

*The use of articles in L2 English:
A phraseological perspective*

Justyna Leśniewska

Jagiellonian University, Cracow, Poland

justyna.lesniewska@uj.edu.pl

Abstract

While it is a well-known fact that speakers of article-less mother tongues, such as Polish, experience problems with articles in English, this study seeks to investigate the problem from a different perspective. Namely, it poses the question of whether the correct use of the article system of the L2 is indeed a purely grammatical task (as it is universally perceived), or whether the correct use of articles is to some extent aided by the mechanisms that underlie the formulaic character of language. The study was conducted with 90 Polish upper-intermediate and advanced users of L2 English, who completed a test on article use, which made it possible to compare patterns of article use between contexts of different collocational strength (defined in terms of the frequency of occurrence in a corpus). The statistically higher success rates for article use in high-frequency collocations (with the grammatical "rule" being the same) indicate that phraseological aspects of language use may indeed play a role in what is usually perceived as the correct application of grammatical rules.

Keywords: articles in ESL; determiners; formulaicity; phraseological aspects of second language acquisition

1. Introduction

1.1. Articles in ESL

Articles are notoriously difficult to acquire for learners of English as a second language (ESL), especially for speakers whose L1 lacks articles. The problem has been extensively researched (a brief overview is provided below), but almost exclusively as an aspect of the development of learners' syntactic competence. The use of articles is well-established as a grammatical topic, and most ESL textbooks and grammars give extensive sets of "rules" which specify how articles are used (Holmes, 1988; Hsu, 2008). The only exceptions are certain untypical uses of articles which occur in idiomatic expressions. Those uses, usually labelled as "fixed" or "idiomatic" (Holmes & Moulton, 1993; Orlando, 2009), include such expressions as *living hand to mouth*, *all of a sudden*, *in front vs. in the back*, or *game of cat and mouse*. This category is made up only of those uses of the definite, indefinite or zero article which cannot be explained by the rules. Consequently, idiomaticity is simply a convenient label for those cases of article use which fall outside the syntactic regularity. The regular uses, which constitute the majority of article uses, are seen as governed by extensive lists of "rules." Major rules, which involve countability, definiteness and specificity (Ekiert, 2007), are accompanied by minor rules, often called "rules of thumb" (Faerch, 1986) in pedagogical grammars. One example is the principle that no articles should be used with names of cities. Those "rules of thumb" always (rather frustratingly for learners) come with exceptions (e.g., *The Hague*).

Therefore, being able to use articles correctly in English is generally seen as the result of the eventual mastery of rules of grammar, with the exception of the idiomatic or fixed uses, which have to be memorized. However, in view of the growing evidence that language processing is to a considerable extent formulaic (see the overview below), this article explores the possibility that even those uses of articles which appear to be rule-governed, that is, syntactically regular, may in fact be aided by the mechanisms which are responsible for formulaicity in language use.

1.2. Phraseological aspects of L2 acquisition and use

It is a widely recognized fact nowadays that formulaicity plays an important role in language processing. The now widely made distinction between two modes of language processing is often attributed to John Sinclair (1987, 1991), who distinguished between the "open choice principle" and the "idiom principle." The first one is

a way of seeing language text as the result of a very large number of complex choices. At each point where a unit is completed (a word or a phrase or a clause), a large range of choice opens up, and the only restraint is grammaticalness . . . It is often called a 'slot-and-filler' model . . . At each slot, virtually any word can occur. (Sinclair, 1991, p. 109)

The open choice principle operates, therefore, like traditional grammar-centered models of language: There are a number of syntagmatic choices available for each slot along the paradigm. On the other hand, the idiom principle holds that "a language user has available to him or her a large number of semi-pre-constructed phrases that constitute single choices, even though they might appear to be analysable into segments" (Sinclair, 1991, p. 110).

A distinction of this kind was made even earlier, in a seminal paper by Pawley and Syder (1983), which sought to explain what the authors saw as "two puzzles for linguistic theory": native-like selection and native-like fluency. What they claimed was that "fluent and idiomatic control of a language rests to a considerable extent on knowledge of a body of 'sentence stems' which are 'institutionalized' or 'lexicalized'" (p. 191). The ability to recall larger chunks from memory does not mean that the chunks are not analysable into segments.

The idiom principle can be seen as enabled by chunking, a concept deriving from psychology (Chase & Simon, 1973; Gobet et al., 2001), which has found support in the field of language. Chunking occurs at all levels of language (Nation, 2001, p. 319). Complex words, for example, are usually processed as wholes, not as combinations of individual morphemes. Morphemes, in turn, are processed as units, not as sequences of individual phonemes. Chunking enables the grouping of smaller units into larger wholes, but also the analysis of the wholes into segments if needed. The main advantage of chunking appears to be reduced processing time, and, therefore, faster language comprehension and production (Ellis, 2001). The disadvantage of chunking is that it takes up storage space: Language users need to store chunks (combinations of items) in addition to the components that are already stored separately (Nation, 2001, pp. 320-321). Therefore, it makes sense that high-frequency items are stored as chunks, reducing processing time, since they occur often enough to make up a large proportion of the overall language produced. Low-frequency items, on the other hand, do not "deserve" separate storage space, they are recreated "by rule" when needed (Aitchison, 1987).

This line of thinking was further developed by Wray (2002a, 2002b), who argued that "formulaic processing is the default," and that "construction out of, and reduction into, smaller units by rule occurs only as necessary" (Wray, 2002b, p. 119). This is an explanation for the existence of irregularities in language:

If we only create and understand utterances by applying rules to words and morphemes, it is difficult to see why irregularity should be tolerated, let alone why an

item or construction should progress from regular, to marked, to antiquated, to a fossilized historical relic. (p. 118)

Hoey's concept of *lexical priming* (2005) is based on the idea that lexical patterns are responsible for the structure of language, and that grammar is merely an outcome of the pervasiveness of collocation. Hoey presents collocation as a psychological concept: The recurrent co-occurrence of words is enabled by priming:

As a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with the contexts and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context. (p. 8)

Grammar emerges from recurrent patterns of word combinations.

