# Examining the Irregularities of Articles and Introducing Minimized NP Systems in Unish

**Sunyoung Park & Silo Chin**
*Sejong University, Korea*

## Abstract

The current research carefully reviews article systems in existing natural and artificial languages. It also shows various denotations of article systems in a series of languages, arguing that irregularities of article use across the languages pose ongoing difficulties to language learners. The current study suggests that the Unish language is advantageous to learn in terms of the article system because it only marks specificity, which can be argued to be minimally represented

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

Silo Chin (Corresponding author)
Visiting Professor, Department of English Language and Literature, Sejong University, Korea
Email: silochin@sejong.ac.kr

by not incorporating external interfaces. The current study claims that in order for the learners to be able to acquire a language easily, representations of grammatical features should be regular and the configuration process should be minimized.

Keywords: universal language, articles, regularity, Unish, interface hypothesis, minimal representation principle

# 1. Introduction

In language acquisition studies including L1, L2, L3 as well as bilingual studies, the 'acquisition of article' has recently received a great deal of attention. One reason for this phenomenon can be because of the ongoing difficulties of article acquisitions. Some researchers attested that even the most advanced language learners cannot reach native-like proficiencies with regard to the article systems (Robertson 2000, White 2003, Ionin & Montrul 2010, among many others). Grammatically, articles fall in the functional category, and in fact, a number of languages (i.e., Korean, Chinese, the majority of Slavic and Baltic languages, and Bantu languages) do not have articles.

Dryer (1989) argues that the article system is not a common grammatical phenomenon, suggesting that one third of the world languages have articles in their languages, while only 8% of them denote definiteness (Calier & De Mulder 2010). Furthermore, even within the languages that denotes definiteness, the notion of definiteness is marked variously across the languages. Languages like English and German mark both definiteness and indefiniteness with definite and indefinite articles, respectively. Some languages mark only definiteness (Arabic, Hebrew), whereas some mark only indefiniteness (Turkish). The large variation of article representation

across languages makes configuration or remapping of article features to the target language difficult for language learners. Furthermore, as will be discussed in Section 2.7, the definiteness feature in an article system involves external interface, posing further difficulty to learners.

Likewise, articles have innate difficulties with their involvement of several sub-linguistic modules including syntax-semantics-discourse knowledge. In addition, variant forms across the languages make it more difficult to acquire. Thus, in artificial languages, which by nature will be the learners' additional languages for several purposes, the configuration process should be minimized. Therefore, the current study adopts 'Minimal Representation Principle' proposed by Park & Tak (2017) and suggests the 'no article system' in Unish, a language currently under development. The organization of the current paper is as follows. Section 2 discusses the irregular use of articles in natural languages, and Section 3 methodically presents Noun Phrase representations in the Unish language. Section 4 provides implication and conclusion.

## 2. Irregular Uses of Articles across Languages

Many natural languages have article systems to encode NPs. Some languages have definite/indefinite distinction, whereas other languages have specific/non-specific distinction.

### 2.1. Definiteness and Specificity

First of all, the definite and specific features are both discourse related, and they are related to the knowledge of speaker and/or hearer

in the discourse. Definiteness is a semantic feature that refers to the knowledge of both the hearer and the speaker regarding a unique referent. On the other hand, specificity is a semantic feature that makes reference to a speaker knowledge about a unique referent in the discourse. The informal distinction of two notions are adopted by Ionin et al. (2004) and provided below.

> (1)    Definiteness and Specificity: Informal Definitions
>         If a Determiner Phrase (DP) of the form [D NP] is…
>         a. [+definite], then the speaker and hearer presuppose the existence of a unique individual in the set denoted by the NP.
>         b. [+specific], then the speaker intends to refer to a unique individual in the set denoted by the NP and considers this individual to possess some noteworthy property.
>
> <div align="right">(Ionin et al. 2004: 5)</div>

In many natural languages, notions of (in)definiteness or (non)-specificity are marked by articles, and their representations vary. Because of the inconsistent/various use of articles, L2 and L3 learners have difficulty when they are confronted by them. Let us first examine different uses of articles in some natural languages.

