

Basic Copula Clauses in Indonesian

David Moeljadi, Francis Bond, and Luís Morgado da Costa

Division of Linguistics and Multilingual Studies

Nanyang Technological University, Singapore

D001@ntu.edu.sg, bond@ieee.org, luis.passos.morgado@gmail.com

1 Introduction

Every language has a copula clause type which may have a copula verb. Some languages lack a copula verb; the copula slot is left blank and we have ‘verbless clauses’. In addition, some languages have more than one copula verb. Most commonly, one will just refer to ‘a state’ and the other to ‘coming into a state’, similar to ‘be’ and ‘become’ in English (Dixon, 2009, p. 175). In this paper, we want to limit our discussion to the stative ‘be’ clause. Indonesian, a Western Malayo-Polynesian language of the Austronesian language family,¹ has multiple copula verbs, distributed over different semantic relations, and at the same time, has ‘verbless clauses’.

Analyses of Indonesian copulas can be found in reference grammars, such as Alwi et al. (2014), Mintz (2002), and Sneddon et al. (2010). Syntactic analysis in Lexical Functional Grammar (LFG) (Kaplan and Bresnan, 1982) was done by Arka (2013). However, to the best of our knowledge, no work has been done on modeling Indonesian copula clauses in Head Driven Phrase Structure Grammar (HPSG) (Sag et al., 2003) and Minimal Recursion Semantics (MRS) (Copestake et al., 2005). This paper aims to fill in this gap, referring to existing HPSG literature on copulas, such as Bender (2001) and Van Eynde (2009). Our analysis is implemented in the Indonesian Resource Grammar (INDRA), a computational grammar for Indonesian (Moeljadi et al., 2015).

Basic copula clauses in Indonesian can roughly

¹Indonesian (ISO 639-3: ind) belongs to the Malayic branch with Standard Malay spoken in Malaysia, Brunei Malay in Brunei, Local Malay in Singapore and other Malay varieties spoken at various places in Indonesia (Lewis, 2009). The Indonesian language is spoken mainly in the Republic of Indonesia as the sole official and national language and as the common language for hundreds of ethnic groups living there (Alwi et al., 2014, p. 1-2). In Indonesia it is spoken by around 22.8 million people as their first language and by more than 140 million people as their second language. It is over 80% cognate with Standard Malay (Lewis, 2009).

be divided into three types, depending on the part-of-speech of the predicate: noun phrase (NP), adjective phrase (AP), or prepositional phrase (PP). The one with NP predicate typically express the notions of ‘proper inclusion’ and ‘equation’, the one with AP predicate expresses ‘attribution’, and the one with PP predicate typically expresses ‘location’ (Payne, 2008, p. 111-123). Table 1 shows an outline of the three types of basic copula clauses in Indonesian.

Relation	Subject	Predicate
Proper inclusion,	<i>Budi</i>	(<i>adalah</i>) <i>guru</i> (NP)
Equation	<i>Budi</i>	is a teacher
Attribution	<i>Budi</i>	\emptyset <i>pandai</i> (AP)
	<i>Budi</i>	is clever
Location	<i>Budi</i>	(<i>ada</i>) <i>di rumah</i> (PP)
	<i>Budi</i>	is at home

Table 1: Three types of basic copula clauses in Indonesian

All three types of basic copula clauses in Table 1 can appear in ‘verbless clauses’. Moreover, ‘Attribution’ is best expressed without a copula verb. The copula verbs shown in Table 1 are *adalah*² for ‘proper inclusion’ and ‘equation’, and *ada* for ‘location’. However, as mentioned before, there are more than one copula for some semantic relations as discussed in Section 2.

2 Basic Data

2.1 Copula clauses with NP predicate

Copula clauses with an NP as predicate may or may not have a copula verb *adalah*, *ialah*,³ or *merupakan*⁴ as predicate (Alwi et al., 2014, p. 358-359). These clauses express the notions of

²*adalah* is derived from the existential verb *ada* and a focus particle *-lah*.

³*ialah* is derived from 3SG *ia* ‘s/he’ and a focus particle *-lah*.

