Generic Reference in English, Arabic and Malay: A Cross Linguistic Typology and Comparison

Eidhah Abdullah AL-Malki[1], Norazman Abdul Majid[1] & Noor Abidah Mohd Omar[1]

[1] Language Academy, UTM, Skudai, Johor Bahru, Malaysia

Correspondence: Eidhah Abdullah AL-Malki, Language Academy, UTM, Skudai, Johor Bahru, 81310, Malaysia. Tel: 60-017-772-4049. E-mail: eidhah.almalki@gmail.com

Abstract

According to the Longman Grammar of Spoken and Written English 1999 by Biber et al. (p. 266) generic article uses are more than twice as common in academic English than in conversation or fiction. This is an area that English for Academic Purpose (EPA) textbooks and teachers would need to target more than general English teaching. This paper is therefore a contribution towards better understanding of what linguistic facts about generics teachers and textbooks of EAP might need to cover in order to deal with them satisfactorily, particularly for learners with Arabic or Malay as L1. This paper is also significant as it is the first to compare the expression of generic meanings by noun phrases in three typologically quite different languages: the Germanic language English, the Semitic language Arabic and the Austronesian language Malay. The contrast between the three languages is substantial in that they have different settings according to the nominal mapping parameter (NMP), which captures some widespread generalizations about the occurrence of mass and countable nouns and articles in the languages of the world. As a part of a bigger project that investigates the acquisition and interpretation of generic reference by speakers of these languages, this article is descriptive and comparative in nature. The main finding is that the rules for mapping forms to generic meanings are more complex in English than in Malay or Arabic, in that English marks the difference between NP level and S level genericity and between established and non-established categories.

Keywords: genericity, generic reference, classifier languages, nominal mapping parameter, articles, typology

1. Introduction

The expression of generic reference is a linguistic universal across all human languages. The precise means used to convey reference to generic meaning, however, is not universal and, as we shall see later, languages possess different ways to express that generic meaning. No language possesses an inflection that is specifically used for generic reference of nouns yet all languages make distinctions between generic and particular noun phrases and in all languages generic expressions depend on the synergy of the verbal and the nominal elements of that language (Dayal, 2005). This paper will introduce the linguistic background of the phenomenon and compare the linguistic elements that are involved in generic expression in the three object languages (English, Arabic and Malay).

Generic reference of NPs is more common in academic English than other varieties of English, though still only 5% of article uses in English according to Biber et al. (1999). Yet, EAP textbooks seen by the researcher do not offer adequate treatment of generics, as they do not cover generic use of articles explicitly at all. Skills for Success by Caplan et al. (2011) is an example of an EAP textbook where generics are rarely used in reading passages and not actually taught in any grammar section. This paper will be useful for EAP textbook writers and teachers to help understand what they should be covering in relation to generics especially for speakers of languages like Malay and Arabic.

2. Generic Reference: A Cross-Linguistic Overview

Generic sentences express general laws and regularities as in example 1.a not particular or episodic facts as in 1.b

(1)   a. Dogs are loyal.

      b. Dogs are going through my garbage.

Both the sentences with generic noun phrases (NPs) in items 2.a.b.c below and the sentences with habitual or characterizing predicates (but NPs that do not in themselves exhibit generic reference) in 2.d.e.f express typical characteristics of their subjects. Such sentence genericity is shown by the tense/aspect of the verb primarily (often the timeless use of present simple or *used to* in the past in English). Generic sentences, then, provide "propositions which do not express specific episodes or isolated facts, but instead report a kind of general property, that is, report a regularity which summarizes a group of particular episodes or facts" (Krifka et al., 1995:2). In more recent work, however, habituals   (as in 2.d.e.f) are by some experts not considered generic since their NPs refer to individuals rather than kinds (Declerck, 1991; Krifka et al., 1995; Behrens, 2000). In other words, for them, in sentences with generic NPs (2.a.b.c) the generalizations are made over the noun, while in habitual sentences (2.d.e.f), the generalizations are made over what the noun does habitually.

(2)   a. The potato grows underground.

      b. A potato is a brown vegetable

      c. Potatoes grow in the ground.

      d. I go to school.

      e. Your dog bites

      f. Mary smokes

In this paper, we are only looking at sentences that do contain a generic NP as our interest is in NP genericity. Episodic sentences (typically with progressive aspect or event use of simple past in English) generally do not contain a generic NP except in rare instances where a kind-selecting predicate is used (*extinct* etc. See section 5.). In this study, then, genericity, generics and generic reference will be used to refer to NP genericity, i.e. NPs in any sentences (like 1.a and 2.a.b.c) that refer to "kinds" and usually occur in characterizing or habitual sentences.