## 2.2. English

Definiteness is marked by the articles 'the' and '(a)n' in English, whereas specificity is not overtly marked in English. The following examples illustrate definiteness and specificity in English discourses.

> (2)    a. [+definite, +specific] 'I would like to talk to the winner of today's race—she is my best friend!'

b. [+definite, –specific] 'I would like to talk to <u>the winner</u> of today's race—whoever that is; I'm writing a story about this race for the newspaper.' (Ionin et al. 2004: 8)

c. [definite, +specific] '<u>A man</u> just proposed to me in the orangery—though I am much too embarrassed to tell you who it was.'

d. [–definite, –specific] '<u>A man</u> is in the women's bathroom—but I haven't dared go in there to see who it is.'

(Fodor & Sag 1982: 359)

In example (2a), the definite article 'the' is used because the existence of a winner is presupposed by the both speaker and the hearer. It is also [+specific] since the speaker has concrete knowledge of the referent and intends to refer to a specific individual in his mind. The winner is [+definite] but [–specific] in example (2b) because both the speaker and the hearer presuppose the existence of a winner, but the speaker lacks specific knowledge of the referent, and the speaker does not intend to refer to a particular individual. In example (2c), a man is [–definite] because no unique referent is shared by the speaker or the hearer from the discourse knowledge or mutual world knowledge.

On the other hand, 'a man' is also [+specific] and the indefinite article 'a' is used because the speaker intends to refer to a particular individual in his mind, but the hearer does not know the man. A man in example (2d) is [–definite] because no unique referent information is shared by either the speaker or the hearer, but it is also [–specific] because the speaker does not intend to refer to a particular individual in mind. As was shown in examples (2a) through (2d), English denotes only definiteness but not specificity in their NP representation.

## 2.3. German

German uses articles on the basis of definiteness similar to English language. German overtly marks the [+definite] feature through the definite articles 'der', 'die', 'das' and the [–definite] feature via the indefinite article 'ein/-e'. However, specificity is not overtly marked in German. Consider the following examples in (3) for definite and specific illustration in German:

(3)   a. [+definite; –specific]
     Er    geht    in <u>die Kirche</u>
     He    go-3PS[1]  to the church
     'He goes to the church.'

    b. [+definite; +specific]
     Er   geht    in   <u>die  Kirche</u> (am Moritzplatz)
     He  go-3PS  to   the  church (on Moritzplatz)
     'He goes to the church (on Moritzpatz).'

    c. [–definite, +specific]
     Ich bringe    auf die Party morgen   <u>eine gut    Freundin</u>
     I   take-FUT to  the party tomorrow a    good  friend
     von mir mit,   wenn  es   dir   recht  ist
     of   mine,    if    it   you  okay  be
     'I will take a good friend of mine with me to the party
     tomorrow, if that is okay with you.'

---

[1] The following abbreviations are used in this paper: ACC (accusative), AGR (agreement), AOR (aorist), ART (article), DEF (definite article), FUT (futuer tense), NEG (negative), PP (past participle), PAST (past tense), PROG (progresive), PUL (plural), 1PS (first person singular), 3PS (third person singular).

d. [–definite, –specific]

| Klaus | hat | eine | <u>Wohnung</u> | gemietet, |
|-------|-----|------|---------|-----------|
| Klaus | have-3SG | a-ACC | flat | rent-PP, |

| aber | ich | weiß | nicht | wo |
|------|-----|------|-------|-----|
| but | I | know | not | where |

'Klaus has rented a flat, but I don't know where.'

(Bisle-Müller 1991: 39)

In example (3a), definite article 'die' is used before the Kirche (church) because the speaker and hearer share knowledge of the existence of the church in their presupposition. Speaker does not intend to refer to a specific church in his mind, thus [–specific]. On the other hand, in example (3b), because the speaker has a specific church in mind, the 'die Kirche' is [+specific]. In example (3c), even though the speaker has a particular friend in mind, the referent 'Frendin' is not familiar to the hearer. Therefore, the indefinite article 'eine' is used in (3c). In example (3d), neither the speaker or hearer have a specific flat (eine Wohnung) in mind, thus [–definite, –specific].

