Classifiers and the definite article in Indonesian*

Carol-Rose Little & Ekarina Winarto
Cornell University

1. Introduction

It has been said that languages with classifiers\footnote{We thank Curt Anderson, Kurt Erbach, Mary Moroney, Sarah Murray, Justin Royer, John Whitman, Mia Wiegand, and audiences at Heinrich-Heine-Universität Düsseldorf, AFLA 25, and the Cornell Semantics Group as well as three anonymous NELS reviewers for comments and discussion. We would especially like to thank Peter Sutton for discussion on Section 3. Data comes from the second author’s native judgements and consultations with other Indonesian speakers. Responsibility for any errors is our own.} do not have overt definite articles (Chierchia 1998b, Bošković 2008). Indonesian, however, is typologically rare in that it has both classifiers (Chung 2000, Sneddon 2010) and a definite article (Macdonald 1976, Rubin 2010, Winarto 2016). Classifiers in Indonesian appear with numerals but even then the classifier is optional. Whereas (1a) has a numeral followed by a noun, (1b) shows that the numeral can also appear with a classifier, in this case the classifier \textit{buah}.

\begin{tabular}{ll}
\hline
(1) & a. lima mangga  \\
 & five mango  \\
 & \textit{‘five mangoes’}\footnote{\textbf{Glosses:} CL: classifier; DEF: definite article; EXT: existential; NEG: negation; PERF: perfective aspect; PROG: progressive aspect; REDUP: reduplication}  \\
\hline
& b. lima \textbf{buah} mangga  \\
& five \textbf{CL} mango  \\
& \textit{‘five mangoes’}  \\
\hline
\end{tabular}

The fact that the classifier in Indonesian is optional has been noted by various authors (Chung 2000, Sneddon 2010, Dalrymple & Mofu 2012), but much of this discussion is on the distribution of classifiers and plural marking.

Our paper builds on Winarto (2016)’s observation that a classifier cannot co-occur with the definite article \textit{-nya}, as shown in (2a). When the classifier is not present, the definite article with the numeral and noun is grammatical as in (2b).

\begin{tabular}{ll}
\hline
\end{tabular}
In this paper, we review arguments from Winarto (2016) that 

\[-nya\] is indeed a definite article and data showing that the definite article \[-nya\] is ungrammatical with classifiers (§2). In §3, we consider analyzing the Indonesian data under a Chierchia-style analysis for classifiers and a standard analysis for the definite article (i.e., Heim & Kratzer 1998). We argue that certain challenges arise if we try to put together these accounts. We then propose an analysis where all nouns in Indonesian are type \(\text{et}\). Our analysis accounts for the following: (i) classifiers occur with numerals; (ii) NPs with classifiers are always indefinite; and (iii) classifiers and the definite article cannot co-occur. We also provide an analysis to account for the optionality of the classifier in Indonesian by adopting a numeral modifier operation building on Dalrymple & Mofu (2012) and Filip & Sutton (2017). Our analysis has implications for how classifiers are analyzed crosslinguistically: namely, we hope to highlight that classifiers can fill various functions across languages.

2. Classifiers and Definiteness in Indonesian

Indonesian is an Austronesian language spoken in Indonesia by about 198 million people. Since the sociolinguistic situation is complicated, we use data from colloquial Jakarta Indonesian, the variety spoken by the second co-author and our consultants.

2.1 Marking definiteness in Indonesian

Like in typical classifier languages, Indonesian bare nouns can get a kind interpretation (3a), generic interpretation (3b), indefinite and definite interpretation (3c).

(3) a. \textbf{Dinosaurus} sudah punah. \\
\textit{dinosaur} PERF extinct \\
‘The dinosaur is extinct.’ Kind

b. \textbf{Anjing} menggonggong. \\
\textit{dog} bark \\
‘Dogs bark.’ Generic

c. Saya melihat \textbf{mobil}. \\
I see \textit{car} \\
(i) ‘I saw a car/cars.’ \\
(ii) ‘I saw the car(s).’ Definite or indefinite depending on context

Standard grammars of Indonesian say that Indonesian bare nouns can be definite or indefinite, as in (3c). However, recently authors have noted that there is an overt definite article \[-nya\] given in (4), historically derived from the third person possessive (Sneddon 2006, Rubin 2010, Winarto 2016).
Anjing-nya menggonggong.
dog-DEF bark
(i) ‘The dog(s) are barking.’
(ii) Not: ‘Dogs bark.’ (Cannot be dog-kind either)

The article, -nya, has properties similar to the definite article the in English. While grammars of standard Indonesian like Sneddon (2010) do not discuss -nya as being a definite article, it is standard in spoken forms of all registers.