Different languages however employ different ways to express NP genericity. There is no better example to illustrate the diversity of expression of generic NP reference across languages than the one given by Behrens (2000). Example 3.c is the original French sentence translated into 8 languages:

(Behrens, 2000, p. 1)

(3).   a. A boa constrictor [IND, SG] is a very dangerous creature, and an elephant [IND, SG] is very cumbersome.

      b. GERMAN: Eine Riesenschlange [IND, SG] ist sehr gefährlich, und ein Elefant [IND, SG] braucht viel Platz.

      c. FRENCH: Un boa c'est [IND, SG] [TOPIC] trés dangereux, et un éléphant c'est [IND, SG, TOPIC] trés encombrant.

      d. HUNGARIAN: Áz óriáiskígyó [DEF, SG] nagyon veszélyes, az elefánt [DEF, SG] roppant terjedelmes.

      e. GREEK: Ο βόας [DEF, SG] είναι τρομερά επικίνδυνος κι ο ελέφαντας [DEF, SG] αρκετά ενοχλητικός.

      f. ARABIC: Al-buwwaa'u [DEF, SG] khatirun giddan, w-al-fiilu [DEF, SG] haa'ilu l-hagmi.

      g. TAGALOG: Lubhang mapanganib ang sawa [TOPIC, Ø NUM/lND], at napakalaki naman ang elepante [TOPIC, Ø NUM/IND].

      h. FlNNISH: Boat [NOM, PL] ovat hyvin vaarallisia, ja elefantti [NOM, SG] vie paljon tilaa.

      i. VIETNEMESE: Một con trăn [CLASS, NUM/IND], thật là nguy hiểm vá một con voi [CLASS, NUM/IND], thì thật là lịch kịeh rầy rà.

In the previous examples, French, English and German utilize the indefinite article with singular nouns. Hungarian, Greek and Arabic utilize the definite singular to make the reference. Tagalog, Finnish and Vietnamese are article-less languages and each language allows a different construction to make the generic reference as noted by Behrens (2000).

Despite the universality of the concept, the difficulty in working with generic reference from a cross-linguistic perspective, lies in the fact that genericity is initially a semantic phenomenon that has syntactic projections. There is dramatic variation and numerous elements that contribute to the marking of a generic reference in a given language. Behrens (2000, p. 8) observes:

"If there is any point linguists working on genericity agree upon, it is the following: genericity is a matter of interpretation which results in utterances from the interaction of a number of variable factors such as the lexical semantics of the constituting elements, pragmatic knowledge and discourse situation, grammatical marking of determination and quantification on the noun phrases, and grammatical marking of tense, aspect, and mood on the predicates, syntactic position of the noun phrases, and so on. It is probably very rarely found in the languages of the world that generic interpretation is encoded in a unique and unambiguous way by the use of exclusively generic forms."

Later in this paper, we will attempt to describe in fuller detail the relevant linguistic properties in the three object languages, English, Arabic and Malay, so as to allow us to examine the role that these properties may play in the interpretation and production of English articles in generic use by Arabic L1 and Malay L1 learners of English.

3. Typological Differences Relevant to NP Genericity between the Three Object Languages

The typological differences between the three languages which are relevant to the expression of generic meaning lie in two main areas. These are determiners, quantifiers and number markers. This section will examine those main differences between these languages in terms of:

1) Marking of in/definiteness through articles in languages that have them (determiner- vs determiner-less languages)

2) Number and quantification of nouns (number markers and classifiers)

These differences are important since articles and number are commonly used to mark genericity in addition to their basic meanings of definiteness, indefiniteness and number. Differences in what markers are available constrain what can be utilised by a language to express NP genericity.

The first typological difference between the three languages is the role determiners play in these languages. Languages either allow or require nouns to appear with an overt in/definite article or allow bare nouns to appear without an article. The first category of languages is determiner languages while the second category of languages is determiner-less languages (There will be more elaboration on the argument surrounding the existence or non-existence of a true determiner-less language later). There are three categories overall that we can define as follows:

1) Determiner languages: Require noun phrases in argument position to be preceded by a determiner.

2) Determiner-less languages: Allow all nouns to appear bare in argument position.

3) Mixed type: Allows some nouns to occupy argument position without a determiner while other nouns cannot.

Arabic fits the first category, Malay fits the second category while English fits the third since English allows singular proper nouns, plural and mass nouns in argument position with no determiner but does not allow singular nouns unless they are licensed by a determiner. Other determiner languages are Romance languages like French and Italian. Other determiner-less languages include Chinese, Japanese and Korean. It should be noted that some theorists, such as Longobardi (1994), hold that all languages have articles in deep syntax, just they do not always surface as a phonological form. This view would make all languages more similar than they appear on the surface, and indeed predict English articles to be easier to learn than they appear to be by learners with very different L1s. We, however, follow Chierchia (1998a,b) and others in regarding nouns in languages without any articles like Chinese as in fact being what they appear: article-less at all levels of grammatical analysis (see further section 4).

Another interrelating classification of languages that is useful for our purpose divides languages into classifier vs. non-classifier languages.