German is similar to English in that it denotes definiteness through definite and indefinite articles, but it should be also pointed out that German is more complex than English because German articles are case, grammatical gender, and number sensitive (see Bisle-Müller 1991 for further information).

## 2.4. Arabic

Both Standard Arabic and the vernaculars encode definiteness but not specificity. Similar to English, definiteness is marked morphologically with definite article 'al-' in Arabic (Watson 2002). In order to mark the definiteness of a noun, article 'al-' is prefixed to singular, plural and mass nouns as in (4).

(4)   a. <u>al-baytu</u>     kabeerun
         the-house   big
         'The house is big.'

      b. <u>al-baytu</u>      al-kabeer
         the-house    the-big
         'the big house'

While Arabic is similar to English and German language in that it has a definite article, it is different in that Arabic does not have an indefinite article. Indefiniteness is marked phonologically in Standard Arabic at the end of the indefinite noun. In most Arabic vernaculars, indefiniteness is unmarked at all. The following examples present how definiteness is encoded independently from specificity.

(5)   a. uriidu       an     atakallama *ma'* <u>al-*fa'izi*</u> ayyan  kan
         1st person  want  to talk        with DEF-winner    whoever
         'I want to talk to the winner whoever he is.'

      b. uriidu       an     atakallama *ma'*   <u>al-*fa'izi*</u>
         1st person  want  to talk        with   DEF-winner
         huwa          Sadiiqi
         he is         my friend
         'I want to talk to the winner; he is my friend.'
                                                    (Abudalbuh 2016: 110)

Example (5a) features [+definite, –specific] noun and definite article 'al-' is used. On the other hand, (5b) is [+definite, +specific] and definiteness is also marked with 'al-' regardless of specificity distinction.

## 2.5. Turkish

Turkish is an articleless language, and it does not have an overt article that marks definite feature. However, Turkish has an (optional) overt quantifier 'bir' which marks indefinite feature, and it can also refer to the numeral 'one' (Lyons 1999, Goad & White 2009). Example 6 presents an illustration of the indefinite article in Turkish.

(6)   Müdür    <u>bir araba</u>   iste-di [–definite; –specific]
      President  a car       ask for-PAST-3SG
      'The President asked for a car.'          (Tura 1973: 123)

As shown in example (6), the indefinite quantifier 'bir' can be used to refer to a particular entity that is unknown to the both speaker and the hearer. In regards to the representation of definite NPs, some factors such as word order, stress and case marking decides the definiteness of bare nouns in Turkish. For instance, bare nouns at the beginning of a sentence can always be read as definite, as presented in example (7).

(7)   <u>Çocuk</u>   yer-de        yat-ıyor-du [+definite]
      Child    ground-LOC    lie-PROG-PAST-3S
      'The child was lying on the ground.'      (Tura 1973: 102)

As in example (7), 'Çocuk (child)', bare NP in the subject position is interpreted as definite. In addition, object NPs with case markers can also be interpreted as definite as in (8) below (Tura 1973, Aygen-Tosun 1999).

(8)   Ben belki   kitab-ı      oku-r-um [+definite]
      I     maybe   book-ACC  read- AOR-1S AGR
      'Maybe I read the book.'          (Aygen-Tosun 1999: 1)

Unlike the previously discussed languages, Turkish denotes specificity by using overt case morphology (Tura 1973, Aygen-Tosun 1999). According to Aygen-Tosun (1999), only object NPs with the quantifier 'bir' and accusative case marker '-(y)i'[2] can be interpreted as specificity as in (9a, 9b).

(9)   a. Ben   bir kitap   oku-du-m (–specific)
         I      a book      read-PAST-1S
         'I read a book.'

      b. Ben   bir kitab-ı      oku-du-m (+specific)
         I      a book-ACC    read-PAST-1S
         'I read one of the books.'       (Aygen-Tosun 1999: 2)

'Kitap (book)' with the preceeding quantifier 'bir' and the accusative case marker '(y)i' is marked with specificity as in (9b), whereas in the absence of an accusative case marker, 'Kitap (book)' with the preceding quantifier 'bir' is interpreted as non-specific as in (9a).

