The Indonesian copula adalah: What it is (not) and why we do (not) need it

Dominik BESIER

Bergische Universität Wuppertal

This paper addresses the structural status, function, and predictability of the Indonesian copula *adalah*. It is analyzed as a non-verbal element that is realized in T. Nevertheless, *adalah* does not behave like a T-auxiliary as it is not sensitive to movement. Therefore, *adalah* is analyzed as a post-syntactic operation at the phonological form interface. The main functions of *adalah* are to help identify the break between the subject and predicate, and to act as an element for a smooth intonation contour. A full prediction of *adalah* remains impossible as its occurrence is influenced by factors such as register, communication setting, participating speakers, etc. Nevertheless, it is possible to identify environments where *adalah* is unlikely, namely when there is no need for prosodic smoothening or disambiguation, mostly due to overt strategies such as T-auxiliaries, particles, or topicalization/focalization.

1. Introduction

An unsolved puzzle in the Indonesian language is its copula *adalah*. Unlike the English copula *be*, which occurs with all kinds of non-verbal predicates, namely adjectives (1),[1] noun phrases (2), preposition phrases (PPs) (3), and clauses (4), Indonesian *adalah* is available for nominal (5) and clausal predicates (6), but generally ruled out with adjectives[2] (7) (if such a class exists) and prepositional phrases describing a location (cf. Moeljadi, Bond and Morgado da Costa 2016:444) (8).

(1) *John is hungry.*

(2) *John is a teacher.*

(3) *John is in London.*

(4) *The problem is that he had to leave early.*

(5) *Udin adalah   se-orang   guru.*
     Udin COP   one-CL.HUM   teacher
     'Udin is a teacher.'

(6) *Masalah-nya adalah bahwa dia harus pergi duluan.*
     problem-det   COP   that   3SG must   go   early
     'The problem is that he*she must leave early.'

(7) **Udin adalah lapar.*
      Udin   COP   hungry
      'Udin is hungry.'

(8) **Udin adalah di Jakarta.*
      Udin   COP   in Jakarta
      'Udin is in Jakarta.'

Additionally, *adalah* is usually optional (Sneddon 1996:237). Thus, sentences with (5) or

---

[1] The number in brackets are references to the examples provided.

[2] In some cases, *adalah* can occur with adjective phrases (APs) (Sneddon 1996:238).

without *adalah* are often grammatical (9).

(9)   *Udin   se-orang   guru.*
       Udin   one-CL.HUM   teacher

Thus, the challenge of the *adalah* puzzle is twofold. On the one hand, it remains unclear what kind of syntactic status *adalah* has. On the other hand, the optionality of *adalah* raises the question of the possibility of predicting the occurrence of *adalah*. Both of these issues will be addressed in this paper.

2. *Adalah* – a borrowed element?

Before focusing on these substantial questions, it is necessary to address the possibility that *adalah* may be a borrowed element. Since *adalah* is more common in – if not exclusive to – formal standard Indonesian, there is the possibility that *adalah* may be an artificial element that has been pushed into Indonesian in the language planning process, but may not actually form part of Indonesian grammar. Although *adalah* is most likely a rather recent invention of the Indonesian/Malay language, it is not necessarily an artificial implementation, which is something that will be discussed in the following section.

According to available copulas, in her studies on copular structures, Pustet (2003) categorized 20 Austronesian languages, among other languages. While not an exhaustive list, among these Austronesian languages, Indonesian was the only language exploiting a copula. Thus, copulas generally do not feature in Austronesian languages. Another hint that *adalah* is a relatively recent development in Indonesian comes from its morphology; *adalah* can be split into two productive morphemes: *ada* (existential verb) and *-lah* (emphasis/focus marker). Nevertheless, the definition of the copula *adalah* does not combine these productive morphemes, but is semantically opaque and, thus, has to be treated as one morpheme (Hopper 1972:130). Thus, this illustrates clearly that the copula *adalah* is a more recent innovation in Indonesian.

Although the emergence of the copula *adalah* probably coincides with intensive Western influence (especially Dutch and English) and language planning, it should not be seen as a simple implementation of a foreign idea in to the Indonesian language. There are at least two indications to argue in this direction. Although its development has been conditioned by copulas in Dutch or English (Hopper 1972:130), *adalah* cannot be seen as simply borrowing from these languages, as it lacks the essential features of these copulas, in other words, the status of being a verb (Hopper 1972:130). Even in its distributional behavior, *adalah* differs from the English and Dutch copula (see examples 1–8), as it usually only occurs with nominal and clausal phrases. If it is simply borrowing from these languages, we should also find *adalah* with adjectival predicates, as it has been shown for Malay in Malaysia (Tadmor 2007:321). *Adalah* in Indonesian, however, is different and, therefore, unique in its behavior. Other than this distributional evidence, there is even a potential grammaticalization path from productive *ada+lah* to the copula *adalah* via a forward-shifted topic (Hopper 1972:131), as imitated in (10).