1) Classifier languages: Generally understood not to have overt plural morphology and therefore require a classifier to count all NPs.

2) Non-classifier languages: Do have overt plural morphology and do not require a classifier to count NPs containing countable nouns.

Malay fits the first category of languages since it has no overt plural morphology (if we exclude reduplication) and requires classifiers while English and Arabic belong to the second category since they do have overt plural morphology and do not require classifiers. In fact a common feature in the East and Southeast Asian area is the way the languages deal with number and quantification and classification of nouns. Classifier languages include Malay (Carson, 2000), Tagalog, Vietnamese, Korean, Chinese, Thai and Japanese (Chierchia, 1998).

In English and Arabic, mass nouns necessitate the use of measure phrases containing a classifier to be countable, but count nouns do not. However, in classifier languages, all nouns behave like mass nouns in requiring a measure word. In English, a non-classifier language, *people* is a countable noun that can combine with the numeral *three* in the NP *three people* with no obligatory need to insert a classifier despite the fact that there are instances of classifier use in English like *five **head** of cattle* and *ten **stem** of roses*. Only mass nouns require classifier use in English in order to be counted, like *two **cups** of rice*. In a classifier language like Chinese a noun like *people* requires a classifier to combine with the numeral *three*, therefore the equivalent of *three people* is *three **classifier** people*. Classifiers therefore render the noun countable (Allan, 1977).

From the above discussion, we have seen that typologically, English and Arabic are more similar to each other than to Malay.   Like most classifier languages, Malay does not have articles, and therefore, the prediction is that learners with this L1 background could face more obstacles in their acquisition of a language with countable nouns and which possesses an article system, such as English.

4. The Licensing Function of Articles (Nominal Mapping Parameter)

Before considering what use our three languages make of their article resources to express NP genericity, we draw attention to an influential theory-based generalisation which captures the situation with respect to articles in languages like our three. This is Chierchia's (1998a, b) Nominal Mapping Parameter (NMP), which is formulated within the framework of Chomsky's Principles and Parameters theory (Chomsky, 1993), and provides a principled general account of the three types of language that we identified above (determiner, determiner-less and mixed).

Chierchia (1998a, 1998b) proposed the idea of a Nominal Mapping Parameter which is a semantic parameter to account for variation found in the syntactic realization of NPs in argument position (i.e. positions like subject and object in the sentence). Nouns, according to this view, are either argumental ([+arg]) or predicative ([+pred]). Nouns [+arg] refer directly to kinds and can appear as arguments without the need to be 'licensed' by articles, while Nouns [+pred] refer to instances of properties which require an article or other determiner in order to be able to appear as subjects etc. This formulation is based on Partee's (1987) proposal in which the traditional distinction between referential, predicative, and quantificational NPs is formally captured in her own notation as individuals, predicates, or generalized quantifiers. Accordingly, this parameter has three semantic values or (settings) and it is claimed that all languages fall into one or other of three types:

a) Languages with exclusively argumental NPs [*+ arg, - pred*], in which bare nouns denote kinds (Carlson 1977) and can be used directly in argument position. Such languages are Malay, Chinese and Japanese where nouns can be used in argument position without determiners, in/definite articles or plural markers, and they are interpreted as definite/ indefinite, singular/ plural, according to the context (or if the speaker chooses, in case of Malay, with use of reduplication for plurality and numeral *se* and demonstrative *ini* for in/definiteness).

(Chierchia, 1998, p. 354)

(4)   wò kànjiàn xióng le (Chinese)

       I see bear ASP

       I saw (some/the) bears

In these languages, according to Chierchia, nouns must refer to kinds. Kinds are mass-like since they do not mark a difference between singular and plural instances. Therefore, [+arg, –pred] languages lack morphological distinctions between singular and plural nouns. The same reasons explain why languages of this type do not allow nouns to combine directly with numerals and make use of classifiers instead (Chierchia, 1998, p. 353).

b) Languages with exclusively predicative NPs [*-arg, +pred*]. This includes languages like Arabic and Spanish where nouns cannot be used in argument position unless they are licensed by determiners i.e. project D to turn NPs into arguments. In these languages, bare nouns cannot occur in argument position:

(5) a.   Alsoqour tastad al3asafeer.

             The falcons hunt the birds.

     *b. *   soqour tastad alasafeer

             Falcons hunt the birds.

Note that the noun *soqour* cannot occur without the determiner *al-* which is the Arabic definite article.

c) Languages with mixed types of nouns [+*arg*, +*pred*] where NPs can either denote kinds or predicates. In Germanic languages such as English and German, some nouns can be [+arg] so occur bare, while others can be [+pred] and so need articles.