Here we review diagnostics providing evidence that -nya is a definite article. Using the law of contradiction/the consistency test (Löbner 1985), the sentence in (5a) provides evidence that anjing-nya ‘the dog’ with -nya indeed refers to the same individual.3 Namely, if ‘Rina likes the dog’ is true for the individual ‘dog’, the opposite cannot be true. Compare to anjing itu ‘that dog’ in (5b) with the demonstrative itu: here each instance of anjing itu does not refer to the same individual, therefore not creating a contradiction.

\[
\begin{align*}
\text{(5) a. } & \text{ # Rina suka anjing-nya, tapi tidak suka anjing-nya.} \\
& \text{Rina like dog-DEF but NEG like dog-DEF} \\
& \text{‘Rina likes the dog but doesn’t like the dog.} \\
\text{b. } & \text{ Rina suka anjing itu, tapi tidak suka anjing itu.} \\
& \text{Rina like dog that, but NEG like dog that} \\
& \text{‘Rina likes that dog but doesn’t like that dog.}
\end{align*}
\]

Furthermore, to indicate unique entities like ‘the sun’ in (6), the definite article -nya can be used, though it is optional. Winarto (2016) also notes that -nya cannot be replaced with the demonstrative itu in (6).

\[
\begin{align*}
\text{(6) } & \text{ Matahari(-nya) panas sekali hari ini.} \\
& \text{sun(-DEF) hot very day this} \\
& \text{‘The sun is very hot today.’ } \quad \text{(Adapted from Winarto 2016: 226)}
\end{align*}
\]

The definite article -nya must be used in bridging contexts as in the product-producer one in (7).

\[
\begin{align*}
\text{(7) } & \text{ Budi baru membeli sepasang sepatu. Designer*(-nya) terkenal di Paris.} \\
& \text{Budi just buy pair shoes designer-DEF famous in Paris} \\
& \text{‘Budi just bought a pair of shoes. The designer is famous in Paris.’ } \quad \text{(Winarto 2016: 227)}
\end{align*}
\]

Another diagnostic for definiteness comes from Milsark (1977)’s ‘Definiteness Restriction’. This restriction states that definites should not be able to appear in theme position of

\footnote{3See, however, Moroney (2019) for potential challenges to this diagnostic.}

\footnote{4One may question if -nya in (7) is really a definite article or if it is a resumptive or possessive pronoun, given that -nya is homophonous to the third person possessive pronoun. It is possible that it is ambiguous here between these two meanings. Though teasing these two meanings apart is beyond the scope of this paper, we note that it may be that -nya historically developed into a definite article through contexts like in (7).}
existential predicates. Indeed, NP-nya is ungrammatical with the existential predicate ada in (8).

(8) * Di tempat parkir ada mobil-nya.
    PREP place park EXT car-DEF
    Intended: ‘There are the cars in the parking lot.’

Given these diagnostics, we take -nya to be an overt definite article in Indonesian.

2.2 Classifiers

It has been reported that Indonesian has as many as 60 classifiers (Dardjowidjojo 1978), but the most commonly used ones today are buah (general), orang (people) and ekor (animals). Classifiers appear with numerals but even then they are optional as indicated in (9a). A classifier and noun without a numeral is ungrammatical as indicated in (9b).

(9) a. lima (buah) mangga
    five CL mango
    ‘five mangoes’

    Furthermore, numerals with classifiers cannot be definite. If a speaker introduces two dogs and three cats into the discourse as in (10a), they cannot refer back to them with the numeral-classifier phrases as in (10b).

(10) a. Saya lihat dua ekor anjing dan tiga ekor kucing.
    I see two CL dog and three CL cat.
    ‘I saw two dogs and three cats.’

    b. # Tiga ekor kucing lari ke taman.
    three CL cat run to park
    Intended: ‘The three cats ran towards the park.’

    Numeral-classifier phrases are also grammatical in existential constructions (11), as expected.

(11) Di tempat parkir ada dua buah mobil.
    PREP place park EXT two CL car
    ‘There are two cars in the parking lot.’

    The example in (11) contrasts to (8): a noun or numeral phrase with -nya is ungrammatical in theme position of existential predicates.