A strong view of the importance of collocational competence has been advocated by Ellis (2001), who argues that language users store chunks of language in long-term memory and acquire the experience of how likely particular items are to co-occur. A crucial role is played by associations between items which are observed to appear in the vicinity of each other. Language users are able to break up the chunks according to the grammar rules of the language, but can produce and comprehend them without reference to those rules. A lot of learning can also be accounted for in terms of learning by association, as a result of encountering certain word combinations.

Recently, the number of studies of various aspects of formulaicity has been growing rapidly. A number of different research strands have been feeding into this trend: traditional phraseological approaches, large-scale corpus analyses of learner language, discourse analysis and historical linguistics, and psycholinguistic investigations into the mechanisms which underlie formulaicity. Despite this fast-growing body of research, the emerging picture is far from clear due to a plethora of different phenomena that are investigated in connection with formulaicity, under a host of names, such as *formulaic sequences*, *multi-word expressions*, *lexical bundles*, *interactional routines*, *language chunks*, and so on.¹

I shall follow here those authors (e.g., Meunier, 2012) who use the term *formulaicity* as an umbrella term to encompass a wide range of language phenomena related to the fact that language production is not based solely on the use of individual lexical items according to syntactic rules.

¹ For an informative and inspiring review of the current state of research on formulaic language, see Wray (2012); for an excellent overview of the issues involved, see Weinert (2010).

It seems likely that formulaicity in language is connected to the frequency of occurrence of certain phrases. While the relationship between frequency of occurrence and formulaicity is by no means a straightforward and simple one, and frequency is by no means the only determinant of formulaicity (see Wray, 2012, for a discussion), there is definitely some relationship between the two (Ellis, 2012). Support for this view comes from recent studies on the speed of processing of lexical bundles, which clearly show frequency effects (Tremblay & Baayen, 2010; Tremblay, Derwing, Libben, & Westbury, 2011).

Even though it has been argued that the psychological validity of the idiom principle has not been empirically proven to a satisfying extent (Siyanova-Chanturia & Martinez, 2015), the pervasiveness of formulaicity in language, as attested for by the vast body of literature mentioned above, strongly suggests that formulaicity may affect the use of articles in L2 English. Currently the only recognized form of interplay between formulaicity and the use of articles is the existence of some odd uses of articles frozen in idiomatic expressions; but the real extent of that interplay is likely to be much broader. It stands to reason that the idiom principle is to some extent a driver of correct article use. The study reported here was carried out in order to provide an initial exploration of this possibility.

1.3. Research on articles in ESL

Among the various research findings pertaining to article use by learners of English as a second language, perhaps the most robust finding is that article use strongly depends on crosslinguistic factors: Speakers of article-less languages find them much more problematic (e.g., Hawkins et al., 2006; Ionin, Zabizarreta, & Maldonado, 2008; Snape, 2008; Zdorenko & Paradis, 2008). Many studies focus on the problems with articles which learners face in the early stages of L2 acquisition. They note that the indefinite article, in particular, is acquired late, and that beginners tend to overuse the definite article (Bitchener & Knoch, 2010; Huebner, 1983; Young, 1996).

It has also been established that difficulty in using English articles is caused to a great extent by the problems learners face when determining the countability of nouns (Butler, 2002; White, 2009). It has also been demonstrated that more errors in article use occur with abstract nouns than with concrete nouns (Hua & Lee, 2005; Ogawa, 2008). An exploration of learners' article use with abstract nouns (Amuzie & Spinner, 2013) has shown that the level of accuracy in article use is related to more nuanced categories within the abstract noun category, based on the nouns' degree of boundedness, which in turn determines their countability. Another important fact about L2 article use is that articles may remain a problematic area of L2 use even at relatively advanced

levels of proficiency (e.g., Díez-Bedmar & Papp, 2008; Master, 1997; Parrish, 1987), thus becoming a marker of nonnativeness in otherwise proficient output.

This state of affairs is usually attributed to the complicated and elusive nature of the rules governing article use in English. It has been suggested (Shintani & Ellis, 2013) that the main source of difficulty is that articles do not comply with the “one to one principle” (Andersen, 1984), because a single morpheme performs multiple functions. Also, as Master (2002, p. 332) notes, articles occur very frequently, which makes continuous rule application more of a challenge. It is also true that, alongside other function words, articles are normally unstressed and may be perceived by learners as less salient in the input.

A large number of studies on articles in SLA revolve around the role of universal grammar. The functionalist perspective in particular has inspired a large number of studies on L2 article use which investigate the influence of putative syntactic, semantic and discourse universals on the systematicity and variability of interlanguage (e.g., Chaudron & Parker, 1990; Huebner, 1983; Parrish, 1987; Thomas, 1989; Young, 1996). Such studies look at, for example, the encoding of definiteness and specificity, as well as the tendency to mark the topic/comment distinction and the distinction between new, continuous, and reintroduced referents (e.g., Jarvis, 2002).