(6)   a. Falcons hunt birds. (Plural)

   b. A falcon hunts birds. (Singular)

   c. Rice is a grain (Mass)

Table 1. Parameter settings of the nominal mapping parameter (adopted from Chierchia, 1998)

Example Languages	Characteristics	Setting
Thai, Chinese, Japanese, Malay	Bare arguments No number distinction in nouns Obligatory use of classifiers for counting All nouns are mass nouns	[+arg, -pred]
French, Italian, Arabic	No bare arguments: determiners required Number distinction in nouns Count/mass distinction	[-arg, +pred]
English, Hindi	Some arguments contain bare nouns Some arguments require determiners Subject to a number distinction	[+arg, +pred]
No known languages because this setting means that nouns have no interpretation at all	Nouns are neither inherently argumental nor predicative	[-arg, -pred]

Our object languages were chosen for this study because they exemplify the three settings and represent the different categories of the Nominal Mapping Parameter. English has a [+arg, +pred] setting of the NMP. Malay has a [+arg, -pred] setting while Arabic has a [-arg, +pred] setting. We see next how this impacts on the expression of generic reference.

5. In/Definiteness, Number Marking and Generic Expression in English

Generic expression is closely associated with the article system in languages with articles. The article system, in turn, is often associated with in/definiteness except in languages where articles mark specificity (which is outside the scope of this account, since none of our three languages mark it). English is a two article language. It has the definite article *the* and the indefinite article *a/an* along with bare forms.

In English, definite singular count nouns, indefinite singular count nouns, bare plural count nouns and bare mass nouns can convey genericity. Definite plurals are not allowed to express generic meaning except in names of nationalities like *the Greeks love frappe* and this has been the standard view of English generics (e.g., the English grammar of Quirk et al., 1985, p. 265; Dahl, 1975; Lyons , 1977; Carlson, 1977b):

(7)   a. The falcon hunts birds (singular definite generic)

   b. A falcon hunts birds. (singular indefinite generic)

   c. Falcons hunt birds. (plural indefinite generic)

   d. Rice is nutritious. (bare mass generic)

   e. *The falcons hunt birds. (plural definite generic)

Krifka et al. (1995) however suggest that in fact two subtypes of genericity are marked in English. The first type requires NPs with kind reference (i.e. generic reference conveyed by the NP itself, see section 2 above) and they propose that only singular definite count nouns, bare plural count nouns and bare mass nouns can carry kind reference. Krifka et al. (1995, p. 10) further argue that certain predicates exclusively allow kind-referring arguments and show this clearly. For example, the subject argument of the predicate *die out* or *be extinct* and the object argument of *invent* can only be kinds (conveyed by NPs with inherently generic reference) since only kinds can *be invented* or *be extinct*. Indeed this applies even when the sentence as a whole is episodic rather than

generic. This has been used as proof that indefinite noun phrases do not usually have kind reference and so that singular indefinite generics are to be treated differently:

(8)   a. The lion will become extinct soon.

        b. Lions will become extinct soon.

        c. *A lion will become extinct soon.

[Krifka et al., 1995, pp. 10-23]

In such sentences *the lion* or *lions* refers to the species lion, as manifested by all lions together, since extinction can apply only to that collective entity, not to individual lions separately. By contrast in (7) it is a characteristic or habit of individual falcons to hunt birds: the species as such does not do it, but its members do it (Cohen, 2001). The generic meaning of *a lion* in such sentences is more 'If something is a falcon, it hunts birds'. Krifka et al. (1995) therefore distinguish between kind referring generics like (7) above where the genericity comes from the NP denoting a whole category and what they term 'characterizing' generic sentences like (8) where the genericity comes from the sentence context which describes what members of a category typically are or do (see section 2).

According to Krifka et al. (1995), then, English generic sentences can have NP-level (kind-denoting) genericity or sentence-level (characterizing or habitual sentence) genericity. Therefore, we have three possibilities with regards to the article system in English:

• The indefinite article *a* can only be used in generic meaning in sentence-level generic contexts.

• The definite article *the* is used generically primarily in singular NP-level generic contexts, and secondarily in sentence level generic contexts (for well-established categories, see below)

• The zero article bare form of mass and plural count nouns can be used with generic meaning in both S-level and NP-level generic contexts.

There is however, a limitation on *the* in characterising sentences that only allows it to occur with well-established categories/kinds.

(9)   a. The bear can be rather dangerous

        b. The brown bear can be rather dangerous

        c. *The angry bear can be rather dangerous

        d. A bear can be rather dangerous

        e. A brown bear can be rather dangerous

        f. An angry bear can be rather dangerous

In these examples *brown bear* refers to a distinct well-established species of bear, not just any brown bear. By contrast *angry bear* is a free combination denoting any bear that happens to be angry, so is not 'established' or institutionalised in the same way.

From the above discussion, we can see that English generic sentences can contain both definite and indefinite NPs singular and plural. The only combination not allowed is definite plurals. However, it is clear that these forms are not interchangeable semantically.