(10)   *Alexander – There was a great general* (Hopper 1972:131)

With the general change from the verb initial word order to subject–verb–object (Cumming 1991), the shifted topic becomes the subject, and *adalah* grammaticalizes into a copula. Therefore, the Indonesian copula *adalah* has 'naturally' made its way into the Indonesian language, or as natural as possible for a semiartificial language.

The last aspect for consideration is that Indonesian, close to a century after its birth, is increasingly losing its semiartificial status. With a growing number of Indonesian native

speakers (generally in a colloquial Indonesian variety), standard Indonesian increasingly changes to a formal register of the Indonesian language. This standard variety does not necessarily reach the standard proposed by prescriptive grammarians; however, it is a language used in formal situations. This formal register also requires an internalized grammar that includes *adalah*. If *adalah* was only needed by prescriptive grammar, there should be a better chance to establish clear rules for its usage. As this is not the case, *adalah* should be treated as part of the tacit knowledge of grammar (internalized grammar). Thus, even if *adalah* was, at some point, intentionally planted into the standardized Indonesian language, it would no longer be an artificial element, but it would become an integrated part of the internalized Indonesian grammar especially in, but not exclusively to, the formal variety.

In conclusion, despite being a recent innovation influenced by foreign languages and language planning, *adalah* is an important and unique part of Indonesian grammar that is worthy of research. We can now address the questions: what is *adalah*, and when it is required?

3. What is the structural status of adalah?

In generative syntax, a clause normally subsumes three essential layers: the V/v layer, the T layer, and the C layer. All of these layers can and have been split into several heads (see Rizzi (1997) for C, Harley (1995) for V, and Cinque (1999) for T). The V/v-layer deals with the argument structure and theta-role assignment normally done by the lexical verb. The T deals with tense and agreement, and the C-layer deals with force and information structure (topic, focus). In this study, only the V- and T-layers are of interest. While V is the place of the lexical verb, there are two competing approaches for the T-position. Guilfoyle, Hung & Travis (1992) proposed two different subject positions in the Indonesian language: the position for the agent inside of V, and the second subject in the specifier position of an inflection phrase. The T-head is filled with voice morphology, such as *meN-* or *di-*. However, Arka (2013a) argues that T is occupied by finite auxiliaries such as *sudah and akan*. To reconcile both approaches, voice-morphology is seen as the realization of a voice-head comparable to Kratzer (1996) (see also Marantz' (1997) little v, Harley's (1995) Event or Bowers' (1993) Pred) in a split VP fashion. Therefore, voice morphology is still part of the V/v-layer, and Arka's auxiliaries are then T-realizations.

The critical question needed to be addressed is if *adalah* is a syntactical part of the V/v-layer or the T-layer.

3.1 *Adalah* is in T not in V

The question of the appropriate layer for the realization of *adalah* can be reduced to the question of the verbality of *adalah*. Whereas Hopper (1972:130) and Musgrave (2001:223) argue that *adalah* shows no verbal characteristics, Arka (2013b:39) and Moeljadi, Bond and Morgado da Costa (2016:447) claim that *adalah* is verbal. Moeljadi, Bond, and Morgado da Costa (2016:447) analyze *adalah* as a transitive verb with a noun phrase (NP) complement, and as raising auxiliaries (Moeljadi, Bond & Morgado da Costa 2016:449–450) in rare occurrences with APs and PPs. Mustaffa (2018) proposes a position between *adalah* in Malay; he argues for the auxiliary-like character of *adalah*; however, he still assigns some verbal component to it. As a result, *adalah* is realized in the v-layer.

This paper will first look at the evidence for the (non-)verbality of *adalah*, before considering Mustaffa's (2018) approach of little v.

Syntactically, *adalah* behaves differently than verbal forms, in contrast to the copular verb *ada*. Whereas *ada* can be embedded below a finite T-auxiliary like *masih* (see Arka 2013a) (11), the combination of *adalah* and *masih*[3] is generally ungrammatical (12).[4] Interestingly, the impossibility of *adalah* to co-occur with tense, aspect, and mood (TAM) markers has also been identified by Moeljadi, Bond and Morgado da Costa (2016:445). They even contrast it to the copula *merupakan*, which can take such a TAM marker (13).

(11) *Adit   masih ada   di   Jakarta.*
      Adit   CONT be   in   Jakarta
      'Adit is still in Jakarta.'

(12) **Adit masih adalah   guru-nya   Udin.*
       Adit   CONT COP   teacher-3SG.POSS Udin

(13) *Ini sudah/akan   *adalah/merupakan hal   yang   luar   biasa.*
      this PERF/FUT   COP   case REL   beyond   ordinary
      'This has been/will be an extraordinary case.' (Moeljadi, Bond & Morgado da Costa, 2016:445)

Despite this different syntactical behavior, they analyze both forms as verbal due to their general interchangeability (Moeljadi, Bond & Morgado da Costa 2016:446). While *merupakan* is undoubtedly a verbal form, which even still bears overt verbal morphology *meN-* and *-kan* (Moeljadi, Bond & Morgado da Costa 2016:444), the mismatch in the behavior of *merupakan* and *adalah* points to a non-verbality of *adalah*.

