Typological Analysis of Articles in World Languages

Sunyoung Park

Sejong University, Korea

Abstract

A number of second and third language acquisition studies have been conducted on the acquisition of articles in order to examine the learnabilities of the feature from various L1 backgrounds, such as Chinese, Korean, Japanese, Spanish, Russian, Turkish among many others. Numerous studies have revealed ongoing difficulties of acquiring (in)definite articles by L2 and L3 language learners even at the most advanced stages of acquisition. Considering the issue of learnabilities of articles, it would be interesting and meaningful to investigate how (in)definiteness is realized in natural languages and figure out whether using articles to denote (in)definiteness is a

Sunyoung Park
Visiting Professor, Department of English Language and Literature, Sejong University, Korea
Email: sunpark@sejong.ac.kr

predominant phenomenon or not. Therefore, the current paper examined the typologlogical database—World Atlas of Linguistic Structures and presented descriptions of (in)definitenesses in natural languages. It was shown that more than half of existing natural languages do not have (in)definite articles and it is realized in other linguistic means. The current study also reviews article uses in artificial languages, such as Esperanto and Unish, concluding that in order for the language learners to be able to acquire a linguistic feature most efficiently, it should be as typologically neutral as possible.

1. Introduction

A number of studies have been conducted in recent decades to investigate the acquisition and learnability of (in)definite articles (Ionin et al. 2004, Trenkic 2008, Ko et al. 2010, Snape 2013, Park 2014, Tuniyan & Slabakova 2017). According to previous studies, articles are tremendously difficult to acquire for second and third language learners, especially when their L1 does not have article or article-like systems. Lardiere (2009) proposed the feature assembly hypothesis, which argues that in order to acquire successful acquisition, language learners should be able to reassemble features of target language which already exist in the learners' L1. In this view, learners should assemble features into new formal configurations rather than selecting features from the universal feature inventory. Therefore, convergence to L2 grammar is highly dependent on whether L1 features have the same morphological properties in the L2 grammar or not. Indeed, the role of L1 transfer has been much discussed to account for the difficulties of an acquiring article system.

The most actively conducted experimental studies include

acquisition of 'English articles' by learners from an articleless background. In addition, bidirectional studies between languages with articles, but with different usages, have been carried out. While a great number of language acquisition studies have been conducted in terms of acquiring (in)definite articles, typological distributions of descriptions on (in)definiteness have been rarely discussed so far. Therefore, the current study aims to provide descriptions of (in)definiteness and its worldwide distribution. Then it further pursues providing implications to the development of artificial languages by suggesting typologically neutral determiner forms.

The organization of the current paper is as follows. Section 2 provides descriptions of expressions of (in)definitenesses in natural languages. Section 3 presents articles in the artificial languages of Esperanto and Unish. Section 4 presents the implications and conclusions of the study.

2. (In)Definite Descriptions in Natural Languages

2.1. Definite Article Uses

The current section explores uses and non-uses of definite articles in natural languages. The data used in the current paper is employed from Dryer (2013) provided in the WALS[1]. Dryer (2013) made five categories for the analysis, and they include i) definite word distinct

---

[1] The following abbreviations are used in this paper: ACC (accusative), BEN (benefactive case), DEC (declarative), DEF (definiteness), ERG (ergative case), FSG (feminine singular), IND (indefiniteness), NP (noun phrase), PERF (perfect), PL (plural), REAL (realis mood), TOP (topic), WALS (World Atlas of Linguistic Structures), 1SG (first person singular), 3SG (third person singular), 3PL (third person plural).

from demonstrative, ii) demonstrative word used as a marker of definiteness, iii) definite affix on noun, iv) no definite article but indefinite article, and v) neither definite nor indefinite article. Table 1 shows properties of definite articles with number of languages in natural languages (Dryer 2013).

