 3 are the categories for non-definiteness. Learners performed better on the definiteness category than on the non-definiteness categories, suggesting that learners were sensitive to semantic prototypicality. Nominal grounding is prototypically overt. Unique instantiation associated with the definite *the* represents the most salient type of nominal grounding. In particular, cue 21 (*second mention with variation* → *the*) and cue 22 (*part of* → *the*) are two prototypical exemplars of definite reference, encoding anaphoric reference, and inferred uniqueness, respectively (Radden and Dirven 2007), and are the best-performing cues in the two learner groups.

The non-definiteness categories are mostly associated with less prototypical types of grounding such as intrinsic grounding and covert grounding, and other idiosyncratic usages. Low-to-intermediate learners scored the same at a low accuracy on Category 2 and Category 3 cues, indicating that they had overall difficulty interpreting the more peripheral grounding relations. In comparison, advanced learners showed significant improvements on the non-definiteness categories. They scored higher on Category 2 (general non-definite) than on Category 3 (idiosyncratic non-definite) cues. This group difference can be attributed to effects of frequency and semantic transparency because general non-definite cues (e.g., *plural* → \emptyset , *non-countable* → \emptyset) are a lot more frequent in usage and semantically analysable than idiosyncratic non-definite cues (e.g., *XX University/College* → \emptyset ; *XX Street/Road/Avenue* → \emptyset), which are much less frequent and semantically opaque. With increasing proficiency, L2 learners showed development on the weaker (non-prototypical) domains of constructional knowledge as a function of input frequency and semantic analysability.

Many of the previous studies on verb-based constructions reported that input frequency was the most important predictor of L2 learning trajectory (Azazil 2020; Ellis and Ferreira-Junior 2009; Römer and Berger 2019). Input frequency and prototypicality often overlap, as “the greater the token frequency of an exemplar, the more it contributes to defining the category, and the greater the likelihood that it will be considered the prototype” (Ellis and Collins 2009, p. 331). But in the case of the English article construction, some definiteness cues encode prototypical grounding relations but may not have high token frequency. Some non-definiteness cues such as *non-countable* → \emptyset represent the most frequent article usage in English but only encode peripheral grounding relations.

The current study showed that input frequency played a role in influencing learners’ performance. General cues have much higher frequency in the English language than idiosyncratic cues (Zhao and MacWhinney 2018). Correspondingly, our advanced learners’ mean accuracy on general cues (in Categories 1 and 2) was generally higher than that of idiosyncratic cues (in Categories 3 and 4). More importantly, the current study showed that semantic prototypicality was the most important factor that affected the categorisation of L2 article constructional knowledge. Wulff et al. (2009) found that beginning L2 learners’ acquisition of English tense-aspect was influenced by the input frequency of the verbs and also the verbs’ prototypicality of lexical semantics. Thus, we can infer that centrality of meaning plays a more important role in influencing learners’ usage-based acquisition of semantically complex systems such as articles, tense-aspect, modality, and prepositions than that of structurally complex systems such as complex clauses. The fact that semantic prototypicality rather than input frequency largely determined the categorisation of L2 article constructional knowledge could also be due to the nature of instructed L2 learning. Explicit grammatical knowledge can demonstrate a powerful effect that can alter the predicted usage-based trajectory of language acquisition (Roehr-Brackin 2014). Chinese learners traditionally get a great deal of explicit grammar and limited authentic input. Explicit instruction on English articles is typically structured by presenting the differences between the definite article and the non-definites (Master 1990, 2002) and can have durable effects on implicit knowledge (Akakura 2012). The binary approach to article instruction may have a powerful effect on learners’ mental categorisation of L2 constructional knowledge.