In Malay, even the combination of *adalah* and *merupakan* is possible (Mustaffa 2018:38).

(14) *Islam adalah   merupakan   agama   yang   syumul.*
      Islam COP   form   religion COMP   holistic
      'Islam is a holistic religion.' (Mustaffa 2018:38)

Further evidence for the non-verbality of *adalah* comes from negation. While the verbal form *ada* can be negated by the negator *tidak* (can only negate verbs and adjectives)[5] (15), the negation of *adalah* with *tidak* is impossible (Musgrave 2001:223) (16).

---

[3] Although *masih* might not be the most prototypical T-auxiliary, it is the most common one with nominal predicates. Prototypical T-auxiliaries like *akan*, *sedang*, and *sudah* are normally not acceptable with nominal predicates even without *adalah*:

(89) **Adit   sudah   guru-nya   Udin.*
      Adit   ANT   teacher-POSS.3SG Udin

(90) **Adit   sedang gurunya   Udin.*
      Adit   PROG   teacher-POSS.3SG Udin

(91) **Adit   akan   gurunya   Udin.*
      Adit   FUT   teacher-POSS.3SG Udin

[4] As pointed out by an anonymous reviewer, the combination of *masih adalah* can occur both in spoken (92) and written discourse (93).

(92) *...sehingga yang   kita tahu ini masih adalah teori.* (https://biologi.unitri.ac.id)
        so.that   NOM   1PL know this CONT COP   theory
        '… so that what we know here is still a theory.'

(93) *…selama ia   masih   adalah warga mamuju* (http://mamujukab.go.id/page/24/?feed)
        as.long 3.sg CONT   COP   citizen Mamuju
        '…as long as he/she is still a citizen of Mamuju'

[5] If a distinction between verbs and adjectives exists in Indonesian.

(15) *Adit   tidak   ada   di   Jakarta.*
      Adit NEG   be   in   Jakarta
      'Adit is not in Jakarta.'

(16) **Adit   tidak   adalah   guru-nya   Udin.*
       Adit   NEG   COP   teacher-3SG.POSS Udin

Here again we find a behavioural mismatch of *merupakan* and *adalah* as well (17).

(17) *Budi   tidak   merupakan/*adalah   guru.*
      Budi NEG   COP   teacher
      'Budi is not a teacher.' (Moeljadi, Bond & Morgado da Costa 2016:452)

In general, nominal phrases/predicates cannot be negated by the negator *tidak*, but with the negator *bukan*. Nevertheless, *adalah* cannot be embedded into a *bukan*-phrase (18). On rare occasions, a *bukan*-phrase can be embedded below *adalah* (19).[6]

(18) **Adit   bukan   adalah   guru-nya   Udin.*
       Adit   NEG   COP   teacher-3SG.POSS Udin

(19) *Bahasa   Indonesia adalah   bukan bahasa   yang   pe-nampil   subyek*
      Language Indonesia COP   NEG   language REL   CAUS.NOM-show subject.
      'Indonesian is not a subject-prominent language.' (Sneddon 1996:238)

Although *bukan* is generally used to negate noun phrases, it must occur before the predicate (20).

(20) *Adit   bukan   guru-nya   Udin.*
      Adit NEG   teacher-3SG.POSS Udin
      'Adit is Udin's teacher.'

Placement in front of the complement (21) is impossible

(21) **Udin   makan   bukan   roti.*
       Udin   eat   NEG   bread

Thus, *bukan* precedes the predicate. Additionally, it can be embedded below T (22), but cannot embed T (23). Therefore, it should be treated as an auxiliary. Auxiliaries must be higher than verbs (but below T (21)). Since *adalah* cannot be embedded by *bukan* (18) but embeds *bukan* in rare cases (19), *adalah* must be higher than the auxiliary and consequently higher than a verb.

(22) *Adit   sudah bukan   guru-nya   Siska   lagi.*
      Adit ANT   NEG   teacher-3SG.POSS Siska   again
      'Adit is not Siska's teacher anymore.'

(23) **Adit bukan sudah guru-nya   Siska   lagi.*
       Adit NEG   ANT   teacher-3SG.POSS Siska   again

Accordingly, *adalah* cannot occur below any other auxiliary like *pernah* (once) (24).[7]

---

[6] An anonymous reviewer has stated doubts concerning the grammaticality of sentence (19). As this example is taken from Sneddon's (1996) reference grammar, I assume that this sentence is acceptable at least to some speakers.

[7] As an anonymous reviewer pointed out, in very rare cases a combination of *pernah* and *adalah* may occur (94).

(94) … *Fry pernah adalah aktor yang …*   (https://potensibisnis.pikiran-rakyat.com)
        Fry once COP   actor REL

(24) *Adit   pernah adalah guru-nya   Siska
     Adit   once   COP   teacher-3SG.POSS   Siska

Finally, it is impossible to embed *adalah* below control verbs like *ingin* (want) (25, 26).