Table 1. Number of Languages Regarding Definite Articles

	Category	Number of Languages
1	Definite word distinct from demonstrative	216
2	Demonstrative word used as definite marker	69
3	Definite affix on noun	92
4	No definite article but indefinite article	45
5	Neither definite nor indefinite article	198
	Total	620

The first category, definite word distinct from demonstratives, involves languages with definite articles that are distinct from demonstrative markersn and 216 languages have this feature. Such languages include English in which definite article '*the*' is used to denote definiteness of the nouns and it is distinct from the demonstratives *this* and *that*. The Lakhota language is an example of such languages.

(1) Lakhota (Ingham 2001: 16)
wic'aṡa   ki   he
man   the   that
'that man'

In the Lakhota language, the definite article 'ki' and demonstrative 'he' are distinct from each other, but they can occur together.

There are also languages in which demonstrative makers are used to express definitenesses, as described in the second category. In many languages, demonstrative words are used anaphorically to refer back to the fore mentioned referents. The position of the demonstrative marker and its frequency of occurrence greatly vary by language. Consider the example of Eastern Ojibwa language in (2).

(2) Eastern Ojibwa (Nichols 1988: 46)

"mii	maanpii	wii-bkeyaanh"	kido	giiwenh
but	here	intend-turn.off.1SG	say.3SG	it.is.said
wa	mko.			
that	bear			

'"Well, this is where I turn off," the bear said.'

As shown in (2), to anaphorically refer back to the noun 'mko (bear)' that was previously mentioned in the discourse, the demonstrative marker 'wa', glossed as 'that', is employed in Eastern Ojibwa language. In languages like Swahili, the demonstrative maker occurs in different positions depending on its usages. In other words, when the demonstrative maker is demonstratively used, it follows nouns, whereas when the demonstrative marker is used to denote definiteness, it precedes nouns. On the other hand, the cases are vice versa in some languages, including Ute (Givón 1980), Shambala (Besha 1993), and Pa'a (Skinner 1979).

The third category features languages where the definite marker is attached pronominally, and it was revealed that 92 languages fall into this category. An example of such languages can be illustrated as shown in Egyptian Arabic.

(3) Egyptian Arabic (Gary & Gamal-Eldin 1982: 59)
     ʔit-tajjaar-a   gaaja.
     <u>the-plane</u>-FSG   come
     'The plane is coming.'

As shown in (3) in Egyptian Arabic, a definite marker '**ʔiṭ**' is attached before noun as an affix to express the definiteness of the following noun 'plane'. What is more, 'clitics' are used to express definiteness in some languages. Definite clitics can appear as an affix form before nouns or as postnominal modifiers. Consider an example of Angami language in (4).

(4) Angami (Giridhar 1980: 93)
     lêsə̂dá   kêvī<u>=ù</u>
     book   good<u>=the</u>
     'the good book'

In Angami, the clitic '*ù*' *is* attached after the adjective '**kêvī**' and modifies preceding nouns to denote definite meaning. The fourth category features languages with no definite article but with an indefinite article. For instance, the Tauya language, spoken in Madang Province in Papua New Guinea, has no definite article, but does have an indefinite article, '*ʔafa*'.

(5) Tauya (MacDonald 1990: 108)
    a.   fanu   <u>ʔafa</u>
        man   <u>IND</u>
        'a man'

b. *nen-ni*   *wate*   *amoʔo=pe*   *ese-i-ʔa.*
    3PL-ERG   house   new=BEN   want-3PL-<u>IND</u>
    'They want a new house.'

As in (5a), Tauyan has an indefinite article, but the use of it is not obligatory as shown in (5b).

The last category involves languages with no (in)definite articles at all, and its number reaches to 198. In these languages, definite distinction is not clear, as exemplified in (6).

(6) Cherokee (Scancarelli 1987: 190)
    <u>*ki:hli*</u>   *u:-skala*   <u>*achu:ca.*</u>
    <u>dog</u>   3SG-bite.PUNCT   <u>boy</u>
    'The/a dog bit the/a boy.'