In sum, Turkish denotes definiteness, and it is encoded differently in the subject and object NP positions. Subject NPs are always definite unless they are overtly marked by 'bir', the indefinite marker. Meanwhile, the definiteness of NPs in the object position should be marked with an accusative case marker. Furthermore, the presence of the over case morphology determines specificity readings. All definite NPs are specific by default, but indefinite NPs can be [+specific] or

---

[2]  'Yi' is a phonological variation of 'i'.

[–specific].

## 2.6. Esperanto

Esperanto is a well-known constructed language based mostly on European languages English and Romance languages with a few influences from other Germanic languages and Slavic languages. Even though Esperanto is an artificial language, it overtly marks definiteness and has a definite article 'la', which is equivalent to 'the' in English. However, it does not overtly mark indefiniteness.

(10)   a. <u>La     knabo</u> kuras
          ART   boy     run
          'The boy runs.'


       b. <u>Knabo</u>  kuras
          boy     run
          'A boy runs.'                    (Kellerman 2009: 24)

Example (10a) shows definiteness marking in Esperanto with definite marker 'la'. On the other hand, indefiniteness is marked by omitting definite article 'la' as in (10b).

We have seen that article systems greatly vary in accordance with the particular language. As discussed above, some languages like English and German have both definite and indefinite marker and they denote definiteness. On the other hand, languages like Arabic mark only definiteness, and languages like Turkish mark only indefiniteness. It seems that article usage is quite diverse in natural languages. Furthermore, Esperanto, an artificial language, overtly marks definiteness in a system similar to the Arabic. In fact, because of the discrepancies between language learners' mother tongues and their L2

or L3 languages, L2 or L3 learners experience difficulty when they acquire article systems in the target language. Thus, we propose that in artificial languages, the article system should be minimally represented.

## 2.7. Articles and Interface Hypothesis

'Interfaces' in L2 research is widely understood as interaction or mapping between linguistic modules or representations. The mapping process between linguistic modules always involves interfaces. For example, the syntax of a sentence should be mapped on semantics for interpretation of the sentence, and that is the syntax and semantics interface. In the same vein, sometimes the syntax of a sentence also has to map to the discourse, and that is called the syntax and discourse interface (Sorace & Serratrice 2009). Many studies have attested to difficulty with interfaces, and there has been increasing emphasis on internal and external interfaces (Paradis & Navarro 2003, Montrul 2004, Sorace & Serratrice 2009). To explain briefly, internal interfaces can be defined as the links between the language system itself (syntax-semantics, syntax-morphology, morphology-phonology, etc.), whereas external interfaces links linguistic modules with other aspects of world knowledge and cognition (syntax-semantics-discourse, syntax-semantics-pragmatics, etc.). A number of L1, L2, L3, and bilingual acquisition studies have shown that properties involving external interfaces pose much more difficulty than those involving internal interfaces (Montrul 2004, Tsimpli & Sorace 2006, Slabakova 2008, Sorace & Serratrice 2009). It was argued that "structures requiring the integration of syntactic knowledge and knowledge from other 'external' domains require more processing resources than structure requiring only syntactic knowledge" (Sorace & Serratrice 2009: 199). In fact, the notion of 'definiteness' involves discourse knowledge

(external interface), thus posing difficulty to the learners. Therefore, to minimize the configuration process of language users, sub-linguistic interfaces should be minimized, and the concept of definiteness does not necessarily need to be overtly marked.