Table 2. Noun forms and their generic readings in English

Readings		
Characterizing sentences Like *X can be dangerous*	KIND-requiring sentences Like *X is extinct*	FORMS
√	√	Bare Plural
√ [1]	√	Definite Singular
*	*	Definite Plural
√	*	Indefinite Singular
√	√	Bare Mass
1 = only with well-established kinds		

6. In/Definiteness, Number Marking and Generic Expression in Arabic

Modern Standard Arabic (MSA) has two articles and no bare form or zero article. The definite article *al-*, a bound morpheme that is prefixed to a following consonant to the front of the noun as in (10):

(10)   amama al-bait

in front D-house

In front of the house

Indefinite nouns are marked by the suffix *-n* written in the form of a nunation () though this indefinite marker is pronounced but often not written. The indefinite noun will have the same cases of nominative, accusative and genitive but with the equivalent nunation (11.b.c.d):

(11)   a.   anta sabi- un

you boy a

You are a boy.

        b.   sabi- un

        c.   sabi- an

        d.   sabi- in

Some Saudi varieties of Arabic still demonstrate this phenomenon while others (like Saudi Hijazi Arabic) have dropped the indefinite marker (*-n*) and replaced it with a single word numeral before the noun *wa:ħid* واحد \ وحده (which means *one-M* and *one-F*) to denote indefiniteness and singularity similar to Emirati Arabic (Ntelitheos, Idrissi, & Tamimi, n.d.) In this case, *wa: ħid* واحد is used not as a numeral but in a depleted or bleached sense to indicate indefiniteness. The dialects that use *wahid* never use *-n*. Most Saudi dialects use neither and only use zero. In fact, most Arabic dialects have dropped the nunation. Only some Bedouin dialects still use –*n*.(and also used in classical/formal Arabic: Fus-ha eloquent)

It is worth mentioning that Fassi Fehri (1993, p. 216) and Lyons (1999, pp. 93-94) argue against the view that (-*n*) is an indefinite article in MSA since proper nouns (which do not need definite or indefinite marking) may also carry this suffix. Proper nouns are often argued to be inherently definite and in some languages have an obligatory definite article (e.g. Greek). This favours the argument that that it is odd to find an indefinite marker on a proper noun.

In MSA count nouns are categorised into singular, dual and plural. The definite article *al* and the indefinite article –*n* are used in combination with all of those categories except the dual which does not take the indefinite article -*n* as shown in table 3 below. The fact that the dual does not take the indefinite article is also taken by some as evidence, among other evidence, that –*n* is actually not an indefinite article.

By contrast SA uses the zero indefinite article with all count noun categories. The indefinite article –*n* is not used any more and bare forms are used if nouns are not definite.

Table 3. In/Definiteness and articles in MSA and Saudi Arabic nouns

Hijaz Saudi Arabic	Most Saudi Arabic	MSA		
al-walad	*al*-walad	*al*-walad-u	Count Singular	Definite
boy-Nom-def	boy-Nom-def	boy-Nom-def		
the boy	the boy	the boy		
al-walad-ain	*al*-walad-ain	*al*-walad-aa-ni	Count Dual	
boys-TwoNom-Def	boys-TwoNom-Def	boys-TwoNom-Def		
the two boys	the two boys	the two boys		
al-awlad	*al*-awlad	*al*-awlad-u	Count plural	
boys-Nom-Def	boys-Nom-Def	boys-Nom-Def		
the boys	the boys	the boys		
al-zait	*al*-zait	*al*-zait-u	Mass	
zait-Nom-Def	zait-Nom-Def	zait-Nom-Def		

the oil	the oil	the oil		
wahid walad	walad	walad-u-n	Count Singular	Indefinite
boy-Nom-Indef	boy-Nom-Indef	boy-Nom-Indef		
a boy	a boy	a boy		
walad-ain	walad-ain	walad-aa-ni	Count Dual	
boys-TwoNom-Indef	boys-TwoNom-Indef	boys-TwoNom		
two boys	two boys	two boys		
awlad	awlad	awlad-u-n	Count Plural	
boys-Nom-Indef	boys-Nom-Indef	boys-Nom-Indef		
boys	boys	boys		
zait	zait	zait-u-n	Mass	
zait-Nom-Indef	zait-Nom-Indef	zait-Nom-Indef		
some oil	some oil	oil		

Arabic, in MSA and most dialects, only allows definite (singular, plural and mass) NPs to express generic meaning. As can be seen in (12.a.c.e), no indefinite NPs (singular 12.a, plural 12.c and mass 12.e) are allowed in generic meaning in characterising sentences (Fehri, 2004, p. 46).