It needs to be emphasised that virtually the entire body of research on articles deals with the problem from the perspective of syntax, as illustrated by the above review. At the time of the writing of this article, a thorough search for studies that specifically address the issue in question—phraseological effects on the seemingly rule-based use of articles in ESL—yielded just one result, a study by Leńko-Szymańska (2012) which tackles this very problem by means of a corpus-based analysis of learner writing, compared with baseline data from a native speaker corpus. Leńko-Szymańska extracted all cases of three-word combinations including articles from the learner corpus and identified 3-grams (combinations of three words which occur together frequently enough to be classified as lexical bundles) with articles in the native corpus, such as *one of the*, *go to the*, *part of the*, *there is a*, *he was a*, *there is a*, and so on. She then looked at how the use of articles in 3-grams compares to the total number of article uses in the corpora. One very interesting finding which emerged from this study is that, in 3-grams, the definite article occurs much more often (ca. 30% of the uses) than the indefinite article (ca. 17% of the uses). In the learner data, *the* also occurs more often in bundles than *a/an*, across all levels of proficiency. A corpus approach like this one has some limitations: It does not take into consideration correctness—some, probably many, uses of articles in the learner corpus may be incorrect; and it does not provide any data about the use of the zero article. However, this approach has the benefit of clearly showing how the use

of articles in lexical bundles becomes more frequent as proficiency increases. In fact, at advanced levels, the frequency of the use of the definite article in bundles by L2 learners reaches that of native speakers, and for the indefinite article it is actually higher than the native norm. At the same time, the rule-based uses of articles fall below the native norm, even at the advanced level. This finding is extremely interesting: It suggests that there may be a phraseological effect at play affecting the way learners of English use articles since articles are more likely to be used by learners if they are part of a frequent combination of words.

2. Research question and predictions

The research question posed by this study is: Does the idiom principle account to some extent for the correct article use by learners of English? The idiom principle as such cannot be directly observed since it refers to learners' mental processes. As was mentioned above, there are reasons to believe that there is a connection between formulaicity and frequency of occurrence. This study, therefore, makes the assumption that the frequency of occurrence of certain phrases in language is roughly indicative of the mode of processing. Very generally speaking, word combinations which are perceived as "typical" and which occur frequently are more likely to be processed using the idiom principle, while the combinations which are rather rare are likely to be processed in the open-choice manner.

The assumptions formulated above imply that when comparing two contexts of the use of an article in which the relevant grammatical rule for article use is the same, but in one instance the article is included in an open-choice combination of words, and in the other, in a combination generated by the idiom principle, learners should be more successful with the use of the article in the latter case. It is therefore expected that the correctness of the use of articles appearing in frequently used word combinations will be significantly higher than for the same articles appearing in relatively rare combinations. It should be noted that the corpus-based frequency information on the selected word combinations most likely does not correspond precisely to how often the phrases were actually encountered by the particular group of L2 English speakers who participated in this study; however, for the purposes of this study, the assumption was made that there is at least a rough, general correspondence between frequency of occurrence and the likelihood of L2 users encountering a certain word combination.

The above research question was explored in a test-based study involving adult learners of English at the B2/C1 level of the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2011). The learners' L1 was Polish, an article-less language.

3. Method

3.1. Participants

The participants were 90 Polish university students majoring in English or linguistics. The students' placement in groups for their English-as-a-foreign-language classes reflects their level of advancement. For the purpose of the study, this general indication of the level of advancement was considered sufficient. The first group (Group 1, $n = 44$) was at the B2 level of the CEFR, whereas the second, more advanced group (Group 2, $n = 46$) was at the C1 level. All participants were between 20 and 22 years of age, with the mean age slightly lower for Group 1.

3.2. Instrument and procedure

The one-page test used in the study (see Appendix A) consisted of sentences in English from which all the articles had been removed. The participants were asked to put in the missing articles in the right places. The tests included a total of 12 pairs of target items (presented in a mixed-up order) which included exactly the same structures with articles: the definite, the indefinite, and the zero article. The pairs are included in Table 1. Grammatically speaking, the reason for the use of the article was identical in Item A and B of each pair, that is, the same grammatical "rule" applied in both cases. For example, Items 5A and 5B both represent partitive expressions (a type of phrasal quantifiers) used to impose countability on noncount nouns (Quirk & Greenbaum, 1973, p. 67). Items 10A and 10B are both examples of the use of the indefinite article with referents that can be classified as countable, indefinite, and nonspecific (Downing & Locke, 1995, p. 429). However, the items differed in one important aspect: The articles in the A items were included in frequently occurring word combinations, whereas the B items were relatively more of an "open choice" type of word combination.

Table 1 Test target item pairs

Pair no.	Item A	Item B
1	<i>a friend of mine</i>	<i>an acquaintance of mine</i>
2	<i>what a shame</i>	<i>what a remarkable player</i>
3	<i>twice a day</i>	<i>five times a semester</i>
4	<i>the sooner the better</i>	<i>the smaller the pot, the more critical the problem</i>
5	<i>a cup of tea</i>	<i>a spoonful of syrup</i>
6	<i>the day I die</i>	<i>the food I brought</i>
7	<i>help the poor</i>	<i>open to the insured</i>
8	<i>hit (someone) in the face</i>	<i>cut in the hand</i>
9	<i>speak English</i>	<i>learn Kurdish</i>
10	<i>get a job</i>	<i>live in a luxury apartment</i>
11	<i>have kids</i>	<i>eat carbohydrates</i>
12	<i>the centre of attention</i>	<i>the ecology of waterways</i>

3.3. Test preparation

Test preparation relied on a combination of researcher intuition, native speaker judgements, and frequency measures from corpus examination. The most challenging step in the process was to find suitable word combination pairs which would qualify as, respectively, more idiom-principle-driven and more open-choice in character.

A brainstorm session between two linguistics researchers aimed at identifying pairs of word combinations that were perceived by the researchers to be more typical and frequent, versus more open-choice combinations. Intuitive ratings thus formed the basis for the initial selection of word combination pairs. Those pairs were then submitted to two colleagues who were native speakers of English, which led to further elimination of pairs for which there was no inter-judge agreement, the replacement and changes to some word combinations, and a resulting group of 20 word-combination pairs.