## 3. Minimal Representation of NPs in Unish

 'Unish' is an artificial language that has been developed by Sejong University in order to remove linguistic barriers in international conversation. 'Unish' is a universal language which aims to a lingua franca in this globalized era. It has been developed based on fifteen representative languages, fourteen natural languages including English, Spanish, Portuguese, Italian, French, German, Russian, Korean, Chinese, Japanese, Arabic, Hindi, Greek and Latin, and one artificial language, Esperanto. It mainly attempts to discover 'commonness' from the aforementioned languages and adopt 'regular' and 'simple' grammatical and pronunciation rules. Kwak (2003) argues that artificial languages are more beneficial for the learners because they are relatively easy to acquire and can be neutral to different language speakers around the world. Lee (2002) previously pointed out the 'regularness' and 'easiness' of Unish in many grammatical features. Further to the previous research, we argue that one of the most distinguishable linguistic features of Unish is the simple NP system. With regards to the number markings of NPs, Unish does not compulsorily mark singularities and pluralities of NPs, but pluralities can be optionally marked with the plural marker '-s' at the end of nouns. Singular nouns can be marked with quantifier/numeral marker 'un'. Example (11) illustrates NPs in the Unish.

(11)   a. We   hav   <u>pesko</u>
            we   have   peach
            'We have a peach(es).'


       b. We   hav   only   <u>un   pesko</u>
            we   have   only   one   peach
            'We have only one peach.'


       c. We   buyed   <u>peskos</u>
            we   buy-PAST   peach-PL
            'We bought peaches.'


In example (11a), the bare noun 'pesko' is used to refer to a peach or some peachess. The numbers of nouns are not overtly marked in the NP. On the other hand, in order to express the number of peaches precisely, one can also use 'un' before the nouns as exemplified in (11b). In the same vein, to adding plural marker '-s' at the end of the nouns is also allowed to emphasize the plurality of the nouns as in (11c). With regard to the definiteness, Unish does not overtly mark definiteness with the nouns, and definiteness can be retrieved from the discourse knowledge. The following example shows denotations of (in)definiteness of NPs in Unish.


(12)   A: U   look   tired.
            'You   look   tired.'

B: I   no   sleeped     wel   yedai    nait,
   I   NEG  sleep-PAST  well  yesterday night,
   koz      dog   barked
   because  dog   bark-PAST
   'I didn't sleep well last night because <u>a dog / the dog /
   dogs</u> barked.'

In example (12), whether the dog in the example is known to the speaker and/or the hearer is unknown from the conversation. If the speaker and the hearer shared knowledge of the dog, the conversation could end there. On the other hand, if the hearer did not know about the dog and wanted to identify the dog that barked last night, the hearer could continue the conversation that asks for more information about the dog. For example, sentences in (13) can be followed after (12).

(13)  A: Which <u>dog</u>?
         'Which dog?'

      B: I's neibor raise <u>dog</u>. It be fiers.
         'My neighbor raises a dog. It is fierce.'

As shown in (13), definiteness of NPscan be easily identified by the conversation. Therefore, one can carefully suggest that 'articles', as function words, do not be necessarily need to be overtly represented in artificial languages in particular. One can rather suggest to denote specificity to refer to a specific referent by the speaker. For example, Unish uses the demonstrative 'da' to express specificity or particularity of the following referent. One can use 'da' without considering the hearer's knowledge. It can be used when the speaker wants to specify

the referent that is being referred.

(14)   a. I need <u>da fon</u>. (not the other phones)
             'I need the phone.'

        b. I need <u>(da) fib</u> dat i buyed.
             'I need the phone that I bought.'

In example (14a), regardless of the definiteness of the phone, 'da' is used because the speaker wants to refer to a specific phone in his mind. On the other hand, the use of 'da' is optional in example (14b) even though the speaker is referring to a specific object. 'Da' is optional in (14b) because the relative clause, 'dat i buyed', provides additional information about the book, and it can be evidence in identifying the referent. In fact, languages like Korean and Japanese use the demonstrative maker 'ku' and 'sono' for specificity.

Adopting the economical and simple features of some natural languages, denotation of definiteness can be reduced to that of specificity to meet the Minimal Representation Principle (Park & Tak 2017). Syntactically, definiteness involves an extra linguistic system which is discourse, but specificity can be decided within the sub-linguistic systems. Therefore, marking only specificity can be economical and less complex, thus easier to learn for artificial language learners.