(12)   a.   *kalb-u-n   yamlik-u   ?arbac-a   ?arjul-i-n

dog-Nom-IND   has-Nom   four-Acc   legs-Gen-IND

A dog has four legs.

   b.   al-kalb-u   yamlik-u   ?arbac-a   ?arjul-i-n

the-dog-Nom   has-Nom   four-Acc   legs-Gen- IND

The dog has four legs.

   c.   *kilaab-u-n   tamlik-u   ?arbac-a   ?arjul-i-n

dogs-Nom- IND   have-Nom   four-Acc   legs-Gen- IND

Dogs have four legs.

   d.   al-kilaab-u   tamlik-u   ?arbac-a   ?arjul-i-n

the-dogs-Nom   have-Nom   four-Acc   legs-Gen- IND

The dogs have four legs.

   e.   *zayt-u-n ġaali-n

oil-Nom- IND (is) expensive- IND

Oil is expensive.

   f.   al-zayt-u   ġaali-n

the-oil-Nom (is)   expensive- IND

Oil is expensive.

Looking now at Krifka's kind/NP-level and characterising/sentence-level genericity distinction, we also find that Arabic makes no overt distinction between them (13, 14). Nor does Arabic make a distinction between established and non well-established kinds (15, 16):

(13)   NP-level generics: Singular NPs

   a.   al-mamooth-u mungaridh-un

DEF-mammoth-SG extinct-SG

The mammoth is extinct.

   b.   *mamuth-u mungaridh-un

mammoth-SG extinct-SG

Mammoth is extinct.

    c.   *mamuthun mungaridh-un

        IND- mammoth-SG extinct-SG

        A mammoth is extinct.

(14)   NP-level generics: Plural NPs

    a.   al-dainasour-aat-u mungaridh-at-un

        DEF- dinosours-FEM-PLU extinct-FEM-PLU

        The dinosours are extinct.

    b.   *dainasour-at-u mungaridh-at-un

        dinosaurs-FEM-PLU extinct-FEM-PLU

        Dinosaurs are extinct.

    c.   *dainasour-at-u-n mungaridh-at-un

        dinosaurs-FEM-PLU-IND extinct-FEM-PLU

        Dinosaurs are extinct.

(15)   Well-Established

    al- dubb-u al-bunni-u khateer-un

    Def-dubb-SG Def-bunni-SG extinct-SG

    The brown bear is dangerous

(16)   Non-Well-Established

    al- dubb-u al-ghathib-u khateer-un

    Def-dubb-SG Def- ghathib-SG khateer -SG

    The angry bear is dangerous

Similar to English, the definite article is used to express NP-level genericity with singular nouns. If we consider the view that the *-n* suffix the Arabic equivalent of the indefinite article in English, Arabic does not allow this form to express genericity. Furthermore, Arabic does not allow bare forms to express genericity. Unlike English which only allows the definite article with singular nouns denoting established kinds in characterizing sentences and in sentences with kind predicates, Arabic requires the use of the definite article with singular, plural, and mass nouns to express genericity. However, in Arabic "*the milk is good for you*", "*the books are good for you*" and "*the milk is good for you*" each has a generic reading but also a specific reading. This makes definite plurals, definite mass nouns and definite singulars ambiguous in characterizing sentences (as in 12, 15, and 16) but not NP-level (kind) sentences (as in 13, 14).

7. In/Definiteness, Number Marking and Generic Expression in Malay

Malay is an article-less language in the strict sense that articles are not obligatory. It does however have the option to use articles *ini* ('this') or *itu* ('that') in a weakened sense as the def art and *se* ('one') in a weakened sense as the indefinite article. The demonstrative determiners 'kata penentu' "words that determine" (Adapted from Hassan, 1993: 54) *itu* (that) and *ini* (this) occupy final position in the NPs.

(17)   a.   Pekerja itu telah tiba.

        (Noun + Det: Dem) has arrived

        The worker has arrived

    b.   Singa ini makan danging

        Lion this eat meat

    c.   The lion eats meat

        Singa itu makan danging

        Lion that eat meat

        The lion eats meat

Marsden (1812, cit. in Chan, 1996) described *itu* as the equivalent of 'that, those, the'. Winstedt's Malay Grammar (cit. in Chan, 1996), also describes *itu* as equivalent to 'the, that, those' with the sense of 'the distant, remote in space and time.' As far as indefiniteness is concerned, there is a pre-determiner as described by Wong and Quek (2007) that occurs in front of nouns in the form of a numeral followed by a classifier that can denote indefiniteness.

(18)   se-orang askar

        one-CLS soldier

        a/one soldier

In Malay, this combination of *se-* + classifier, which is originally derived from the numeral *satu* 'one', as in *se-orang* above, is described as fulfilling the role of the indefinite article *a*. (Guilfoyle et al., 1992).

(19)   Terdapat se-buah pokok yang renek dikawasan itu

        There is one-CLS tree small area that

        There is a small tree in that area.

Interestingly, English *a/an* originated in old English as a use of the number word *one* in a weakened indefinite article sense. *The* originated also as a demonstrative meaning 'that' and developed the weakened definite article sense. So what we see in *seorang* and *itu* in Malay and *wahid* in Arabic is all very similar historically.