⁴*merupakan* is derived from a noun *rupa* ‘form, figure, appearance, sort’, an agent-trigger prefix *me-*, and an applicative suffix *-kan*. The original meaning is ‘to form, to shape, to constitute’.

‘proper inclusion’ and ‘equation’. Indonesian does not distinguish these notions syntactically, as shown in Example [1a] and [1b]. The three copula verbs behave the same way.

Since *ialah* is historically derived from 3SG *ia*, it only occurs with a third person subject (Sneddon et al., 2010; Mintz, 2002). Example [1c] shows that *saya* “1SG” cannot be the subject of a copula clause with *ialah*.

The copula verb *merupakan* is a verb (see Footnote 4) which is in the process of becoming a copula. At the present stage it cannot appear if the NP predicate is a specific referent, such as a proper name, demonstrative, or pronoun, as shown in Example [1d]. However, it can precede a unique referent NP with a definite marker or a possessive marker as shown in Example [1b]. This has been confirmed in the Indonesian text in the Nanyang Technological University — Multilingual Corpus (NTU-MC) (Tan and Bond, 2012), which contains 2,975 sentences from three sources: Singapore Tourism Board website *www.yoursingapore.com*, a Sherlock Holmes short story “the Speckled Band”, and a short story written by Akutagawa Ryunosuke “The Spider’s Thread”.

- (1) a. *Budi (adalah/ialah/merupakan) guru.*
 Budi COP teacher
 “Budi is a teacher.”
- b. *Budi (adalah/ialah/merupakan) guruku.*
 Budi COP teacher=1SG
 “Budi is my teacher.”
- c. **Saya ialah guru.*
 1SG COP teacher
 Intended meaning: I am a teacher.
- d. **Orang itu merupakan Budi.*
 person that COP Budi
 Intended meaning: That person is Budi.

2.2 Copula clauses with AP predicate

Copula clauses which express the notion of ‘attribution’ are the ones which have an AP as the main semantic content. They are called ‘predicate adjectives’ in Payne (2008, p. 120-121). A copula is usually absent in predicate adjectives, as shown in Example [2a]. As Sneddon et al. (2010, p. 246-247) note, a copula may be used by some speakers in adjective clauses, as illustrated in Example [2b]. However, since not all speakers agree

with this and we did not find any occurrence of predicate adjectives with copulas in NTU-MC, we will not discuss this particular clause in this paper. Arka (2013, p. 31, 33) notes that a copula cannot precede an adjective.

- (2) a. *Budi pandai.*
 Budi clever
 “Budi is clever.”
- b. *?Pernyataan itu (adalah/ialah) benar.*
 statement that COP true
 “That statement is true.” (based on Sneddon et al. (2010, p. 247))

2.3 Copula clauses with PP predicate

Copula clauses which express the notion of ‘location’ are the ones which have a PP as the main semantic content. They are called ‘predicate locatives’ in Payne (2008, p. 121-123). An existential verb *ada* or *berada* may be used optionally in predicate locatives.

- (3) *Budi (ada/berada) di rumah.*
 Budi COP at home
 “Budi is at home.”

There is another ‘benefactive’ clause in which the main semantic content of the predication is realized in a PP and its syntactic pattern usually follows the one of predicate locatives (Payne, 2008, p. 122). In Indonesian, an optional copula verb *adalah* or *ialah* may appear in this ‘benefactive’ clause, as shown in Example [4a]. Sneddon et al. (2010, p. 246) state that in Indonesian prepositional clauses, the PP occurs as predicate.

Regarding *ialah*, for the same reason mentioned in Section 2.1, it can only appear with a third person subject. Example [4b] shows that *engkau* “2SG” cannot be the subject of *ialah*.

- (4) a. *Ini (adalah/ialah) untuk Budi.*
 This COP for Budi
 “This is for Budi.”
- b. **Engkau ialah untukku.*
 2SG COP for=1SG
 Intended meaning: You are for me.

3 Analysis

3.1 Copula clauses with NP predicate

Using the HPSG framework (Sag et al., 2003), we analyzed copula verbs *adalah*, *ialah*, and *merupakan* as transitive verbs, denoting a relation. Our

analysis follows the Montagovian treatment as mentioned in Van Eynde (2009, p. 368).