Restructuring of the cognitive interpretation of functional categories is a necessary step in language emergence and acquisition. The most salient difference in categorisation between the two proficiency groups is that advanced learners demonstrated an increased sensitivity to semantic idiosyncrasy. The emergence of Category 4 means that advanced learners learned to differentiate general cues and

idiosyncratic cues within the definites. Low-to-intermediate learners failed to do so. It was likely that the low-to-intermediate learners assigned the definite article as the default for all idiosyncratic cues as a function of frequency distribution. It is true that many idiosyncratic cues such as 'geographical feature names → the', 'construction names → the', and 'political/military institution → the' typically take the definite article, while only a small set of exceptional uses of these cues take the zero article (Zhao and MacWhinney 2018). The association between entity names and the definiteness interpretation may have overshadowed the learning of the exceptional instances. As learners moved to advanced proficiency, they have clearly differentiated articles' general usages from the idiosyncratic usages. They were aware that not all entity names take the definite article. However, their similar performance on idiosyncratic *the* cues and idiosyncratic \emptyset cues suggested that they had not been able to attune to the idiosyncrasy of entity names in their specific contexts of usage. The lack of contextualised constructional knowledge (the ability to associate constructions to particular contexts) constituted one of the major knowledge gaps between native speakers and L2 learners (Ellis et al. 2008; Jach 2018).

In short, the findings of the current study reveal the process that L2 learners went through in acquiring the English article construction. The results support the usage-based assumptions on construction learning. The findings of the study have to be seen in light of some limitations. The study investigated one learner group (Chinese-L1) by using one off-line processing-based selection task. Future studies can sample other learner populations, including learners in English as a second language (ESL) contexts, and from other L1 backgrounds. It will be interesting to compare learners from article-less L1 backgrounds with learners whose L1s have an article system. It is worth exploring whether the existence of an article system in the L1 influences learners' mental categorisation of L2 constructional knowledge. Meanwhile, other varieties of processing and production tasks can be employed to measure learners' article usage. The current cloze test created obligatory sentential contexts that allow only one appropriate article choice. As a result, optional article choices have been purposefully avoided, such as in "Jennifer bought the/ \emptyset earrings she had read about". Future research can use on-line or off-line acceptability judgment tasks or self-paced reading tasks (Ahn 2019) to measure learners' sensitivity to the optionality in article usage. We also need longitudinal productive data to explore how learners gradually form the target schematic representations of the article construction and how input frequency and semantic prototypicality influence this process over time.

9. Conclusions

Structural and generative theories assume the existence of abstract rules and language speakers' ability to apply the rules to novel productions. The evidence brought forward in the current usage-based study on the English article construction showed that abstract patterns can be accomplished through local analogies to existing exemplars, without reference to abstract rules. Through observing changes in the emergence of L2 categorisation, we found that learners formed separate functional categories for definiteness (prototypical grounding relation) and non-definiteness (non-prototypical grounding relation). The formation of categories is not achieved through some top-down rules, but through local analogies, differentiation, and abstraction. With increased proficiency, learners demonstrated stronger competence in differentiating individual form-function mappings and forming abstract categories. Restructuring of the mental categories is an important way of regularization that learners go through to acquire polysemous and semantically complex systems like articles.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Table A1. Target article cues with metalinguistic explanations and exemplars.