As shown in (6), 'dog' and 'boy' are used with no (in)definite articles, and their meaning is vague in terms of (in)definiteness. To sum up, based on the data provided by Dryer (2013), the number of languages expressing definiteness either with a definite article or an affix can be counted at 308. Meanwhile, other languages without distinct definite marker add up to 312 languages, and these categories include languages with i) demonstrative word used as a definite marker (69), ii) no definite article, but indefinite article (45), iii) neither definite nor indefinite article (198). Considering the number of languages, it appears that the number of languages with a definite article and affix is 49.67%, and languages without a definite article account for 50.33%. Before discussing further, an examination of the uses of indefinite articles in natural languages is needed.

2.2. Indefinite Article Uses

The current section explores distribution of indefinite articles. A morpheme that denotes pragmatically indefinite NP is considered to be an indefinite article. Dryer (2013) made five categories for the analysis and they include: i) indefinite word distinct from numeral 'one', ii) numeral 'one' used as indefinite article, iii) indefinite affix on noun, iv) no indefinite article, but definite article, and v) neither indefinite nor definite article. Table 2 shows the number of languages in relation to the properties of indefinite articles (Dryer 2013).

Table 2. Number of Languages Regarding Indefinite Articles

	Category	Number of Languages
1	Indefinite word distinct from numeral 'one'	102
2	Numeral 'one' as indefinite article	112
3	Indefinite affix on noun	24
4	No indefinite article but definite article	98
5	Neither indefinite nor definite article	198
	Total	534

What follows is a review of some languages in each category, investigating how indefiniteness marking is realized in natural languages in terms of indefiniteness. The first category is the most frequently researched area, and it includes English language the most representative example. The number of languages in the first category is reported to be 102.

(7) English

    I saw   <u>a dog</u>   on the street.

    I see-past   <u>IND dog</u>   on DEF street

    'I saw a dog on the street.'

(8) Kobon (Davies 1981: 61)

    nibi   <u>ap</u>

    woman   <u>INDEF</u>

    'a woman'

In example (7), 'a' is used to refer to an entity that has not been introduced to the both hearer and listener. Likewise, in example (8), 'ap' is used to denote the indefiniteness of the preceeding noun in Kobon language.

The second category involves languages in which employ numeral 'one' instead of an indefinite article, accounting for 112 languages. In languages like German, the numeral marker 'one' is obligatory to mark indefiniteness of the noun phrases. Consider example (9):

(9) German

    Ich   habe   <u>ein</u>   Buch.

    I   have   <u>one</u>   book

    'I have a/one book.'

The interpretation of 'ein' can be vague as to whether it is used as indefinite article or as a numeral 'one' in written German. In the mean time, in spoken German, these two interpretations can be distinguished by phonological stress. When 'ein' is used as numeral 'one' it is stressed. On the other hand, in language like Dutch, the pronunciation of numeral 'een (one)' varies according to its uses. In

other words, when the speakers intend to express the meaning of the numeral one, they will pronounce 'een' with a full vowel [en], whereas when the speakers intend to use 'een', it is pronounced as reduced vowel [ən]. Furthermore, in some languages like Turkish, the numeral 'bir (one)' is used in a different position when it functions as an indefinite article.

(10) Turkish

   a. <u>bir</u>   beyaz   köpek
       <u>one</u>   white   dog
       'one white dog'

   b. beyaz   <u>bir</u>   köpek
       white   <u>a</u>   dog
       'a white dog'

As in (10a), the numeral 'bir' functions as a numeral marker, it locates before prenominal adjective position, whereas when 'bir' is used as an indefinite article it appears after prenominal adjectives as in (10b).

The third category includes language with indefinite affixes on nouns, and 24 languages were found in this category. Consider an example sentence in (11).