The copula *ialah* is an instance of *v_np_cop_3_le* which inherits from *v_np_cop_noasp_le* with a constraint: the subject should be third person. The copula *adalah* is an instance of *v_np_cop_noasp_le* which inherits from *cop-verb-lex* with a constraint: it cannot occur with any aspect or tense marker. The copula *merupakan* also inherits from *cop-verb-lex*, but with a different constraint: the head of the complement should be a common noun, not a proper noun, a pronoun, or a demonstrative. We divided *noun* into *commonnoun*, *propername*, and *pronoun*. Finally, the type *cop-verb-lex* inherits from *transitive-verb-lex* with an obligatory complement (see Figure 1).

We use MRS (Copestake et al., 2005) as semantic framework. The MRS representation is the same as the one for transitive sentences (see Figure 2). The value of ARG0 of the semantic head daughter *_cop_v_ialah_rel* is an event (*e2*) which is equated with the INDEX. Its ARG1 has a constraint: the value of the PNG.PERNUM is *3sg*. The value of ARG0 of *named_rel* “*budi*” (*x3*) and *_guru_n_rel* (*x9*) refer to the value of the ARG1 and ARG2 feature of the semantic head daughter respectively.

For zero copula clauses, we made a pumping rule which pumps or converts an NP to a VP (see Figure 3). This rule is a unary rule having only one daughter (Copestake, 2002, p. 120). This syntactic structure is similar to the one in Arka (2013, p. 38) where any lexical category (VP, NP, AP, and PP) can be a predicate XP, the NP subject takes this XP to make an Indonesian clause. Our analysis is also corresponding to ‘Constructional analysis II’ in Bender (2001, p. 101-118). There are three kinds of facts which make it unsuccessful to deal with African American Vernacular English (AAVE) copula absence: the possibility of copulaless existentials, a curious interaction of negation and ellipsis, and the possibility of complement extraction (Bender, 2001, p. 107). These three things do not exist in Indonesian: Indonesian has an obligatory existential verb *ada* with no copula, compared with AAVE which has *there* and a zero copula in existential sentences; AAVE has the possibility of copula ellipsis in case it strands *not*, while Indonesian uses a different expression or a construction in which the complements of the copula not elided; finally, AAVE has a long dis-

tance dependency in which the complement of the silent copula can be extracted, while Indonesian does not have this kind of phenomenon. In short, because of differences in syntactic structure, the constructional analysis which does not work for AAVE, can be implemented for Indonesian.

This rule introduces a predicate *cop_v_zero_rel* with the subject as the first argument and the NP predicate as the second argument, denoting a relation of coreference between them, covering both equational (identificational) and proper inclusion (predicational) relations. The MRS is similar to the one with a copula verb *adalah*, *ialah*, or *merupakan*.

3.2 Copula clauses with AP predicate

As mentioned in Section 2.2, the predicate and the main semantic content of copula clauses with AP predicate is the AP. We treat adjectives as intransitive predicates which take an NP as the subject. Figure 4 shows the parse tree of Example [2a]. The MRS representation is the same as the one for intransitive sentences (see Figure 5). The value of ARG0 of the semantic head daughter *_pandai_a_rel* is an event (*e2*) which is equated with the INDEX. Its ARG1’s value (*x3*) refers to the value of ARG0 of its subject *named_rel* “*budi*”.

3.3 Copula clauses with PP predicate

Predicate locatives have a PP as the main semantic content and an optional verb *ada* or *berada*. We treat *ada* and *berada* as auxiliaries which do not introduce a predicate of their own. The head of the subject is a noun and the head of the complement is a preposition. Figure 6 shows the parse tree of Example [3] with an existential verb *ada*.

For predicate locatives without copula, we made a pumping rule which pumps or converts a PP to a VP (see Figure 7). This rule is similar to the one in Section 3.1. The difference is that this pumping rule for PP does not introduce an additional *cop_v_zero_rel* predicate. The MRS is exactly the same as the one for predicate locatives with *ada* (see Figure 8).

In the MRS representation, the value of ARG0 of the semantic head daughter *_di_p_rel* is an event (*e2*) which is equated with the INDEX. The value of its ARG1 and ARG2 refer to the value of ARG0 of *named_rel* “*budi*” (*x3*) and *_rumah_n_rel* (*x9*) respectively.