# 4. Concluding Remarks


The current research has reviewed article systems in a number of existing natural and artificial languages and revealed inconsistent and

irregular uses of article systems across the languages. As article systems are presenting great difficulties to L2, L3 and bilingual learners, the current study suggests simple and regular article systems in the newly developed language, Unish. Following the Minimal Representation Principle, Unish provides the easiest and simplest representation of NPs by not denoting definiteness but specificity. While the current study has limitation in that no empirical evidence is presented with regard to the acquisition of NPs in Unish, it provides a new topic for a future research on language acquisition.

# References

Abudalbuh, M. 2016. The Acquisition of English Articles by Arabic L2-English Learners: A Semantic Approach. *Arab World English Journal* 7.2, 104-117.

Aygen-Tosun, G. 1999. Specificity and Subject-Object Positions/Scope Interactions in Turkish. *Proceedings of the 1st International Conference in Turkic Linguistics.* Manchester University. Available at <http://ww.fas.harvard.edu/~lingpub/misc/tosun/Manch%20Proceeed.pdf>.

Bisle-Müller, H. 1991. *Artikelwörter im Deutschen: Semantische und Pragmatische Aspekte Ihrer Verwendung*. Tübingen: Max Niemeyer Verlag.

Carlier, A. & W. De Mulder. 2010. The Emergence of the Definite Article: Ille in Competition with Ipse in Late Latin. In K. Davidse et al. (eds.), *Subjectification, Intersubjectification and Grammaticalization* 241-275. The Hague: De Gruyter Mouton.

Dryer, S. 1989. Article-Noun Order. *Proceedings of the 25th Annual Regional Meeting of the Chicago Linguistic Society* 83-97. Chicago, IL: Chicago Linguistic Society.

Fodor, J. & I. Sag. 1982. Referential and Quantificational Indefinites. *Linguistics and Philosophy* 5, 355-398.

Goad, H. & L. White. 2004. Ultimate Attainment of L2 Inflections: Effects of L1 Prosodic Structure. In S. Foster-Cohen et al. (eds.), *EuroSla Yearbook* 119-145. Amsterdam: John Benjamins.

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

Ionin, T. & S. Montrul. 2010. The Role of L1 Transfer in the Interpretation of Articles with Definite Plurals in L2 English. *Language Learning* 60.4, 877-925.

Kellerman, I. 2009. *A Complete Grammar of Esperanto: The International Language.* SC: BiblioLife.

Kwak, E. 2003. Comparisons between Pidgins and 'Unish'. *Journal of Universal Language* 4.1, 17-31.

Lee, D. 2002. A Comparison of Unish Grammar with Esperanto. *Journal of Universal Language* 3.2, 57-74.

Lyons, C. 1999. *Definiteness*. New York: CUP.

Montrul, S. 2004. Subject and Object Expression in Spanish Heritage Speakers: A Case of Morphsyntactic Convergence. *Bilingualism: Language and Cognition* 7.2, 125-142.

Paradis, J. & S. Navarro. 2003. Subject Realization and Crosslinguistic Interfernce in the Bilingual Acquisition of Spanish and English: What is the Role of the Input? *Journal of Child Language Acquisition* 30.2, 371-393.

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

Robertson, D. 2000. Variability in the Use of the English Article System by Chinese Learners of English. *Second Language Research* 16.2, 135-172.

Slabakova, R. 2008. *Meaning in the Second Language*. Berlin: Mouton de Grutyer.

Sorace, A. & L. Serratrice. 2009. Internal and External Interfaces in Bilingual Language Development: Beyond Structural Overlap. *International Journal of Bilingualism* 13.2, 195-210.

Tsimpli, I. & A. Sorace. 2006. Differentiating Interfaces: L2 Performance in Syntax-Semantics Syntax-Discourse Phenomena. *Proceedings of the 30th Annual Boston University Conference on Language Development* 653-664. Somerville, MA: Casadilla Press.

Tura, S. 1973. A Study on the Articles in English and Their Counterparts in Turkish. Ph.D. Dissertation, University of Michigan.

Watson, J. 2002. *The Phonology and Morphology of Arabic*. New

York: OUP.

White, L. 2003. *Second Language Acquisitino and Universal Grammar.* Cambridge: CUP.