The frequency of co-occurrence for those initial intuition-based pairings was verified using two corpora: the British National Corpus (2007; BNC) and the Corpus of Contemporary American English (Davies, 2008-; COCA). Those corpora were deemed adequate due to their size (100 million and 450 million words, respectively) and representative character.² The BNC is made up of written (90%) and spoken (10%) language, and contains texts from a wide range of sources (for example, different kinds of journals, periodicals, newspapers, academic books, popular fiction), in order to represent a wide cross-section of British English. The COCA is also a balanced corpus, made up of texts representing spoken language, fiction, popular magazines, newspapers, and academic texts.

While both corpora were used in the initial search in order to locate suitable pairs for the test, a specific threshold was set with reference to the COCA corpus, the frequency findings from which were considered more reliable because of its larger size. The frequent combination in each pair had to occur at least 40 times more often than its rare counterpart in order to be included in the test. The rare items had a frequency of 0.02 per million words or less. The frequent items had a frequency of 0.18 or more. While it was impossible to determine a perfect set of criteria which could be applied if there was a way to extract the items automatically, this frequency requirement was considered to provide sufficient support for the intuitive judgements.

Not all intuition-based pairs corresponded to corpus-based frequency data, nor were frequency counts always similar in both corpora. Consequently, the 12

² The facts that the BNC is no longer being updated and that the texts come from before 1994 were considered of no importance in the case of this investigation because none of the items that were selected were sensitive to language change or technological advancement.

pairs which showed the most convincing difference in the frequency of occurrence were retained. The full list of items and frequencies is provided in Appendix B.

The 12 pairs (24 target items) were presented hidden among other sentences in the test, which not only helped to provide more context for the target items but also to make the relationship between the pairs of the target items less noticeable. It should be noted that all the articles were removed from all the sentences included in the test. Only some of the missing articles in the test were the actual target items. Since all articles were removed, the number of missing articles was larger than the number of the target items under investigation. For example, in the case of pair no. 3, which tested the use of the indefinite article in expressions of frequency, the more frequent of the two combinations, *twice a day*, appears in Item 7: "By midsummer, herbs and vegetables in containers may need water twice day." Whereas its counterpart, the open-choice combination *five times a semester*, can be found in Item 4: "We meet regularly, five times semester, at departmental meeting." The noun phrase *departmental meeting* also requires an article, but whether the participants inserted it or not was not taken into consideration, as this noun phrase was not one of the target items. Such missing articles outside the target items helped distract the test takers from any pattern in the test design they might be able to discern.

An initial version of the test was piloted with three native and three nonnative speakers of English to ensure that the removal of articles did not create ambiguous or incomprehensible sentences, as well as to check if, for all the target items, all the native speakers always provided the same response. The target items which did not meet this criterion were replaced. Variation in the native speakers' choices of articles in the test outside the target items was considered of no importance. Rare or difficult lexical items were avoided in the test. Care was taken to ensure that both the frequent and the rare combinations of words were composed of "ordinary," relatively frequent lexical items which are expected to be known to learners of English at the intermediate and higher levels. Two experienced teachers of English were consulted about the likelihood that all the words used in the test would be known by our target audience. Teachers were convinced that all items would be known by our test participants, and posttest conversations with a few participants confirmed that no lexical item in the test was new to them. Difficult words, due to their greater length and other difficulty-inducing factors, could affect the processing of the test sentences in ways which could not entirely be controlled for, and they could interfere somehow with article use. For the same reason, the test was composed in such a way as to avoid false cognates (for speakers of L1 Polish) or any ambiguity.

In contrast to most tests on article use, which tend to have the classic format of a cloze test, the instrument used in this study elicited article use in a

slightly different way: The text did not have gaps indicating where the participants needed to provide articles. The rationale for choosing this test design was that it is more similar to the actual use of articles than a cloze test. In a cloze test, the test taker receives a signal that an article may be missing at a specific location. In the case of those tests where the zero article is one of the options, the difference between the two formats is admittedly minor, but it still exists as in the gapped version the test taker is specifically prompted, or encouraged, to consider using an article at a specific place, and in the design employed in this study there is nothing in the test that suggests the need for an article at a specific place.

4. Analysis, results and discussion

In the analysis of the data, dichotomous scores were compiled: 1 point was awarded for inserting a correct article and 0 points for failing to insert an article or for inserting an incorrect one. In the case of test items with the zero article, 1 point was given for not providing an article and 0 points for providing the indefinite or definite article.

The mean item score was calculated for the frequent and for the rare uses for all 90 participants, as well as for each group separately. Those mean item scores are presented in Table 2. A *t* test was performed to compare means.

Table 2 Mean item test scores for frequent and rare uses

	Rare combinations	Frequent combinations	<i>t</i> test
All participants (<i>N</i> = 90)	0.68	0.85	$t = 9.50; p < 0.00001$
Group 1 (less proficient, <i>n</i> = 44)	0.53	0.75	$t = 7.44; p < 0.00001$
Group 2 (more proficient, <i>n</i> = 46)	0.82	0.94	$t = 6.64; p < 0.00001$

As shown by the very low *p* value yielded by the *t* test for all 90 participants as a group, the mean for the frequent combinations was significantly higher than that for the rare ones. In other words, for the same articles and the same grammatical rule, the participants tended to be more successful when using the articles in those combinations that occur more frequently and less likely to be correct with the less frequent combinations.

When analysed separately, both groups showed higher success rates in the case of the frequent combinations than the rare ones. In both cases the difference was statistically significant. However, in the case of Group 2 (more proficient) the difference between rare and frequent combinations was smaller: a

difference in means of only 0.12, compared to a difference of 0.22 in the case of the less proficient Group 1.