Regarding ‘benefactive’ clauses, our analysis is the same as the one for predicate locatives. We

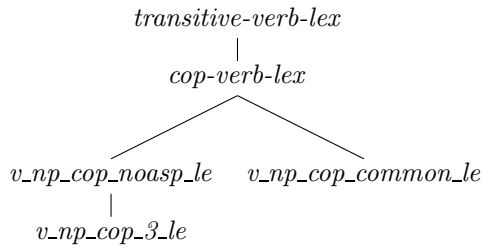


Figure 1: Type hierarchy of Indonesian copula verbs

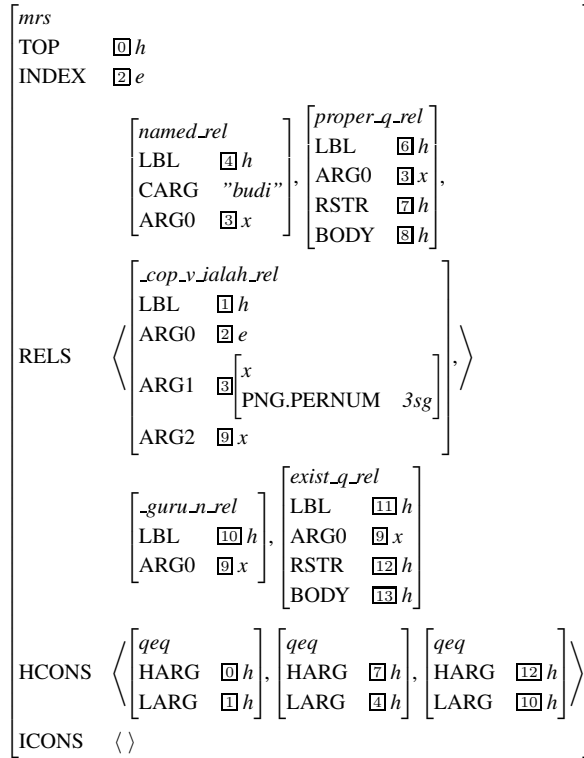


Figure 2: MRS representation of *Budi ialah guru* “Budi is a teacher”

treat *adalah* and *ialah* in these clauses as auxiliaries which do not introduce a predicate. The MRS representation is similar to the one in Figure 8.

4 Conclusion

Our analyses of Indonesian copula clauses are similar to Arka (2013)’s LFG analysis but cover more copula verbs with a refined type hierarchy. Because of differences in syntactic structure between AAVE and Indonesian, the analysis that builds a VP out of a predicative NP, which does not work for AAVE, can be successfully implemented for Indonesian.

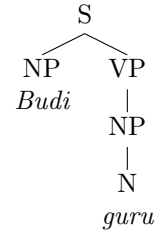


Figure 3: Parse tree of *Budi guru* “Budi is a teacher”

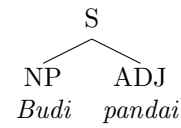


Figure 4: Parse tree of *Budi pandai* “Budi is clever”

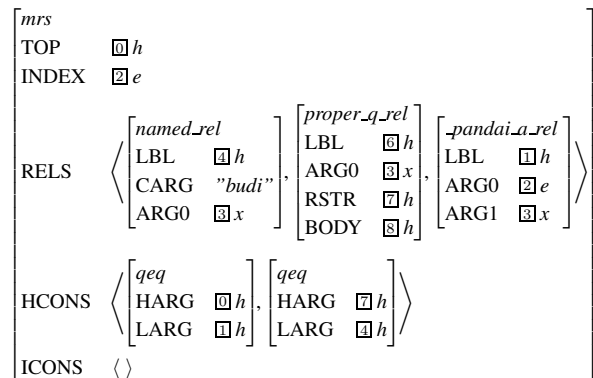


Figure 5: MRS representation of *Budi pandai* “Budi is clever”

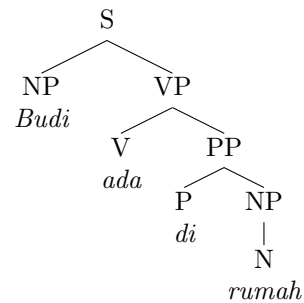


Figure 6: Parse tree of *Budi ada di rumah* “Budi is at home”