Cue No.	Cue Name	Metalinguistic Explanation and Selected Example
1	<i>non-countable with post-modifiers → the</i>	Use <i>the</i> when a non-countable noun is post-modified by a relative clause or a prepositional phrase. (<i>They try to protect <u>the</u> land they own at all cost.</i>)
2	<i>non-countable → Ø</i>	Use \emptyset with unmodified non-countable nouns. (<i>They try to protect <u>Ø</u> land at all cost.</i>)
3	<i>plural with post-modifiers → the</i>	Use <i>the</i> when the plural noun is post-modified by a relative clause or a prepositional phrase and is uniquely identifiable. (<i>I have found you <u>the</u> people who used to work as Chinese-English translators.</i>)
4	<i>plural → Ø</i>	Use \emptyset with plural nouns unless they are uniquely identifiable. (<i>Do you need <u>Ø</u> people who can translate English to Chinese?</i>)
5	<i>singular countable with post-modifiers → the</i>	Use <i>the</i> when the singular countable noun is post-modified by a relative clause or a prepositional phrase and is uniquely identifiable. (<i>I thanked <u>the</u> travel agency who helped me find a cheap air ticket.</i>)
6	<i>singular countable → a/an</i>	Use <i>a/an</i> when the singular countable noun is not made concrete or instantiated by any modifier. (<i>I am working with <u>a</u> travel agency to find a cheap air ticket.</i>)
7	<i>geographical feature names → the</i>	Use <i>the</i> with major geographic features such as rivers, oceans, seas, and deserts. (<i><u>the</u> Ohio River, <u>the</u> Pacific Ocean, <u>the</u> Red Sea, <u>the</u> Sahara Desert</i>)
8	<i>exceptions of geographical feature names → Ø</i>	Use \emptyset with the names of individual lakes, bays, and mountains. (<i><u>Ø</u> Lake Michigan, <u>Ø</u> Hudson Bay, <u>Ø</u> Mount Whitney</i>)
9	<i>construction names → the</i>	Use <i>the</i> with the names of hotels, theatres, bridges, and buildings. (<i><u>the</u> Hilton Hotel, <u>the</u> Majestic Theatre, <u>the</u> Brooklyn Bridge, <u>the</u> Empire State Building</i>)
10	<i>exceptions of construction names → Ø</i>	Use \emptyset with the names of halls, stadiums, and hospitals. (<i><u>Ø</u> Carnegie Hall, <u>Ø</u> Yankee Stadium, <u>Ø</u> Hillsdale Hospital</i>)
11	<i>XX University/College → Ø</i>	Use \emptyset when the name of a university has this structure: XX University. (<i><u>Ø</u> Harvard University</i>)
12	<i>the University/College of XX → the</i>	Use <i>the</i> when the name of a university has this structure: the University of XX. (<i><u>the</u> University of Edinburgh</i>)
13	<i>XX Street/Road/Avenue → Ø</i>	Use \emptyset when the names of streets, roads, avenues, lanes, or boulevards have this structure: XX Street/Road/Avenue/Boulevard. (<i><u>Ø</u> Fifth Avenue</i>)
14	<i>the Street/Road/Avenue of XX → the</i>	Use <i>the</i> when the names of streets, roads, avenues, lanes, or boulevards have this structure: the Street/Road/Avenue/Boulevard of XX. (<i><u>the</u> Avenue of Stars</i>)
15	<i>the architecture of XX → the</i>	Use <i>the</i> when describing constructions that are made specific through the addition of post-modifiers. (<i><u>the</u> Statue of Liberty</i>)
16	<i>XX's architecture → Ø</i>	Use \emptyset when the names of constructions are modified by possessive nouns (such as London's). (<i><u>Ø</u> New York's Statue of Liberty</i>)
17	<i>political/military institution → the</i>	Use <i>the</i> when established institutions (e.g., political and military institutions) are used alone. (<i><u>the</u> World Bank</i>)
18	<i>political/military institution as adjective → Ø</i>	Use \emptyset when established institutions (e.g., political and military institutions) are used as adjectives to modify other nouns. (<i><u>Ø</u> World Bank loans</i>)

Table A1. Cont.

Cue No.	Cue Name	Metalinguistic Explanation and Selected Example
19	'go to' habitual location → ∅	Use ∅ when the word 'go' is used with a habitual location or a habitual transport method. (<i>go to ∅ school</i>)
20	'go to' recreation activity → a/an	Use a/an when the word 'go' is used with recreational activities. (<i>go for a dance</i>)
21	second mention with variation → the	Use <i>the</i> when the noun has already been mentioned before, and the second way in which it is mentioned is slightly different from the first. (<i>I saw a peacock at the zoo. The bird had beautiful feathers.</i>)
22	part of → the	Use <i>the</i> when describing an object that is a unique part of some overall scene, event, or object being discussed. (<i>I'm returning this coat for a refund. The zipper broke after one day.</i>)
23	disease names → ∅	Use ∅ with the names of diseases (except the flu, the measles, and the mumps). (<i>His uncle has ∅ cancer.</i>)