(25) *Adit   ingin   adalah se-orang   dokter
     Adit   want   COP   one-CL.HUM   doctor

(26) Adit   ingin   ada di Jakarta.
     Adit   want   be   in Jakarta
     'Adit wants to be in Jakarta.'

In conclusion, based on this evidence, *adalah* behaves, at no point, verbally. Therefore, the realization in the V/v-layer is unlikely.

Although Mustaffa (2018:28) argues against the verb analysis and for an auxiliary-like character of *adalah* in Malay, he still argues for a verbal component in v that is moved further to T. The main argument of this analysis is that these predicates require a verbal layer as they can be modified by verb phrase (VP)-modifiers such as (temporal) adverbials (Mustaffa 2018:35–37).

(27) Dia seorang   guru   pada suatu ketika dahulu.
     3SG one.CL.HUM   teacher at   one   time   before.
     'She was a teacher once upon a time.' (Mustaffa 2018:36)

Although I agree with Mustaffa's analysis of having a verbal layer in copula structures, I see no necessity to have *adalah* as this v-element. If *adalah* would be a little v, then one might want to argue against the v-layer in sentences which do not contain *adalah* as it is normally optional. However, the evidence presented for the v-layer is actually in cases without overt *adalah*. Hence, the v-layer is available but must remain zero. If the v-layer remains zero, one cannot assume that *adalah* occupies this position with the occurrence of a v-layer.

With so much evidence that *adalah* is non-verbal, I consequently see no necessity for *adalah* to be in the v-layer but assume that this layer remains null. I still, like Mustaffa (2018), expect a v-layer even for nominal predicates.

Having argued against the v-layer as the structural position for *adalah*, we can now explore the possibility of T.

As already illustrated, embedding *adalah* below a T-auxiliary-like *masih* is generally impossible. This leads to the possibility that *adalah* is in complementary distribution with T-auxiliaries, as they cannot co-occur.

(28) Adit   masih guru-nya   Siska.
     Adit   CONT   teacher-3SG.POSS   Siska
     'Adit is still Siska's teacher.'

(29) Adit adalah   guru-nya   Siska.
     Adit COP   teacher-3SG.POSS   Siska
     'Adit is Siska's teacher.'

(30) *Adit   adalah   masih guru-nya   Siska
     Adit   COP   CONT   teacher-3SG.POSS   Siska

'Fry once has been an actor that …'

Although occurring, an informant doubts the grammaticality of this sentence due to *pernah adalah*.

(31) *Adit   masih adalah guru-nya   Siska
     Adit   CONT COP   teacher-3SG.POSS   Siska

Consequently, if *adalah* and *masih* (which is a T-auxiliary) are in complementary distribution, they should structurally occupy the same position in the tree, namely, in T.

Besides this distributional evidence, another factor points to the T-analysis. Arka (2013b:34) shows that *adalah*-less [8] (32) and *adalah*-constructions (33) with NP predicates are non-future.

(32) *Dia   akan dokter
     3SG   FUT   doctor

(33) *Dia   akan adalah dokter.
     3SG   FUT   COP   doctor

Thus, he contrasts *adalah* with the inchoative form *menjadi*, which leads to the opposition of the present stative and future inchoative. Nevertheless, stative copulas are not wholly non-future in Indonesian, as *merupakan* can be embedded below *akan*.

(34) *Ini akan   merupakan hal yang luar   biasa.*
     this FUT   COP   case REL   beyond ordinary
     'This will be an extraordinary case.' (Moeljadi, Bond & Morgado da Costa 2016:445)

Thus, the information on non-future must be realized in null-T (30). With *adalah* being incompatible with *akan* (32) or any T-element in general, it is plausible that the optionality of *adalah* is in T and, as such, *adalah* itself (if realized) is the bearer of the non-future tense information.

Based on these data, it is plausible that *adalah* is realized higher than the V/v-layer. The candidate proposed here is T. Nevertheless, the exact position of the generation of *adalah* might ultimately be less relevant, and should be seen as a post-syntactic element.

3.2 *Adalah* is post-syntactic

Although there is evidence for *adalah* to be realized in T, it also does not behave like a T-auxiliary, such as *masih*. Despite the possibility of being in complementary distribution with the T-auxiliaries, it is still different in its syntactic behavior since it does not allow further movement in focalization. In Indonesian, it is possible to focus on the predicate by fronting it. This focalization goes along with a change in intonation (35) or an overt *-lah* (36). With an overt *-lah*, T-auxiliaries can even be focalized on their own (37). All of these structures should involve movement into the C-layer. Whether this is head movement to C or remnant movement into a focus phrase is not relevant here.

(35) *Masih pacar-nya   Siska / Adit.*
     CONT   boyfriend-3SG.POSS Siska Adit
     'It is still the case that Adit is Siska's boyfriend.'

(36) *Masih pacar-nya   Siska-lah Adit*
     CONT   boyfriend-3SG.POSS Siska-FO Adit
     'It is still the case that Adit is Siska's boyfriend.'