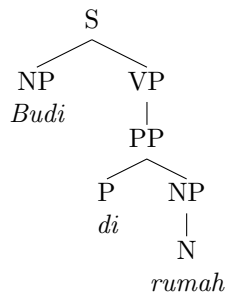


Figure 7: Parse tree of *Budi di rumah* “Budi is at home”

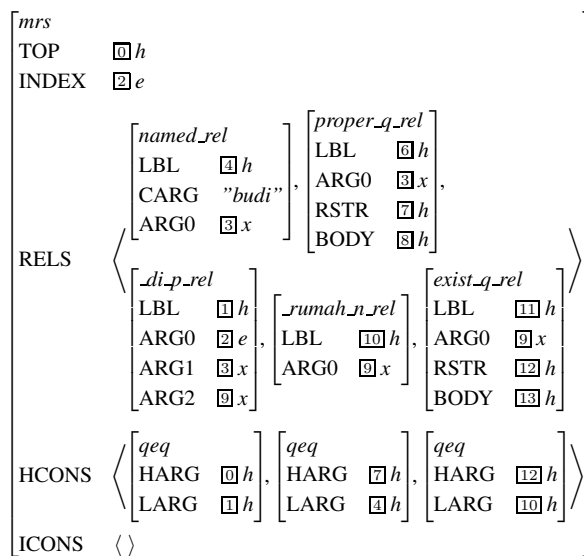


Figure 8: MRS representation of *Budi (ada) di rumah* “Budi is at home”

References

- Hasan Alwi, Soenjono Dardjowidjojo, Hans Lapoliwa, and Anton M. Moeliono. 2014. *Tata Bahasa Baku Bahasa Indonesia*. Balai Pustaka, Jakarta, 3 edition.
- I Wayan Arka. 2013. Nonverbal predicates in Austronesian and Papuan languages: an LFG perspective. pages 30–46, Unud, Bali.
- Emily M. Bender. 2001. *Syntactic Variation and Linguistic Competence: The Case of AAVE Copula Absence*. Ph.D. thesis, Stanford University.
- Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan A. Sag. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language and Computation*, 3(4):281–332.
- Ann Copestake. 2002. *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford.
- R. M. W. Dixon. 2009. *Basic Linguistic Theory 2: Grammatical Topics*. OUP Oxford, Oxford.
- Ronald Kaplan and Joan Bresnan. 1982. Lexical Functional Grammar: A formal system for grammatical representation. In *The Mental Representation of Grammatical Relations*, pages 173–281. the MIT Press, Cambridge.
- M. Paul Lewis. 2009. *Ethnologue: Languages of the World*. SIL International, Dallas, Texas, 16 edition.
- Malcolm W. Mintz. 2002. *An Indonesian and Malay Grammar for Students*. Indonesian / Malay Texts and Resources, Perth, 2 edition.
- David Moeljadi, Francis Bond, and Sanghoun Song. 2015. Building an HPSG-based Indonesian Resource Grammar (INDRA). In *Proceedings of the Grammar Engineering Across Frameworks (GEAF) Workshop, 53rd Annual Meeting of the ACL and 7th IJCNLP*, pages 9–16.
- Thomas E. Payne. 2008. *Describing Morphosyntax: A Guide for Field Linguists*. Cambridge University Press, New York.
- Ivan A. Sag, Thomas Wasow, and Emily M. Bender. 2003. *Syntactic Theory: A Formal Introduction*. CSLI Publications, Stanford, 2 edition.
- James Neil Sneddon, Alexander Adelaar, Dwi Noverini Djenar, and Michael C. Ewing. 2010. *Indonesian Reference Grammar*. Allen & Unwin, New South Wales, 2 edition.
- Liling Tan and Francis Bond. 2012. Building and annotating the linguistically diverse NTU-MC (NTU-multilingual corpus). 22(4):161–174.
- Frank Van Eynde. 2009. On the copula: from a fregean to a montagovian treatment. In Stefan Müller, editor, *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar, Georg-August-Universität Göttingen, Germany*, pages 359–375.