The fact that the difference between the rare and frequent combinations became smaller as the level of proficiency increased is understandable: Ultimately, with very advanced language competence, there would be very little difference as articles would be used mostly correctly in all cases for both the frequent and rare word combinations.

It should be noted that the predicted higher means for frequent items were not obtained in the case of all the pairs on the test, as shown in Table 3. Out of the 12 pairs of frequent-versus-rare items, the differences between the means for the frequent items and their rare counterparts as shown by a *t* test was statistically significant (at $p < .05$) for nine item pairs, and not significant for three pairs. The items for which the effect was not observed included: *what a shame* and *what a remarkable player*, *get a job* and *live in a luxury apartment*, and *the centre of attention* and *the ecology of waterways*. In the case of the first pair it is relatively easy to come up with a possible explanation for the observed lack of any effect of formulaicity. While the phrase *what a shame* is definitely much more frequent than the rather open-choice word combination *what a remarkable player*, the nouns *player* and *shame* differ in the degree to which they are countable. First of all, *player* is a concrete and *shame* an abstract noun, and, as was noted in the literature review above (see e.g., Amuzie & Spinner, 2013), the degree of success in article use depends on this distinction (with abstract nouns being more difficult to use correctly with articles) but also on other more nuanced distinctions which result from the degree of boundedness of a given noun. It is, therefore, possible that the abstract and less countable character of *shame* reduced the phraseological advantage which was expected on the basis of the phrase *what a shame* being frequent. For the pair *get a job* and *live in a luxury apartment*, one plausible explanation is that the combination *live in a luxury apartment* was generally very easy for the test takers, with *apartment* being a clearly countable, concrete noun. The mean scores for both items were very high (0.89 and 0.90, respectively), which means that the effect of formulaicity, if any, may not have registered because of a kind of ceiling effect for the rare combination. For the last pair which did not show a difference, *the centre of attention* and *the ecology of waterways*, it is difficult to provide a plausible explanation for this fact.

Table 3 Test scores compared for item pairs (N = 90)

Version	Article	Target item	M	SD	t	p(t)
A		<i>a friend of mine</i>	0.80	0.40		
B	<i>a</i>	<i>an acquaintance of mine</i>	0.48	0.50	4.75	<.0001
A		<i>what a shame</i>	0.79	0.41		
B	<i>a</i>	<i>what a remarkable player</i>	0.81	0.39	-0.37	.71
A		<i>twice a day</i>	0.89	0.31		
B	<i>a</i>	<i>five times a semester</i>	0.79	0.41	1.83	.04
A		<i>the sooner the better</i>	0.84	0.36		
B	<i>the</i>	<i>the smaller the pot, the more critical the problem</i>	0.23	0.43	10.35	<.0001
A		<i>a cup of tea</i>	0.99	0.11		
B	<i>a</i>	<i>a spoonful of syrup</i>	0.90	0.25	1.94	.02
A		<i>the day I die</i>	0.94	0.23		
B	<i>the</i>	<i>the food I brought</i>	0.68	0.47	4.83	<.0001
A		<i>help the poor</i>	0.73	0.45		
B	<i>the</i>	<i>open to the insured</i>	0.52	0.50	2.99	<.01
A		<i>hit (someone) in the face</i>	0.69	0.47		
B	<i>the</i>	<i>cut in the hand</i>	0.40	0.49	4.04	<.0001
A		<i>speak English</i>	1.00	0.00		
B	<i>zero</i>	<i>learn Kurdish</i>	0.93	0.25	2.52	.01
A		<i>get a job</i>	0.89	0.32		
B	<i>a</i>	<i>live in a luxury apartment</i>	0.90	0.30	-0.24	.81
A		<i>have kids</i>	1.00	0.00		
B	<i>zero</i>	<i>eat carbohydrates</i>	0.89	0.32	3.34	<.01
A		<i>the centre of attention</i>	0.61	0.49		
B	<i>the</i>	<i>the ecology of waterways</i>	0.57	0.50	0.6	.55

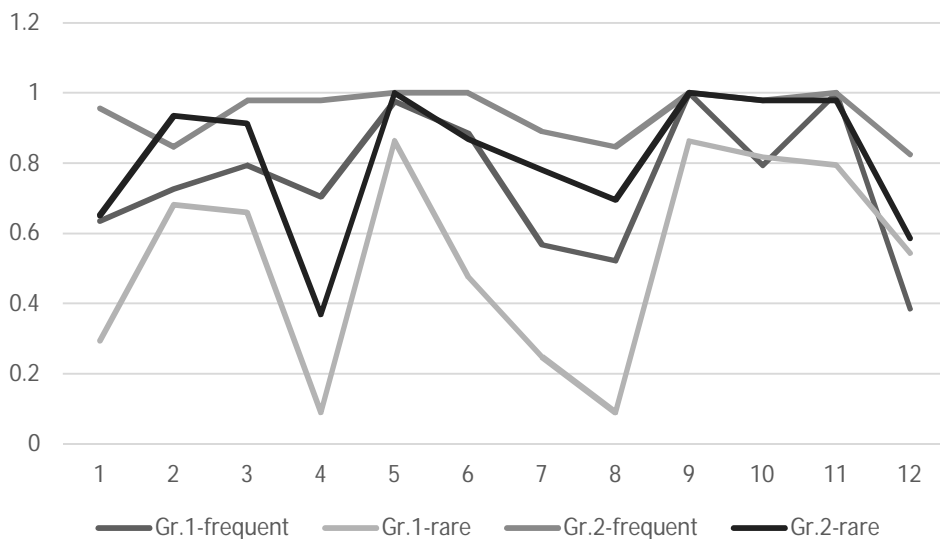


Figure 1 Means for the frequent and the rare combinations for both groups

Table 3 shows mean scores for rare and frequent items for all the participants. The means were also calculated for the two groups separately, and the results are presented in Figure 1. As can be seen in this figure, response patterns were very similar for both groups.