---

5 This sentence is grammatical with the meaning ‘His/her cars are in the parking lot.’
6 Because many classifiers come from nouns, buah by itself simply means ‘fruit’ so (9b) is grammatical if the intended meaning is ‘mango fruit’.
7 Classifiers can also appear with some weak quantifiers like beberapa ‘a few/several’ but not with strong quantifiers like semua ‘all’. In this paper, we concentrate on their appearance with numerals.
2.3 Classifiers and -ny a cannot co-occur

Numerals modifying nouns can co-occur with the definite article -nya in (12a). But, as noted by Winarto (2016: 232), classifiers may not co-occur with -nya (12b).

\[(12)\]
\[
\begin{array}{l}
\text{a. lima mangga-nya} & \text{b. * lima buah mangga-nya} \\
\text{five mango-DEF} & \text{five CL mango-DEF} \\
\text{‘the five mangoes’} & \text{Intended: ‘the five mangoes’}
\end{array}
\]

There is one construction, given in (13), where the definite article appears with the classifier, however we take this to be a case of adjunction and distinct from (12b).

\[(13)\]
\[
\text{kucing-nya dua ekor} \\
\text{cat-DEF two CL} \\
\text{‘the cats, two of them’}
\]

While -nya cannot attach to the noun with a prenominal numeral and classifier in (12b), in (13) the numeral and classifier are postnominal. As seen in (13), the postnominal numeral-classifier phrase can co-occur with -nya. We analyze the numeral-classifier phrase as an adjunct and we concentrate on the incompatibility of the classifier and definite article in (12b).

Finally, though the definite article historically comes from the third person possessive (Sneddon 2006, Rubin 2010), possessives, unlike the definite article, are grammatical with classifiers as in (14) for the first person and third person possessive.

\[(14)\]
\[
\begin{array}{l}
\text{a. dua orang anak saya} & \text{b. dua orang anak-nya} \\
\text{two CL child 1.POSS} & \text{two CL child-3.POSS} \\
\text{‘my two children’} & \text{‘his/her two children’} \\
\end{array}
\]

We take the example in (14b) as evidence that -nya as a third person possessive and -nya as the definite article are homophonous and therefore two distinct morphemes.\(^8\)

\(^8\)Unlike in English, numerals with possessives in Indonesian do not introduce uniqueness presuppositions. For example, (1b) is a felicitous follow up to (1a), showing that there is no entailment on the number of children the speaker in (1) has, unlike ‘my two children’ does in English.

\[(1)\]
\[
\begin{array}{l}
\text{a. Dua orang anak saya tinggal di Amerika.} \\
\text{two CL child my PREP America} \\
\text{‘My two children/two of my children live in America.’} \\
\end{array}
\]
\[
\begin{array}{l}
\text{b. Saya juga ada satu anak lain yang tinggal di Australia.} \\
\text{I also have one child other that lives PREP Australia.} \\
\text{‘I also have another child that lives in Australia.’}
\end{array}
\]
2.4 Summary

Indonesian is typologically rare in that it has both classifiers and a definite article. However, they cannot co-occur. In the remainder of this paper, we propose an analysis that accounts for the following:

- Classifiers occur with numerals.
- NPs with classifiers are always indefinite.
- Classifiers and the definite article cannot co-occur.

In the next section, we first explore combining a Chierchia-style analysis for classifiers (i.e., classifiers \textit{qua} individuating functions (Chierchia 1998b)) and standard analyses for definite determiners (e.g., Heim & Kratzer 1998). We discuss consequences of this type of analysis for noun denotations in Indonesian. We then propose an alternative analysis where the classifier takes both a numeral and a predicate as its arguments, and indefiniteness is encoded in the classifier. We demonstrate that encoding indefiniteness into the classifier blocks the application of the definite article.

3. Analysis

3.1 Two systems in one?

Here we consider what would happen if we take the view that classifiers are individuating (Chierchia 1998b) and a standard view of the definite determiner (Heim & Kratzer 1998). Perhaps Indonesian has a reflection of these two systems in its grammar.

Chierchia (1998b) proposes that classifiers are function from kinds to the set of all parts of the kind in a world (15a) and represented by $\exists$. This allows the numerals in (15b) to combine with nouns mediated by the classifier.

(15) Classifiers and numerals in Indonesian in a Chierchia-style analysis

\begin{enumerate}
  \item $\llbracket \text{buah} \rrbracket = \text{\textsc{u}}$
  \item $\llbracket \text{lima} \rrbracket = \lambda P \lambda x. \text{[} P(x) \text{ ] } \& \text{[} \mu_y (x) = 5 \text{ ]}$
\end{enumerate}

A standard semantics for definite determiners is given in (16), i.e., the Heim & Kratzer 1998-view of definite articles as type $\langle(\text{et})\text{e}\rangle$.