---

[8] Arka (2013b) calls those clauses verbless.

(37) *Masih-lah Adit pacar-nya   Siska.*
     CONT-FOC Adit boyfriend-3SG.POSS Siska
     'It is still the case that Adit is Siska's boyfriend.'

None of these inversions, which result from movement operations, are possible with *adalah*. *Adalah* allows focalization with *-lah*, neither as a single item (38) nor part of the predicate (39). Focalization without overt *-lah* is also ungrammatical (40). The same observation can be made with the *-kah* forms required for questions (41–42).[9] No fronting was observed.

(38) **Adalah-lah Adit   pacar-nya   Siska.*
       COP-FOC   Adit   boyfriend-3SG.POSS Siska

(39) **Adalah pacar-nya   Siska-lah   Adit.*
       COP   boyfriend-3SG.POSS Siska-FOC   Adit

(40) **Adalah   pacar-nya   Siska / Adit.*
       COP   boyfriend-3SG.POSS Siska   Adit

(41) **Adalahkah   Adit   pacar-nya   Siska?*
       COP-QUEST   Adit   boyfriend-3SG.POSS Siska

(42) **Adalah   pacar-nya   Siska-kah   Adit?*
       COP   boyfriend-3SG.POSS Siska-QUEST Adit

Thus, *adalah* is not sensitive to movement; it is therefore, interpreted as a post-syntactic phenomenon. As such, it does not have to be sensitive to movement. The idea of a post- or non-syntactic behavior of *adalah* is strengthened by the fact that *adalah* does not contribute to the meaning of a sentence. The following examples have identical meanings with (43) or without (44) overt *adalah*.

(43) *Adit adalah gurunya Udin.*
     'Adit is Udin's teacher.'

(44) *Adit gurunya Udin.*
     'Adit is Udin's teacher.'

As there is no difference in information structure, the function or necessity of *adalah* can only be at the Phonological Form (PF) interface.

4. What is the function of *adalah*?

With its post-syntactic character, *adalah* cannot be a syntactic requirement. Even in the approaches of Moeljadi, Bond and Morgado da Costa (2016) and Mustaffa (2018), which place *adalah* in the V/v-layer, the occurrence of *adalah* is not a syntactic requirement for a V-/v-layer because this layer is also assumed to be in *adalah*-less 'copula structures'. *Adalah* does not contribute to the meaning but is "semantically vacuous" (Mustaffa 2018:29). However, what then is the function of *adalah*? The function of *adalah* is two-fold. First, it avoids obscurity concerning the break of the subject and predicate (Hopper

---

[9] A similar inversion is possible in Malay (Mustaffa 2018: 29):

   (95) *Adakah Ali se-orang   guru?*
        COP   Ali one-CL.HUM teacher
        'Is Ali a teacher?' (Mustaffa 2018: 29)

However, here the *-lah*-suffix is replaced by the question suffix *-kah*. Thus, the main empirical evidence brought for the little-v analysis in Malay is not available in Indonesian.

1972:130, Musgrave 2001:223). Second, it is "stylistic" (Hopper 1972:131) to "add a smoothness" (Sneddon 1996:237) to the construction. This paper will now closely examine these two functions.

4.1 *Adalah* avoids ambiguity

The following sentence is ambiguous.

(45) *Guru-nya   Lila Mina   Surawati.*
      teacher-3SG.POSS Lila Mina   Surawati

It is unclear whether the teacher is Lila Mina Surawati (46), Lila's teacher is Mina Surawati (47), or if Lila Mina's teacher is Surawati (48).

(46) *Gurunya / Lila Mina Surawati.*

(47) *Gurunya Lila / Mina Surawati.*

(48) *Gurunya Lila Mina / Surawati.*

In spoken register, the subject and predicate could be identified as two different intonation phrases with both a pitch event (raising tone) and a pause between them. Written language lacks intonational cues. Thus, *adalah* is a strategy to avoid ambiguity. The sentences (49–51) are no longer ambiguous, even in written form.

(49) *Guru-nya   adalah Lila Mina Surawati.*
      teacher-3SG.POSS COP   Lila Mina Surawati
      'Her teacher is Lila Mina Surawati.'

(50) *Gurunya   Lila adalah Mina Surawati.*
      teacher-3SG.POSS Lila COP   Mina Surawati
      'Lila's teacher is Mina Surawati.'

(51) *Gurunya   Lila Mina adalah Surawati.*
      teacher-3SG.POSS Lila Mina COP   Surawati
      'Lila Mina's teacher is Surawati.'

Therefore, written language is most likely responsible for the development of *adalah* (Hopper 1972:130).

Thus, one function of *adalah* is to "make sure that both the subject and predicate are easily identifiable" (see Hopper 1972:130). As such, *adalah* is an "anti-ambiguity device" (Eid 1983:197) to mark the division of subject and predicate (Sneddon et al. 2010:247).