Two item-related issues need to be addressed. One is the possible effect of adjectival premodification on the use of articles with nouns. In two of the test items the noun happened to be premodified by an adjective (*a remarkable player, a luxury apartment*), which introduced the possibility of another variable confounding the results as there are reasons to believe that adjectival premodification may somehow interplay with article use by L2 English learners. Trenkic (2008) found that learners from article-less language backgrounds tend to omit articles more in adjectivally premodified (Art+Adj+N) than in nonmodified contexts (Art+N). She also offered a “syntactic misanalysis account” (Trenkic, 2007), which links the failure to use articles to the fact that articles are treated as adjectives.

In this study, the two items which included premodified nouns in the target items belonged to the “rare” category, thus potentially contributing to the expected lower scores for those items because of a variable that was not taken into consideration. However, for the two pairs in which the two items occur, *what a shame* and *what a remarkable player*, and *get a job* and *live in a luxury apartment*, the expected effect was *not* observed. In other words, the learners were similarly successful in providing an article in both the rare and the frequent item, *despite* the fact that the rare item was additionally more likely to be more difficult due to the use of an adjective. Thus, in this study, the issue of adjectival premodification did not appear to play a role in article use, at least as far as one can tell on the basis of the two target items which featured adjectival premodification.

Another issue which needs to be addressed is the level of difficulty of some of the words. It is true that some of the “rare” items feature words of somewhat lower frequency than the “frequent” combinations. However, all the lexical items in both types of expressions were expected to be familiar to the learners, as explained in the instrument and procedure section.

An interesting finding concerns the types of wrong test answers provided by the participants. As stated above, in the process of compiling dichotomous scores, 1 point was awarded for a correctly supplied article, and 0 points were given for failing to insert an article or for inserting an incorrect one. Out of the 512 answers for which the score was zero, an overwhelming majority—475 answers, almost 93%—were answers which were wrong because no article was provided. Only 37 answers were cases in which a wrong article was supplied. This indicates that, regarding article use by learners from articleless L1 backgrounds, failing to provide an article is much more common than providing an incorrect one. Of course, failing to use an article can also be seen as a case of

wrong article choice, namely, the choice of the zero article. However, it is impossible to distinguish between the use of the zero article and failing to use any article (cf. Leńko-Szymańska, 2012), nor is it certain that making a distinction of this kind is feasible. The concept of the “zero” article is in itself problematic and not universally recognized by linguists (Berezowski, 2009). It should be noted that the format of the test used in this study, which did not provide a prompt to use an article in specific places in the text (as, for instance, a gapped article test would do) may have contributed to the notable underuse of articles. As far as the present analysis goes, whatever the reason for failing to use an overt article, it remains an interesting finding in its own right that the participants were much more inclined not to use an article than mistaken as to which of the overt articles (*a(n)* or *the*) should be provided.

5. Conclusion

This article argues that the perception of article use (outside of idiomatic uses) as being purely rule-governed may be incomplete and should be broadened to include what is here called the phraseological perspective. The study presented here provides some initial evidence in support of this claim.

The results presented above do offer support for the view that frequency-driven conventionality in language plays a role in the use of articles in L2 English. The overall results show that the Polish learners’ use of articles is consistently more successful in the case of those word combinations that are frequent. The difference in mean scores between rare items and their frequent counterparts was very clear and significant. There is, therefore, some learner sensitivity to the frequency of linguistic forms in the input, which is here interpreted to be a sign of the open-choice principle at work. The exact nature of the psycholinguistic reality behind this phenomenon is beyond the scope of the present discussion. Here, it can only be said that there is *some* formulaicity-related mechanism at work which affects the use of articles by L2 learners. This is an important point because accounts of article use in L2 do not generally take that mechanism into consideration and treat the use of articles as purely grammatical processes.

In this study, this phraseological effect appears to be more visible in less advanced learners of English, which was to be expected as with rising competence the learners’ performance with respect to both categories of article use should be gradually improving. It is worth noting that, even though the gap between the scores for rare and frequent combinations is smaller in the case of Group 2 than Group 1, it is nevertheless statistically significant, which provides further evidence for the fact (mentioned in the literature review) that articles remain an area of difficulty even at advanced levels of English proficiency.

This study is not without shortcomings, such as the fact that the pairing of *what a shame* with *what a remarkable player* inadvertently introduced an additional difference, that of the degree of concreteness/boundedness of the noun. Further testing will require adjustment in this regard and should also offer better control of the syntactic context in which the test items appear. The topic addressed by this study, however, is an intriguing one and definitely deserves further inquiry, possibly with more rigorously designed or fine-tuned tests. The present study relied on researcher intuition in designing the research instrument, for lack of any other viable method of constructing the tool needed to investigate the issue. A test constructed without reliance on researcher intuition would be superior to the one used in this study. Also, it is possible that other measures of formulaicity could be used, for example, the mutual information (MI) score of the words in a string, which has been found to be more closely/strongly related to the processing speed of native speakers than the raw frequency of the string as a whole (Ellis, Simpson-Vlach, & Maynard, 2008). Since the present study showed higher success for frequent combinations over rare ones by Polish ESL learners, a similar phenomenon would likely be observed in the case of speakers of other article-less languages. Research with such populations is thus warranted. It would also be interesting to see if phraseology-related effects obtain in the case of learners from other L1 backgrounds.