(11) Korowai (van Enk & de Vries 1997: 75)
     uma-té-do   <u>abül-fekha</u>   khomilo-bo.
     tell-3PL.REAL   <u>man-IND</u>   die.3SG.REAL-PERF
     'They told that a certain man had died.'

In (11), the affix 'fekha' is employed to express indefiniteness while

'certain' is used to express that the referent was known to the speaker only, and is thus specific.

The fourth type of is languages with definite article but without indefinite article. Ninety-eight languages fall in this category, and many Arabic and Icelandic languages are with this character. Example (12) illustrates an example in Arabic.

(12) Standard Arabic

ištret	<u>kita:b</u>	alsbuʕ al-amadi.	<u>al-kitab</u>	hadjija
bought-I	<u>book</u>	week the-last.	<u>the-book</u>	present
la-<u>rfiq</u>	ʔaziz.			
to-<u>friend</u>	dear			

'I bought a book last week. The book is a present to a dear friend.'

(Mayo & Hawkins 2009)

In (12), definite article 'al-' is used to express definiteness of the referent, whereas indefiniteness is realized without an indefinite article by using bare NPs.

The last category is languages that have no articles at all, nodefinite or indefinite articles, and the number of languages in this category reaches to 198. Many Asian languages, including Korean and Japanese, are examples of such languages.

(13) Korean

Ana-nun	<u>chek</u>-ul	sat-da.
Anna-TOP	<u>book-</u>ACC	buy-PAST-DEC.

'Anna bought a book.'

As shown in (13), Korean language do not have article to refer to the indefinite referent, and bare nouns can be used to refer to the

indefinite NP. Some example languages in each category have been briefly reviewed showing evidence that definiteness and indefinites can be expressed in various means.

To sum up, the first three categories can be regarded as languages with indefinite words, and it represents 238 languages: i) indefinite word distinct from numeral 'one' (102), ii) numeral 'one' as indefinite article (112), and iii) indefinite affixed to a noun (24). The rest of the categories can be counted as languages without any distinct indefinite word and the total number of such languages can be added up as 296 languages: no indefinite article but definite article (98) and neither indefinite nor definite article (198). In other words, the number of languages with indefinite article takes up 44.5%, and those without an indefinite article accounts for 55.5%. Recall the percentages of definite articles, and it was shown that the number of languages without definite article was slightly higher than with definite article, showing 50.55% and 49.67%, respectively. More than half of the existing natural languages do not have (in)definite words, implying that articles or (in)definite words are not necessarily required to express (in)definiteness in utterances. Therefore, one can argue that the typologically neutral NP form of denoting (in)definiteness can be the form of bare nouns without any (in)definite words.

3. Typological Implications to Artificial Languages

Artificial languages, or so-called constructed languages, have been created with the intention of providing a language for interlingual communication. For example, Esperanto was designed with the purpose of becoming a universal *lingua franca*, and it was supposed to be an easy-to-learn language for people from various ranges of

linguistic backgrounds. While Esperanto could not become a universal language, it can be regarded as the most successful international auxiliary language so far. According to Ethnologue, approximately 2 million people speak Esperanto, and most of them speak the language as their second or third language (Lindstedt 2006, Wandel 2015). Zamenhof, the Polish ophthalmologist, who created Esperanto claimed that the grammar of the language is highly regular and thus could be learned in one hour, and that assumably applies to the learners with European and Slavic language backgrounds.

In fact, Charters (2015) raised a question as to whether Esperanto language is really easy to learn for everyone. It is widely known that the closer a target language is to the learners' native language, the easier it is to learn. In other words, as Esperanto is highly based on European and Slavic languages, it is natural to expect that it would be easier for learners from those backgrounds to learn, whereas perhaps it can be exotic from typological perspectives. Parkvall (2010) examined Esperanto against the typological properties of natural languages catalogued in the WALS, and it was concluded that the language is 'too European' and thus less accessible to speakers of non-European backgrounds. When Esperanto was developed by Zamenhof, a typological database, such as WALS, did not exist, thus he did not have the benefit of having contemporary knowledge on linguistic typology in the creation of the language. Theoretically speaking, finding and defining a typologically neutral form from the existing typological data of natural languages can lead to easier acquisition of the language. Therefore, let us now review articles in Esperanto in terms of typological perspectives and investigate whether it is realized in the most typologically neutral means.