Bare nouns in Malay, without classifiers or reduplication, can denote singularity or plurality (Carson, 2000). However, number can also be explicitly marked in the nominal phrase through the use of classifiers or reduplication of the noun. Carson (2000) argues that all nouns in Malay are syntactically mass nouns. The examples below show that all nouns behave similarly on the syntactic level i.e. can be reduplicated (for plurals) and require a classifier to combine with numerals. Unlike Chinese, Malay only requires classifiers in noun phrases to combine with numerals while Chinese requires classifiers in noun phrases with numerals and with demonstratives (Carson, 2000, p. 13).

(20)   a.   Sup-sup itu panas

              Soup-PL the hot

              The varieties of soup are hot

        b.   Tiga mankuk sup panas

              Three bowl soup hot

              Three bowls of soup are hot

        c.   *Tiga sup panas

              Three soup hot

              Three soups are hot

        d.   Kuda-kuda itu panas

              Horse-PL the hot

              The horses are hot

        e.   *Tiga kuda panas

              Three horse hot

              Three horses are hot

        f.   Tiga ekor kuda panas

              Three CLS horse hot

              Three horses are hot

As noted by Carson (2000), classifiers and reduplication cannot co-occur, though determinera *itu* and *ini* can occur with reduplication:

(Adapted from: Carson, 2000, p. 7)

(21)   a.   Buku-buku ini berat

              Book-PL the heavy

Those books are heavy

    b.   Dia ada tiga buah buku

        She has three CLS book

        She has three books

    c.   *Dia ada tiga buah buku-buku

        She has three CLS book-PL

        She has three books

Although syntactically all nouns are mass, semantically, Malay makes a distinction between atomic nouns like *cars* (which contains individual entities), and non-atomic nouns like *oil* which does not. This distinction is made through the use of a different set of classifiers for each. Sortal classifiers are used for atomic nouns while mensural classifiers are used with nouns that are non-atomic or do not have individual units.

(22)   Sortal

        aTiga orang guru

        three CLS teacher

        three teachers

(23)   Mensural

        Dua timbun pasir

        Two CLS sand

        Two piles of sand

Like some English nouns that can have both count and mass class membership, some Malay nouns can be realized as being both and therefore can combine with both sortal and mensural classifiers. Carson (2000) proposes that Malay grammar does not make syntactic or semantic distinctions between count and mass nouns and that the distinction shown above between nouns which have natural units and nouns that do not is not grammaticalized and that all classifiers put strict semantic selectional restrictions on the properties of the nouns they classify with atomicity being only one of those restriction.

Turning now to the expression of genericity, only the unreduplicated unclassified bare noun without *se* or *ini* or *itu* can express this. The reduplicated noun cannot be used with generic meanings in Malay (Rafferty, 2002). Instead it carries a plural indefinite specific interpretation as in (24.a). The bare noun in (24.b) has a generic interpretation and has neither definite nor indefinite reading. In (24.c), the mass noun *anjing* is classified by the classifier *se-ekor* (meaning 'one') and has an indefinite nongeneric interpretation and a specific reading where only the speaker knows the dog in reference but not the hearer. It is used in a story context for example as in *There was a dog once*. The noun *anjing* in (24.d) is modified by the demonstrative *itu* that gives the noun a definite nongeneric reading.

(24)   a.   Anjing-anjing [Ind, Pl] makan daging

            dog-dog   eat   meat

            Dogs are eating meat.

    b.   Anjing [ØInd/Def Ø Num] makan daging

            Dog   eat   meat

            A dog/ the dog/ Dogs eat meat.

    c.   Se-ekor anjing   [Ind, Sng] makan daging

            one   dog   eat   meat

            A dog is eating the meat.

    d.   Anjing [Def, Ø Num]   itu makan daging

            dog   that   eat   meat

            That/the dog is eating meat.

With regard to the kind vs characterizing distinction, Malay does not distinguish kind (NP level) generics as below in (25) from characterizing (S level) generics as above in (24b) nor between established and unestablished categories. Unreduplicated, bare nouns are used for both types of generics.

(25)   Sepeda ditmukan oleh Kirkpatrick Macmillian

bicycle invented by Kirkpatrick Macmillian

The bicycle was invented by Kirkpatrick Macmillian.

8. Conclusion

To sum up, table 4 summarizes generic expression in English, Arabic and Malay with regard to in/definiteness and number.

Table 4. In/definiteness and number in Generic expression in the object languages

Malay Generic	Arabic Generic	English Generic	NP
	(+definite) *al-*	(+/-definite) *the/a*	Singular
	(+definite) *al-*	(-definite) *Ø*	Plural
Ø (neutral as to def/indef)	(+definite) *al-*	(-definite) *Ø*	Mass

In the previous discussion, we have reviewed how generic reference is made in the three languages. English and MSA both have a definiteness-based article system and have more than one article form potentially available for generic use. Malay on the other hand is an article-less language, at least it has no obligatory articles, and so only the bare form is available for generic reference. Malay however only allows simple bare forms to express genericity: the bare reduplicated noun does not have a generic interpretation. From its available two articles, Arabic only allows definite NPs to express generic meaning, with all types of nouns, singular, plural and mass. English, however, exploits all its article resources and allows definite, indefinite and bare NPs to express generic meaning with count nouns except the plural definite. With mass nouns, only the bare form is used to express generic meaning.