As Sneddon (1996:237) highlights, *adalah* typically occurs between two long nominal phrases. This observation can be explained by the concept of an anti-ambiguity device in combination with an intonation structure. In sentences (46) to (48), intonation is sufficient for disambiguation. Here, *adalah* is unlikely in the spoken register. However, with two long nominal phrases, intonation alone cannot disambiguate sufficiently. Therefore, *adalah* can also function as an anti-ambiguity device in spoken register. Therefore, a closer examination of the intonation structure is necessary.

Before addressing the *adalah* case, some general notes on Indonesian intonation structures are required. I adopt Liberman's (1975) and Pierrehumbert's (1980) ideas that prosodic units are marked by tones or pitch accents. The prosodic units relevant to this study are intonational and intermediate phrases. The main difference is that after an intonational phrase, a pause can be added (Nespor and Vogel 1986:188). In a standard

Indonesian sentence like (52)[10] or (53), intermediate phrases end on a high phrasal tone; for example, at the end of *meninggal* and *lalu* in (52) and the end of *menari* and *Bandung* in (53).
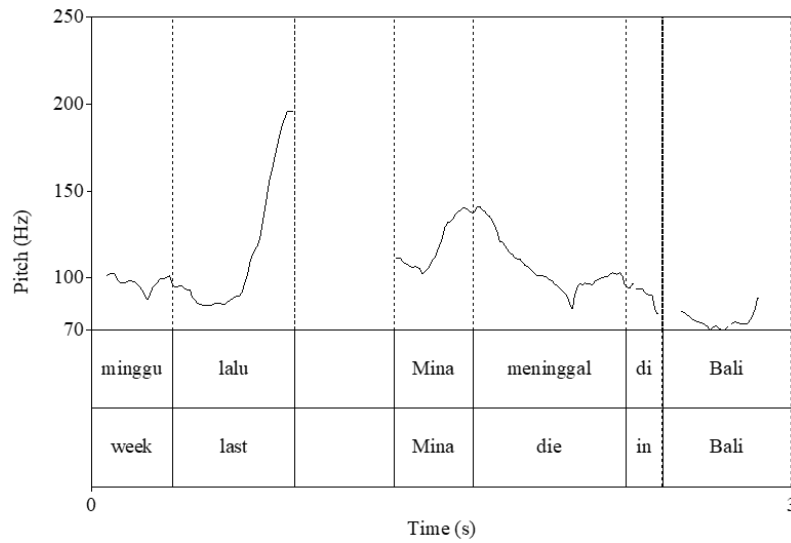
(52)
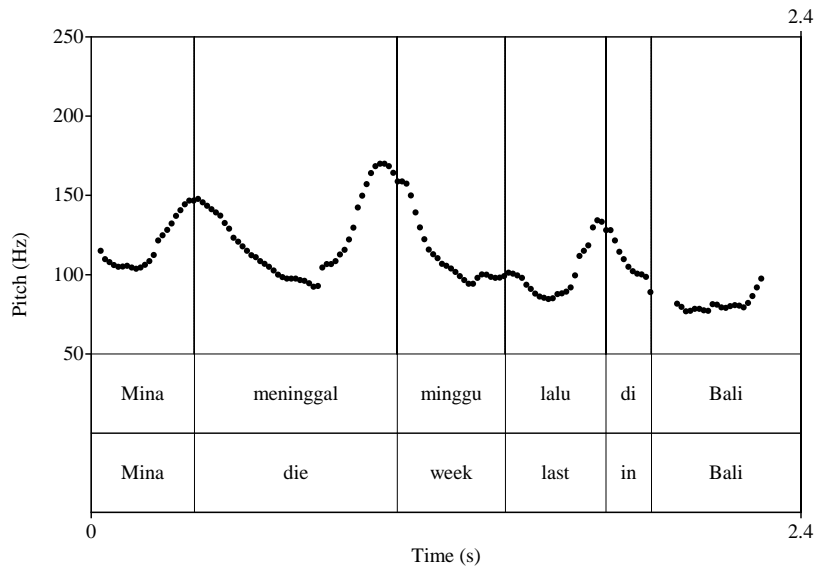
(53)
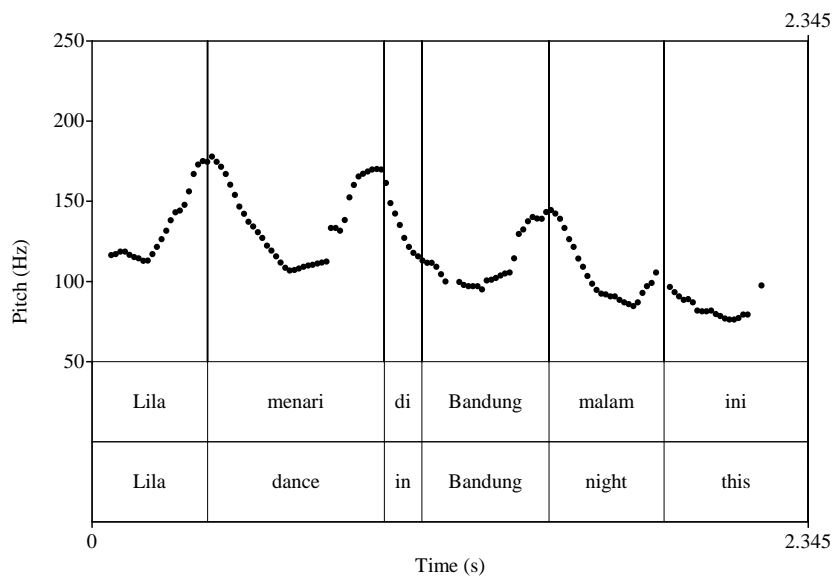
Intonation phrases also have a high boundary tone (after *lalu*). An additional pause is possible in (54).