3.1. Articles in Esperanto

The current section reviews how (in)definiteness is realized in Esperanto. Esperanto was constructed in a highly multilingual environment, mainly based on European languages. In terms of articles, Esperanto has a definite article 'la' but no indefinite article. Consider example (14).

(14) Esperanto
    a. Mi aĉetis   <u>libron</u>   hieraŭ.
        I   buy-past <u>book</u>   yesterday.
        'I bought a book yesterday.'

    b. <u>La libro</u> estas donaco al mia amiko.
        <u>DEF book</u> be present to my friend.
        'The book is a present to my friend.'

As shown in (14a), Esperanto does not have an indefinite article as the bare NP is used for indefiniteness. On the other hand, to express definiteness of the referent, definite article 'la' can be used in Esperanto as in (14b). Considering the number of natural languages in the typological categorization, the number of languages with definite articles and without definite articles is 98 out of 534 languages in terms of expressing indefiniteness. Therefore, one can suggest that the article system in Esperanto is not necessarily typologically neutral.

3.2. Unish Articles

Unish is an international auxiliary language developed by a research team at Sejong University in Korea. Unish was created to alleviate communication problems caused by language barriers and

overcoming the weaknesses of Esperanto. It sought to distribute a fair universal language to all language users in the world by investigating properties that are most commonly found in fourteen major natural languages and one artificial language, including Chinese, Spanish, English, Hindi, Arabic, Portuguese, Russian, Japanese, German, Korean, French, Italian, Greek, Latin, and Esperanto. Unish grammar and vocabularies have been chosen based on the following principles of commonality, short-word length and simplicity (Wikipedia, https://en.wikipedia.org/wiki/Unish). Let us examine how (in)definiteness is marked in Unish in terms of article usage.

Unish does not have article systems at all. Speakers and hearers can imply the (in)definiteness of the referent through pragmatic knowledge. Consider example (15).

(15) Unish
    a. I buyed   <u>buk</u>   yedai.
        I buy-PAST   <u>book</u>   yesterday
        'I bought a book yesterday.'

    b. <u>Buk</u>   be for i's   frend.
        <u>Book</u>   be for I-POSS   friend
        'The book is for my friend.'

The conversation in Example (15) shows that Unish does not use articles to express (in)definiteness of the referent. When the indefinite book was introduced firstly into the conversation as in (15a), the bare noun form of 'buk' was used, and when knowledge of the book was shared between both speaker and the hearer as in (15b), the bare noun was also employed without a definite marker. Unish appeared to be using a typologically neutral means in terms of denoting (in)definiteness,

and this is in line with the proposed 'minimal realization principle' required in developing artificial languages (Park & Tak 2017, Park & Chin 2020).

4. Conclusion

The current paper reviewed a number of natural languages based on the typological data in the WALS. It attempted to invest the realization and description of (in)definiteness across natural languages and discover whether uses of articles are the most popular means of expressing (in)definiteness in terms of a typological approach. Following the categorization of Dryer (2013), it was found that (in)definiteness description can be expressed in a range of ways, and using articles is not the most efficient and popular way of denoting (in)definiteness. In fact, it was found that the number of languages without (in)definite articles is greater than those with (in)definite articles. Therefore, one can suggest that having no article system can be considered as typologically neutral, and this can provide implications for constructed languages. The current study reviewed article uses in Esperanto and Unish, and the latter language was proven to offer a more efficient means of expressing (in)definiteness by not having articles at all. Although this study is limited to the uses of articles, it will be meaningful to develop typologically neutral languages by exploiting the available contemporary database in future artificial language research.