(16) $\llbracket \text{-nya} \rrbracket = \lambda P ty. [P(y)]$

A consequence of proposing the definitions in (15) and (16) is that nouns in Indonesian must be kind-denoting as per (17a) in order to combine with the classifier in (15a), or nouns are predicate-denoting as in (17b) so that they can combine with the definite article in (16).
An ambiguity proposal for nouns

a. \[ [\text{mangga}_1] = \text{^}\text{MANGO} \] (the mango-kind)

b. \[ [\text{mangga}_2] = \lambda x. [\text{MANGO}(x)] \] (the set of individual mangoes)

The kind-denoting noun reflects analyses for classifier languages that need a classifier to individuate the noun to combine with a numeral. The predicate-denoting noun reflects languages with definite articles where the definite article combines with predicates to create individuals. The ambiguity proposal in (17) accounts for why the definite article can combine directly with a predicate-denoting noun (i.e., \text{mangga}_2). It also accounts for why classifiers are needed as they must mediate between \text{mangga}_1 and numerals. Though numerals may also combine directly with \text{mangga}_2, thus accounting for the optionality seen with classifiers.

However, some challenges arise with this analysis. First, it creates redundancy in the lexicon as it would mean needing to posit two denotations for every noun. Secondly, it does not capture why the definite article can’t co-occur with the classifier. After application of the classifier and numeral, the phrase is type \langle \text{et} \rangle so the definite article should be able to combine with that phrase. Finally, this lexical ambiguity account does not explain why classifiers are dependent on numerals: the classifier semantics in (15a) assumes that classifiers are needed for nouns, not numerals. Next, we propose an analysis that we argue can account for these empirical facts.

3.2 Proposed analysis

We propose that nouns in Indonesian are predicates of type \langle \text{et} \rangle in (18) following other work (Chung 2000 and Dalrymple & Mofu 2012 for Indonesian and Nomoto 2013 for Malay). The * operator in (18) indicates that bare nouns are interpreted as complete semilattices and thus include single atoms as well as the sums of those atoms in the denotation, capturing that nouns in Indonesian have general number (Corbett 2000) or are number neutral (Rullmann & You 2006, Wilhelm 2008, Chierchia 1998a).^9

\[ [\text{mangga}] = \lambda x. [^\text{MANGO}(x)] \]

The denotation in (18) generates an atomic set with countable atoms and groupings of those atoms.

We propose that the definite article -\text{nya} is a function from predicates to individuals (\langle \langle \text{et} \rangle \text{e} \rangle) (following Sharvy (1980), Landman (2004), where definiteness with pluralities can come through a maximality condition) in (19).

\[ [\text{-nya}] = \lambda P. \text{sup}(P) \] if defined; undefined otherwise
where \text{sup}(P) = 1x[P(x) \& \forall y[P(y) \rightarrow y \leq x]]

^9In this paper, we do not address the issue of the mass/count distinction in Indonesian. We refer readers to Dalrymple & Mofu (2012) who claim, using data with the reduplicative plural and numeral modification, that there is no mass/count distinction in Indonesian.
Since numeral–classifier phrases are inherently indefinite, we propose that the classifier enclosure this indefiniteness via a choice function \((f)\) of type \((\langle \text{et} \rangle e)\) subject to existential closure in (20a) (see Reinhart (1997), Winter (1997) for choice function analyses of indefinites).\(^{10}\) The classifier takes both a numeral \(n\) in (20b) and a predicate \(P\) in (18) as its arguments. Finally, \(\mu_h\) is a measure function from a group to the cardinality of that group.

\[
(20) \quad \text{Proposed analysis for Indonesian numerals and classifiers}
\]

a. \([\text{buah}] = \lambda n \lambda P. f(\lambda x. [\mu_h(x) = n & P(x)])\) \(f\) is bound by existential closure  
b. \([\text{lima}] = 5\)

The classifier first takes a numeral as its argument in (21) and then a noun.

\[
(21) \quad \text{lima buah mangga ‘five CL mangoes’}
\]

The resulting NP in (21) is an individual of type \(e\) and therefore the definite article, which takes a predicate of type \(\langle \text{et} \rangle\) cannot combine with the NP in (21) as it results in a type mismatch.\(^{11}\) This type mismatch is exemplified in (22).

\(^{10}\)We use a choice function analysis for indefinites here that, for Reinhart, may be closed at multiple places in the derivation. This may overgenerate possible scope readings in Indonesian. The point here is that the classifier encodes indefiniteness, which blocks the application of the definite article. We return briefly to the question of indefiniteness in Indonesian in the conclusion.