---

[10] The examples (52), (53) and (54) are from a male Indonesian speaker in his early thirties. All these items were recorded purposefully and are not part of spontaneous speech.

(54)

With this prerequisite, a copular clause with a nominal predicate can now be examined (55).

(55) *Mina Mega Budiman (adalah) ibunya Muhammad Mauliman*

Since phrasal tones at the boundaries of intermediate phrases (*Mina* and *ibunya*), and boundary tones at the boundaries of intonational phrases (*Budiman*) are high boundary tones (56),[11] the boundary tone between the subject and predicate is consequently less prominent.

(56)
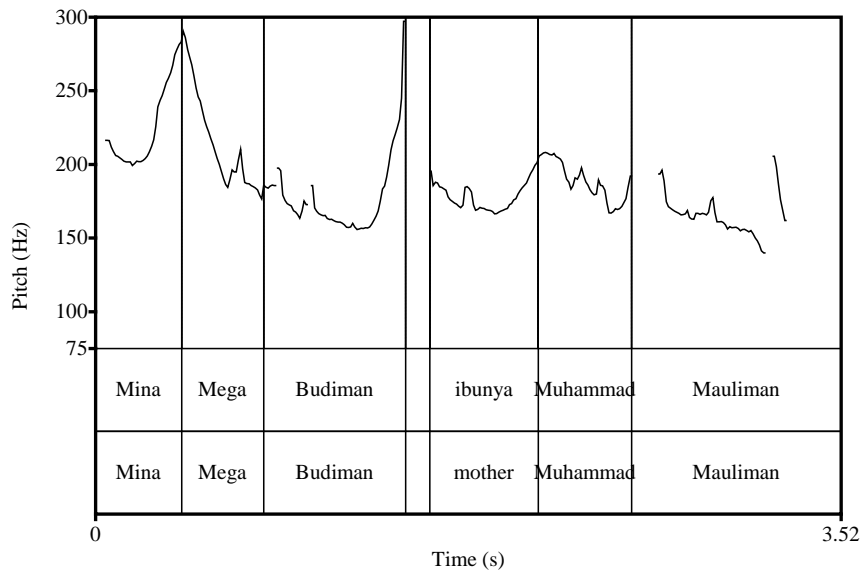
Since prosodically *adalah* builds its own intermediate phrase (56), in such a case, it can provide an additional cue and thus, avoid ambiguity.

---

[11] Example (56) and (57) are from a female Indonesian speaker in her early thirties.

(57)

The pitch track shows the following aligned tiers:

Mina	Mega	Budiman	adalah	ibunya	Muhammad	Mauliman
Mina	Mega	Budiman	be	mother	Muhammad	Mauliman
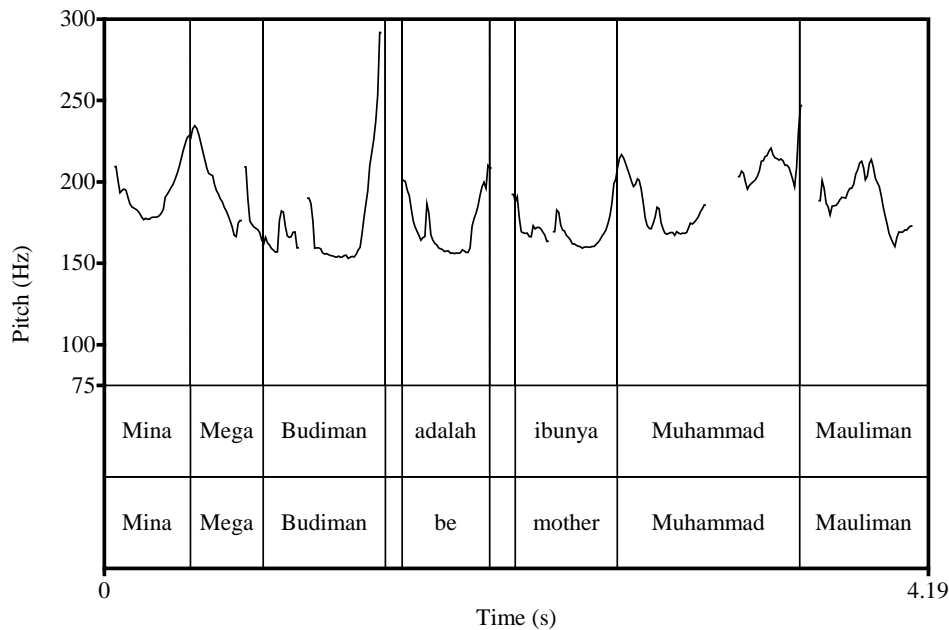
4.2 *Adalah* as 'stylistic' element

However, *adalah* should not be reduced to a function to avoid ambiguity. Sentences (58) and (59) are at no point ambiguous, and the identification of the subject and predicate is easily possible. Nevertheless, omitting *adalah* has not been a favored option, even in formal spoken register. Thus, *adalah*-insertion is preferred in (58) and (59).