Only English makes syntactic distinctions between NP level (kind) and sentence level (characteristic/ habitual) generic sentence contexts and between established and non-established categories. This makes the rules governing the expression of genericity more complicated for English than for Arabic and Malay. The predictions of the previous analyses are that if Arabic L1 learners transfer the rules of their first language they will overuse the definite article to express generic reference in situations that are not permitted in English. Similarly, Malay learners might overuse zero article to express genericity in contexts not allowed by English.

EAP textbooks, such as Gramer et al. (2011), for example, tend to cover articles only in non-generic use as markers of definiteness and indefiniteness. Under our analysis in this paper, teaching and teaching materials could benefit from a more comprehensive approach to teaching articles that include generic reference and that takes into consideration the complexity of generic expression in English and indeed how diverse it is across languages particularly for learners with L1 like Arabic and Malay.

References

Allan, K. (1977). Classifiers. *Language, 53*, 285-311.

Behrens, L. (2000). *Typological parameters of genericity. Arbeitspapier Nr. 37* (Neue Folge) (pp. 1-110). Institut für Sprachwissenschaft, Universität zu Köln.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Grammar of spoken and written English*. Harlow, London UK: Pearson Education.

Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English* (Vol. 2). London/ New York MIT Press.

Caplan, N. A., & Douglas, S. R. (2011). *Q: Skills for Success: Reading and Writing*. London Oxford University Press.

Carlson, G. N. (1977). *Reference to kinds in English* (Ph.D. thesis, University of Massachusetts, Amherst).

Carson, J. C. (2000). *The semantics of number in Malay noun phrases*. University of Calgary.

Chan, K. (1996). *"Itu" and the notion of definiteness in Malay* (Unpublished dissertation (MMLS)). Universiti Malaya, Kuala Lumpur, Malaysia.

Chierchia, G. (1998a). Plurality of mass nouns and the notion of "semantic parameter". *Events and grammar*, *70*, 53-103.

Chierchia, G. (1998b). Reference to kinds across language. *Natural language semantics*, *6*(4), 339-405.

Cohen, A. (2001). On the generic use of indefinite singulars. *Journal of Semantics*, *18*(3), 183-209.

Dahl, O. (1975). "On Generics." In E. Keenan (Ed.), *Formal Semantics of Natural Language* (pp. 99-111). London & New York: Cambridge University Press.

Dayal, V. (2005). Semantic variation and pleonastic determiners: The case of the plural definite generic. Proceedings of *GLOW Asia*.

Declerck, R. (1991). A *Comprehensive Descriptive Grammar of English*. Tokyo: Kaitakusha.

Fassi Fehri, A. (1993). *Issues in the structure of Arabic clauses and words*. Dordrecht: Kluwer.

Fassi Fehri, A. (2004). Nominal classes, reference, and functional parameters, with particular reference to Arabic. *Linguistic Variation Yearbook*, *4*(1), 41-108.

Guilfoyle, E., Hung, H., & Travis, L. (1992). Spec of IP and Spec of VP: Two subjects in Austronesian languages. *Natural Language & Linguistic Theory*, *10*(3), 375-414.

Hassan., A. (1993). *Tatabahasa Pedagogi Bahasa Melayu*. Kuala Lumpur: Utusan Publications and Distributors Sdn. Bhd.

Idrissi, A. (2009). *Acquisition of Definiteness in Emirati Arabic*. The 10th UAE University Annual Research Conference. United Arab Emairates University in Al-Ain, United Arab Emirates.

Krifka, M., Pelletier, F. J., Carlson, G., ter Meulen, A., Chierchia, G., & Link, G. (1995). Genericity: an introduction. *The generic book*, 1-124.

Longobardi, G. (1994). Reference and proper names: A theory of N-movement in syntax and logical form. *Linguistic inquiry*, *25*, 609-665.

Lyons, J. (1977). *Semantics* (vols I & II). Cambridge: Cambridge University Press.

Musgrave, S. (2013). Functional categories in the syntax and semantics of Malay. *NUSA: Linguistic studies of languages in and around Indonesia*, *55*, 135-152.

Rafferty, E. (2002). *Reduplication of Nouns and Adjectives in Indonesian*. Tenth Annual Meeting of the Southeast Asian Linguistics Society, 317-332. Phoenix USA: Arizona State University.

Wong, B. E., & Quek, S. T. (2007). Acquisition of the English definite article by Chinese and Malay ESL learners. *Electronic Journal of Foreign Language Teaching*, *4*(2), 210-234.