References

- Aitchison, J. (1987). *Words in the mind: An introduction to the mental lexicon*. Oxford: Blackwell.
- Amuzie, G. L., & Spinner, P. (2013). Korean EFL learners' indefinite article use with four types of abstract nouns. *Applied Linguistics*, 34, 415-434.
- Andersen, R. W. (1984). The one to one principle of interlanguage construction. *Language Learning*, 34, 77-95.
- Berezowki, L. (2009). *The myth of the zero article*. London: Continuum.
- Bitchener, J., & Knoch, U. (2010). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, 19(4), 207-217.
- Butler, Y. 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, 451-480.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55-81.
- Chaudron, C., & Parker, K. (1990). Discourse markedness and structural markedness: The acquisition of English noun phrases. *Studies in Second Language Acquisition*, 12, 43-64.
- Council of Europe (2011). *Common European framework of reference for languages: Learning, teaching, assessment*. Brussels: Council of Europe.
- Davies, M. (2008-) *The corpus of contemporary American English: 450 million words, 1990-present*. Retrieved from <http://corpus.byu.edu/coca/>
- Díez-Bedmar, M. B., & Papp, S. (2008). The use of the English article system by Chinese and Spanish learners. In G. Gilquin, S. Papp, & M. B. Díez-Bedmar (Eds.), *Linking up contrastive and learner corpus research* (pp. 147-175). Amsterdam: Rodopi.
- Downing, A. & Locke, P. (1995). *A university course in English grammar*. New York: Prentice Hall.
- Ekiert, M. (2007). The acquisition of grammatical marking of indefiniteness with the article a in L2 English. *Columbia University Working Papers in TESOL and Applied Linguistics*, 7, 1-43.
- Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33-68). Cambridge: Cambridge University Press.
- Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal Teddy bear. *Annual Review of Applied Linguistics*, 32, 17-44.
- Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42, 375-396.

- Faerch, C. (1986). Rules of thumb and other teacher formulated rules in the foreign language classroom. In G. Kasper (Ed.), *Learning, teaching and communication in the foreign language classroom* (pp. 65-79). Aarhus: Aarhus University Press.
- Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5, 236-243.
- Hawkins, R., Al-Eid, S., Almahboob, I., Athanasopoulos, P., Chaengchenkit, R., Hu, J., . . . & Velasco-Zarate, K. (2006). Accounting for English article interpretation by L2 speakers. *EUROSLA Yearbook*, 6, 7-25.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.
- Holmes, V. L. (1988). Doubt and certainty in ESL textbooks. *Applied Linguistics*, 9, 21-44.
- Holmes, V. L., & Moulton, M. R. (1993). Idioms in context: The real McKoy. *TESL Reporter*, 26(2), 41-45.
- Hsu, J.-Y. T. (2008). Role of the multiword lexical units in current EFL ESL textbooks. *US-China Foreign Language*, 6(7), 27-39.
- Hua, D., & Lee, T. H.-T. (2005). Chinese ESL learners' understanding of the English count-mass distinction. In L. Dekydtspotter, R. A. Sprouse, & A. Liljestrang (Eds.), *Proceedings of the 7th Generative Approaches to Second Language Acquisition Conference* (pp. 139-149). Somerville, MA: Cascadilla Proceedings Project.
- Huebner, T. (1983). *A longitudinal analysis of the acquisition of English*. Ann Arbor, MI: Karoma.
- Ionin, T., Zabizarreta, M. L., & Maldonado, S. B. (2008). Sources of linguistic knowledge in the second language acquisition of English articles. *Lingua*, 118, 554-576.
- Jarvis, S. (2002). Topic continuity in L2 English article use. *Studies in Second Language Acquisition*, 24(3), 387-418.
- Leńko-Szymańska, A. (2012). The role of conventionalized language in the acquisition and use of articles by Polish EFL learners. In Y. Tono, Y. Kawaguchi, & M. Minegishi (Eds.), *Developmental and crosslinguistic perspectives in learner corpus research* (pp. 83- 104). Amsterdam: John Benjamins.
- Master, P. (1997). The English article system: Acquisition, function, and pedagogy. *System*, 25, 215-232.
- Master, P. (2002). Information structure and English article pedagogy. *System*, 30, 331-348.
- Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics*, 32, 111-129.
- Nation, I. S. P. (2001.) *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

- Ogawa, M. (2008). The acquisition of English articles by advanced EFL Japanese learners: Analysis based on noun types. *Journal of Language and Culture*, 3, 133-151.
- Orlando, M. E. (2009). *The frequency and collocation of modal verbs in English as a second language textbooks as compared to standard English corpora* (Unpublished master's thesis). Université du Québec à Montréal, Canada.
- Parrish, B. (1987). A new look at methodologies in the study of article acquisition for learners of ESL. *Language Learning*, 37, 361-383.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191-226). London: Longman.
- Quirk, R., & Greenbaum, S. (1973). *A concise grammar of contemporary English*. New York: Harcourt Brace Jovanovich.
- Shintani, N., & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, 22(3), 286-306.
- Sinclair, J. M. (1987). Collocation: A progress report. In R. Steele & T. Threadgold (Eds.), *Language topics: Essays in honour of Michael Halliday* (pp. 319-332). Amsterdam: John Benjamins.
- Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sivanova-Chanturia, A., & Martinez, R. (2015). The Idiom Principle revisited. *Applied Linguistics*, 36, 549-569. doi: 10.1093/applin/amt054
- Snape, N. (2008). Resetting the nominal mapping parameter in L2 English: Definite article use and the count-mass distinction. *Bilingualism: Language and Cognition*, 11(1), 63-79.
- The British National Corpus. (2007). Version 3 (BNC XML Edition). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Retrieved from <http://www.natcorp.ox.ac.uk/>
- Thomas, M. (1989). The acquisition of English articles by first- and second-language learners. *Applied Psycholinguistics*, 10, 335-355.
- Tremblay, A., & Baayen, H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151-173). London: Continuum.
- Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61, 569-613.
- Trenkic, D. (2007). Variability in second language article production: Beyond the representational deficit vs. processing constraints debate. *Second Language Research*, 23(3), 289-327.