References

- Ahn, Hyunah. 2019. From interlanguage grammar to target grammar in L2 processing of definiteness as uniqueness. *Second Language Research*. [CrossRef]
- Akakura, Motoko. 2012. Evaluating the effectiveness of explicit instruction on implicit and explicit knowledge. *Language Teaching Research* 16: 9–37. [CrossRef]
- Ambridge, Ben, and Elena Lieven. 2015. A constructivist account of child language acquisition. In *The Handbook of Language Emergence*. Edited by Brian MacWhinney and William OGrady. Malden: Wiley-Blackwell, pp. 478–510.
- Azazil, Lina. 2020. Frequency effects in the L1 acquisition of the catenative verb construction—evidence from experimental and corpus data. *Cognitive Linguistics* 31: 417–51. [CrossRef]
- Bickerton, Derek. 1981. *Roots of Language*. Ann Arbor: Karoma.
- Butler, Yuko Goto. 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: 451–80. [CrossRef]
- Bybee, Joan. 2008. Usage-based grammar and second language acquisition. In *Handbook of Cognitive Linguistics and Second Language Acquisition*. Edited by Peter Robinson and Nick C. Ellis. New York and London: Routledge, pp. 216–36.
- Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Cole, Tom. 2000. *The Article Book: Practice Toward Mastering a, an, and the*, 2nd ed. Ann Arbor: The University of Michigan Press.
- Croft, William, and Alan D. Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Crossley, Scott, and Thomas Lee Salisbury. 2011. The development of lexical bundle accuracy and production in English second language speakers. *International Review of Applied Linguistics in Language Teaching* 49: 1–26. [CrossRef]
- Davies, Mark. 2008. *The Corpus of Contemporary American English: 520 Million Words, 1990–Present*. Provo: Brigham Young University, Available online: <http://corpus.byu.edu/coca> (accessed on 1 June 2017).
- Ellis, Nick C. 2003. Constructions, chunking, and connectionism: The emergence of second language structure. In *Handbook of Second Language Acquisition*. Edited by Catherine Doughty and Michael H. Long. Oxford: Blackwell, pp. 33–68.
- Ellis, Nick C. 2006. Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, saliency, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics* 27: 164–94. [CrossRef]
- Ellis, Nick C. 2008. Usage-based and form-focused language acquisition: The associative learning of constructions, learned attention, and the limited L2 end state. In *Handbook of Cognitive Linguistics and Second Language Acquisition*. Edited by Peter Robinson and Nick C. Ellis. New York: Routledge, pp. 372–405.

- Ellis, Nick C. 2016. Online processing of verb-argument constructions: Lexical decision and meaningfulness. *Language and Cognition* 8: 391–420. [CrossRef]
- Ellis, N. C., and Laura Collins. 2009. Input and second language acquisition: The roles of frequency, form, and function. *Modern Language Journal* 93: 329–35. [CrossRef]
- Ellis, Nick C., and Fernando Ferreira-Junior. 2009. Construction learning as a function of frequency, frequency distribution, and function. *Modern Language Journal* 93: 370–85. [CrossRef]
- Ellis, Nick C., Rita Simpson-Vlach, and Carson Maynard. 2008. Formulaic language in native and second language speakers: Psycholinguistic, corpus linguistics, and TESOL. *TESOL Quarterly* 42: 375–96. [CrossRef]
- Eskildsen, Søren W. 2008. Constructing another language: Usage-based linguistics in second language acquisition. *Applied Linguistics* 30: 335–57. [CrossRef]
- Eskildsen, Søren W. 2012. L2 negation constructions at work. *Language Learning* 62: 335–72. [CrossRef]
- Eskildsen, Søren W. 2015. What counts as a developmental sequence? Exemplar-based L2 learning of English questions. *Language Learning* 65: 33–62. [CrossRef]
- Evans, Vyvyan, and Melanie Green. 2006. *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- Fodor, Janet Dean, and Ivan A. Sag. 1982. Referential and quantificational indefinites. *Linguistics and Philosophy* 5: 355–98. [CrossRef]
- García Mayo, Maria del Pilar, and Roger Hawkins. 2009. *Second Language Acquisition of Articles: Empirical Findings and Theoretical Implications*. Amsterdam and Philadelphia: John Benjamins.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, Adele E. 1999. The emergence of the semantics of argument structure constructions. In *The Emergence of Language*. Edited by Brian MacWhinney. Mahwah: Lawrence Erlbaum, pp. 197–212.
- Goldberg, Adele E. 2019. *Explain me this: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton and Oxford: Princeton University Press.
- Hawkins, Roger, Saleh Al-Eid, Ibrahim Almahboob, Panos Athanasopoulos, Rangsiya Chaengchenkit, James Hu, Mohammad Rezai, Carol Jaensch, Yunju Jeon, Amy Jiang, and et al. 2006. Accounting for English article interpretation by L2 speakers. *EUROSLA Yearbook* 6: 7–25. [CrossRef]
- Huebner, Thom. 1983. *A Longitudinal Analysis of the Acquisition of English*. Ann Arbor: Karoma.
- Ionin, Tania, Heejeong Ko, and Kenneth Wexler. 2004. Article semantics in L2 acquisition: The role of specificity. *Language Acquisition* 12: 3–69. [CrossRef]
- Jach, Daniel. 2018. A usage-based approach to preposition placement in English as a second language. *Language Learning* 68: 271–304. [CrossRef]
- Jespersen, Otto. 1933. *Essentials of English Grammar*. Boston: George Allen & Unwin.
- Jiang, Nan. 2013. *Conducting Reaction Time Research in Second Language Studies*. New York: Routledge/Taylor & Francis.
- Lakoff, George. 1987. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Stanford: Stanford University Press, vol. I.
- Langacker, Ronald W. 1991. *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Berlin and New York: Mouton de Gruyter.
- Langacker, Ronald W. 2008. *Cognitive Grammar: A Basic Introduction*. New York: Oxford University Press.
- Li, Peiwen, Søren W. Eskildsen, and Teresa Cadierno. 2014. Tracing an L2 learner's motion constructions over time: A usage-based classroom investigation. *Modern Language Journal* 98: 612–28. [CrossRef]
- Lieven, Elena, Julian M. Pine, and Gillian Baldwin. 1997. Lexically-based learning and early grammatical development. *Journal of Child Language* 24: 187–219. [CrossRef]
- MacWhinney, Brian. 1987. The Competition Model. In *Mechanisms of Language Acquisition*. Edited by Brian MacWhinney. Hillsdale: Lawrence Erlbaum, pp. 249–308.
- Master, Peter. 1990. Teaching the English articles as a binary system. *TESOL Quarterly* 24: 461–78. [CrossRef]
- Master, Peter. 2002. Information structure and English article pedagogy. *System* 30: 331–48. [CrossRef]
- R Core Team. 2013. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