\(^{11}\)Recall that NPs like (21) can also appear in theme position of the existential predicate. We believe this is not a problem as the existential predicate is the same as the locative predicate. Crosslinguistically, it is common for locative and existential predicates to be the same (Freeze 1992). One possibility is that when (21) composes with \(\text{ada}\) (the existential predicate), \(\text{ada}\) actually acts as a locative.
Classifiers and the definite article in Indonesian

(22) *lima buah mangga-nya Intended: ‘the five CL mangoes’

Our proposed analysis for nouns, classifiers and numerals captures that: (i) classifiers occur with numerals as the classifier takes a numeral as one of its arguments; (ii) numeral-classifier-noun phrases are indefinite by having a choice function in the semantics of the classifier; and (iii) the definite article may not apply to a numeral-classifier-noun phrase as they are already arguments of type e so, this would incur a type mismatch. Our proposed analysis in (20) is similar to Krifka (1995)’s classifier semantics for Mandarin, but, for Krifka, nouns in Mandarin are kind-denoting. Our proposal here is also consistent with the view that classifiers are for numerals, not nouns (see discussion in Bale & Coon (2014)).

Next, we posit a way to account for the optionality of the classifier.

3.3 Accounting for optionality of the classifier

To account for the optionality of Indonesian classifiers, we posit that there is a null modifying operator similar to that posited in Filip & Sutton (2017). This operator turns the numeral into a modifier that can then take a predicate nominal as an argument. This is also similar to the proposal in Dalrymple & Mofu (2012), who account for the optionality of classifiers in Indonesian by positing a contextually supplied classifier CL when there is no overt classifier. Indefiniteness is not encoded into the semantics of this operator in (23) so the numeral noun phrase is type e and therefore does not block application of the definite article.

(23) Numeral shifting operation simplified\(^{12}\) from Filip & Sutton (2017: 352)

\[
\text{MOD} = \lambda n \lambda P \lambda x. [P(x) \& \mu_\#(x) = n]
\]

The operator MOD mediates between the numeral and noun in (24) generating a set, rather than an individual as was proposed in (21).

\(^{12}\)Filip & Sutton (2017) give the following definition for MOD:

(1) \[
\text{MOD} = \lambda n \lambda P \lambda x. [\pi_1(P(x)), \mu_{\text{card}}(x, \pi_2(P(x))) = n, \text{QUA}(\pi_2(P(x)))]
\]
Little & Winarto

(24) *lima mangga* ‘five mangoes’

\[
\lambda x. \left[ \text{MANGO}(x) \land \mu_x(x) = 5 \right]
\]

We therefore propose encoding indefiniteness into the overt classifier but not into the MOD operator. This proposal captures that the numeral-classifier-noun phrase is incompatible with the definite article but numeral-noun phrases are not. The definite article in (19) may combine with the numeral-noun phrase of type \( \langle \text{et} \rangle \) generated in (24).

4. Conclusions

In this paper, we have provided evidence that Indonesian has both numeral classifiers and a definite article. However, prenominal classifiers cannot appear with the definite article. We considered combining a Chierchia-style analysis for classifiers and a Heim & Kratzer-style analysis for definite determiners. We argued that this would create unnecessary redundancy in the lexicon and would not derive the incompatibility of the definite article and classifiers. In our proposed analysis, we take the view that all nouns in Indonesian are predicate-denoting. We use a standard analysis of the definite article that it is a function from predicates to individuals. Classifiers encode an indefiniteness function and take both numerals and nouns as arguments. Our proposed analysis accounts for the data as the indefiniteness function in the classifier blocks the application of the definite article. When the numeral and classifier combine with the noun, it generates an argument which then cannot combine with a definite article. We posited a null MOD operator (Filip & Sutton 2017) to account for the optionality seen in classifiers in Indonesian. This null MOD operator mediates between the numeral and the noun but crucially does not include indefiniteness into its semantics. Therefore, it does not block the application of the definite article in Indonesian to numeral-noun phrases.

For future research, we hope to explore the different dimensions of indefiniteness in Indonesian. In Indonesian, bare nouns can be interpreted as indefinite. Additionally *se-orang*, composed of the reduced form of the numeral one and the human classifier *orang*, gives rise to an indefinite interpretation. The numeral *satu* ‘one’ without the classifier is another option for expressing indefiniteness. Preliminary data suggest that *satu* forces a wide scope reading whereas bare nouns tend to have narrow scope.
Finally, the co-occurrence of definite articles and classifiers in one language may not be as uncommon as previously thought (see, e.g., Jiang 2018 for discussion of the Sino-Tibetan language Nuosu Yi). We hope that with this research we have highlighted that classifiers can fulfill various functions crosslinguistically and interact differently with definiteness.

References


Carol-Rose Little, Ekarina Winarto
crl223@cornell.edu, ew385@cornell.edu