- Trenkic, D. (2008). The representation of English articles in second language grammars: Determiners or adjectives? *Bilingualism: Language and Cognition*, 11(1), 1-18.
- Weinert, R. (2010). Formulaicity and usage-based language: Linguistic, psycholinguistic and acquisitional manifestations. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 1-22). London: Continuum.
- White, B. (2009). Accounting for L2-English learners' article choices. *MSU Working Papers in SLS*, 1, 14-24.
- Wray, A. (2002a). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Wray, A. (2002b). Formulaic language in computer-supported communication: Theory meets reality. *Language Awareness*, 11, 114-131.
- Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*, 32, 231-254.
- Young, R. (1996). Form-function relations in articles in English interlanguage. In R. Bayley & D. R. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 135-175). Amsterdam: John Benjamins.
- Zdorenko, T., & Paradis, J. (2008). The acquisition of articles in child second language English: Fluctuation, transfer or both? *Second Language Research*, 24(2), 227-250.

APPENDIX A

The test used in the study

The text below does not have articles. Write the articles in the correct places, as in the example.

^{the}
✓

Example: He is ^{the} most wonderful person I've ever met.

1. Motorised boats harm ecology of waterways, unless their use is kept at low level.
 2. Glucose, or blood sugar, is produced in our bodies when we eat carbohydrates.
 3. We meet regularly, five times semester, at departmental meeting.
 4. Time matters. Please try to send it in as soon as possible - sooner better.
 5. I want to choose foreign language that few people want to study. Maybe I'll learn Kurdish.
 6. Plants in pots and containers require more water than you actually might think, smaller pot more critical problem. By midsummer, herbs and vegetables in containers may need water twice day.
 7. You should give him spoonful of this syrup every three hours.
 8. I'll remember you until day I die.
 9. I see that you haven't eaten any of food I brought you two days ago. Can I make you cup of tea?
 10. Old leftist political parties are re-emerging to demand that government again expand its role in economy to help poor, even at price of discouraging foreign investors.
 11. I was lucky ball didn't hit me in face.
 12. New version of insurance policy makes number of alternatives open to insured.
 13. Do you speak English?
 14. I was recently asked about my hopes for future by friend of mine. What I know is that I'd like to have kids. And I'd like to live in luxury apartment one day.
 15. Immediately after graduation I need to get job. It doesn't necessarily have to be in my field, and I'm prepared to move anywhere where I can find work. Acquaintance of mine was recently offered position in Berlin and he moved there without moment's hesitation.
 16. What remarkable player he is. His performance today really impressed me. What shame he didn't get picked for team.
 17. Every member of Royal Family enjoys star status; they are used to being centre of attention and there is strong unstated rivalry between them.
 18. He was cut in hand in same fight, according to testimony.
-

APPENDIX B

Test items and their frequency in the BNC and COCA

Pair	Version A/B	Article	Target phrase	Frequency				
				Label	BNC (100,000,000 words)		COCA (450,000,000 words)	
					Raw	Per million	Raw	Per million
1	Version A:	a	<i>a friend of mine</i>	High	230	2.30	1,327	2.95
	Version B:		<i>an acquaintance of mine</i>	Low	1	0.01	33	0.07
2	Version A:	a	<i>what a shame</i>	High	120	1.20	173	0.38
	Version B:		<i>what a remarkable player</i>	Low	0	0.00	0	0.00
3	Version A:	a	<i>twice a day</i>	High	142	1.42	754	1.68
	Version B:		<i>five times a semester</i>	Low	0	0.00	0	0.00
4	Version A:	the	<i>the sooner the better*</i>	High	28	0.28	135	0.30
	Version B:		<i>the smaller the pot, the more critical the problem</i>	Low	1	0.01	0	0.00
5	Version A:	a	<i>a cup of tea</i>	High	619	6.19	876	1.95
	Version B:		<i>a spoonful of syrup</i>	Low	0	0.00	1	0.00
6	Version A:	the	<i>the day I die</i>	High	11	0.11	81	0.18
	Version B:		<i>the food I brought</i>	Low	0	0.00	1	0.00
7	Version A:	the	<i>help the poor</i>	High	21	0.21	241	0.54
	Version B:		<i>open to the insured</i>	Low	2	0.02	0	0.00
8	Version A:	the	<i>hit (someone) in the face</i>	High	26	0.26	115	0.26
	Version B:		<i>cut in the hand</i>	Low	0	0.00	2	0.00
9	Version A:	zero	<i>speak English</i>	High	174	1.74	1,328	2.95
	Version B:		<i>learn Kurdish</i>	Low	0	0.00	0	0.00
10	Version A:	a	<i>get a job</i>	High	299	2.99	1,749	3.89
	Version B:		<i>live in a luxury apartment</i>	Low	0	0.00	0	0.00
11	Version A:	zero	<i>have kids</i>	High	42	0.42	1,158	2.57
	Version B:		<i>eat carbohydrates</i>	Low	0	0.00	11	0.02
12	Version A:	the	<i>the centre of attention**</i>	High	85	0.85	392	0.87
	Version B:		<i>the ecology of waterways</i>	Low	1	0.01	0	0.00

Notes. * The frequency count includes both punctuation versions: *the sooner the better* and *the sooner, the better*; ** the frequency count includes both spellings: *center* and *centre*.