- Radden, Günter, and René Dirven. 2007. *Cognitive English Grammar*. Amsterdam: John Benjamins.
- Roehr-Brackin, Karen. 2014. Explicit knowledge and processes from a usage-based perspective: The developmental trajectory of an instructed L2 learner. *Language Learning* 64: 771–808. [[CrossRef](#)]
- Römer, Ute, and Cynthia M. Berger. 2019. Observing the emergence of constructional knowledge: Verb patterns in German and Spanish learners of English at different proficiency levels. *Studies in Second Language Acquisition* 41: 1089–110. [[CrossRef](#)]
- Sung, Min-Chang, and Hyunwoo Kim. 2020. Effects of verb-construction association on second language constructional generalizations in production and comprehension. *Second Language Research*. [[CrossRef](#)]
- Tarone, Elaine, and Betsy Parrish. 1988. Task-related variation in interlanguage: The case of articles. *Language Learning* 38: 21–44. [[CrossRef](#)]
- Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press.
- Trenkic, Danijela. 2008. The representation of English articles in second language grammars: Determiners or adjectives? *Bilingualism: Language and Cognition* 11: 1–18. [[CrossRef](#)]
- Verspoor, Marjolijn H., and Nguyen Thi Phuong Huong. 2008. Cognitive Grammar and teaching English articles to Asian students. In *Du Fait Grammatical au Fait Cognitive*. Edited by Jean-Remi Lapaire, Guillaume Desagulier and Jean-Baptiste Guignard. Bordeaux: Presses Universitaires de Bordeaux, vol. 1, pp. 249–68.
- Wulff, Stephanie, Nick C. Ellis, Ute Römer, Kathleen Bardovi-Harlig, and Chelsea J. Leblanc. 2009. The acquisition of tense-aspect: Converging evidence from corpora and telicity rating. *Modern Language Journal* 93: 354–69. [[CrossRef](#)]
- Zhao, Helen, and Brian MacWhinney. 2018. The instructed learning of form-function mappings in the English article system. *Modern Language Journal* 102: 99–119. [[CrossRef](#)]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).