References

Besha, R. 1993. *A Classified Vocabulary of the Shambala Language, with Outline Grammar.* Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Charters, D. 2015. The Teaching and Learning of Esperanto. *Interdisciplinary Description of Complex Systems* 13.2, 288–298.

Davies, J. 1981. *Kobon. Lingua Descriptive Studies* 3. Amsterdam: North-Holland.

Dryer, M. 2013. Chapter Indefinite Articles. In: M. Dryer & M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Gary, J. & S. Gamal-Eldin. 1982. Cairene Egyptian Colloquial Arabic. *Lingua Descriptive Studies* 6. London: Croom Helm.

Giridhar, P. 1980. *Angami Grammar*. Mysore: Central Institute of Indian Languages.

Givón, T. 1980. The Binding Hierarchy and the Typology of Complements. *Studies in Language* 4.3, 333–377.

Ingham, B. 2001. *English-Lakota Dictionary*. Richmond: Curzon.

Ionin, T. et al. 2004. Article Semantics in L2 Acquisition: The Role of Specificity. *Language Acquisition* 12.1, 3–69.

Ko, H. et al. 2010. The Role of Presuppositionality in the Second Language Acquisition of English Articles. *Linguistic Inquiry* 41.2, 213–254.

Lardiere, D. 2009. Some Thoughts on the Contrastive Analysis of Features in Second Language Acquisition. *Second Language Research* 25.2, 173–227.

Lindstedt, J. 2006. Native Esperanto as a Test Case for Natural Language. *SKY Journal of Linguistics* 19, 47–55.

Mayo, M. & R. Hawkins. 2009. *Second Language Acquisition of Articles: Empirical Findings and Theoretical Implications.* Amsterdam: John Benjamins.

MacDonald, L. 1990. *A Grammar of Tauya*. Berlin: Mouton de Gruyter.

Nichols, J. 1988. *An Ojibwe Text Anthology.* London: Centre for Research and Teaching of Canadian Native Languages.

Park, S. 2014. L2 Acquisition of Genericity in English Articles: The Case of Korean Adult Learners of L2 English. Ph.D. Dissertation, University of Sheffield.

Park, S. & J. Tak. 2017. Articles in Natural Languages and Artificial Languages. *Journal of Universal Language* 18.1, 105–127.

Park, S. & S. Chin. 2020. Examining the Irregularities of Articles and Introducing Minimized NP Systems in Unish. *Journal of Universal Language* 21.1, 69–88.

Parkvall, M. 2010. How European is Esperanto?: A Typological Study. *Language Problems and Language Planning* 34.1, 63–79.

Scancarelli, J. 1987. Grammatical Relations and Verb Agreement in Cherokee. Ph.D. Dissertation, University of California.

Skinner, M. 1979. Aspects of Pa'anci Grammar. Ph.D. Dissertation, University of Wisconsin.

Snape, N. 2013. Japanese and Spanish Adult Learners of English: L2 Acquisition of Generic Reference. *Studies in Language Sciences: Journal of the Japanese Society for Language Sciences* 12, 70–94.

Trenkic, D. 2008. The Representation of English Articles in Second Language Grammar: Determiners or Adjectives? *Bilingualism: Language and Cognition* 11.1, 1–18.

Tuniyan, E. & R. Slabakova. 2017. L2 Acquisition of Definiteness in English: Non-target Mapping of Anaphoricity onto 'the'. Paper

presented at The 13th International Conference of the Generative Approaches to Language Acquisition, Palma de Mallorca.

van Enk, G. & L. de Vries. 1997. *The Korowai of Irian Jaya: Their Language in its Cultural Context*. Oxford: OUP.

Wandel, A. 2015. How Many People Speak Esperanto? Or: Esperanto on the Web. *Interdisciplinary Description of Complex Systems* 13.2, 318–321.