(58)   *Susilo   Bambang   Yudhoyono ?(adalah)   jenderal.*
       Susilo   Bambang   Yudhoyono COP   general.
       'Susilo Bambang Yudhoyono was a general.'

(59)   *Sukarno ?(adalah)   presiden Indonesia   yang   pertama.*
       Sukarno COP   president Indonesia REL   first.
       'Sukarno was the first president of Indonesia.'

Both (58) and (59) have a strong asymmetry in length between the subject and predicate. One explanation would be that a pause after *Yudhoyono* (58) or *Sukarno* (59) would disrupt the flow and produce an odd break. Since *adalah* builds its own intermediate phrase (57), it can smoothen the intonation without requiring a break. Hence, in (58) and (59), *adalah* has a pure stylistic function. Additionally, *adalah* even allows for a break before or after it and hence, provides new possibilities for prosodic cues.

Therefore, the function of *adalah* is two-fold. First, it avoids ambiguity and provides an additional cue to distinguish between the subject and predicate. Second, it smoothens intonation. Thus, *adalah* is a post-syntactic operation at PF.

5. *Adalah* and when it is (not) needed

Even as a PF operation, the question of when *adalah* is necessary remains. In general, *adalah* is reported as optional. However, is there any possibility of predicting the occurrence of *adalah*? This paper turns this perspective and discusses environments where *adalah* is unlikely, if not impossible. The goal is to predict the non-usage of *adalah*.

5.1 Overt T

As we have seen before, *adalah* is in complementary distribution with overt Ts (60).

(60) **Udin masih adalah guru-nya Siska.*
     Udin CONT COP teacher-3SG.POSS Siska

Therefore, *adalah* insertion normally cannot co-occur with an overt T-element. This distributional behavior has two reasons.

First, *adalah* insertion occurs in T; thus, there is a competition for the same position. Second, in linear order, the subject is left to T and the predicate on its right. Therefore, there is no need for disambiguation because the T element perfectly marks the subject and predicate. Hence, *adalah* is ungrammatical if there is an overt T. It does not matter if the auxiliary is base generated in T like *masih* (60) or moves to T like *sering* (61). Whenever the T-head ends up being filled with an overt element, *the adalah* insertion is blocked due to competition. If we turn it around, no overt T is a prerequisite for *adalah* insertion.[12]

(61) **Udin sering adalah guru-nya Siska*
     Udin often COP teacher-3SG.POSS Siska

Adverbs such as *juga* (62) that can occur between subject and predicate diminish the possibility of *adalah* insertion; however, they do not rule out insertion completely.

(62) *Udin juga #adalah guru-nya Siska.*
     Udin also COP teacher-3SG.POSS Siska
     'Udin is also Siska's teacher.'

In (62), *juga* does not occupy the T-position. Therefore, competition was not given, and *adalah* insertion remained possible. However, since *juga* also takes over the function of disambiguation and prosodic smoothening (additional intermediate phrase), *adalah* is very unlikely.[13]

5.2 Marking the edge of subject or predicate

Another strategy that makes *adalah* superfluous is overt topicalization or focalization. In both cases, either the subject or predicate is given a pragmatic role. In potential *adalah*-cases, this highlights either the subject or predicate, and the boundary between them cannot remain obscure.

Topicalization typically involves the subject. In sentence (63), *Joko Widodo* is a topic marked with the topic marker *itu*.[14] This overt topic marker *itu* marks the right edge of the subject and provides a sufficient cue to avoid any obscurity of this break.

---

[12] In rare cases the combination of T (e.g., *masih*) and *adalah* and even rarer cases of an auxiliary (e.g., *pernah*) and *adalah* is possible. Since V/v is null in these cases, for some speakers it should then be acceptable to perform *adalah*-insertion in the V-layer. In general, *adalah*-insertion takes place at T.

[13] An anonymous reviewer noted the possibility of '*Udin adalah juga gurunya Siska*'. Since *juga* is an adverb, which is not in the T position, it can either precede or follow *adalah*. Therefore, the sentence is possible. Depending on the situation and the speaker *adalah* can be used, however should be generally less preferred than the *adalah*-less version.

[14] Although *itu* is generally a demonstrative (Sneddon 1996: 130), Kaswanti Purwo (1984:65) identifies also a topic marker function for *itu*. As a demonstrative reading might be unexpected with a proper noun, the topic marking is here the most likely analysis.

(63) *Joko Widodo   itu   #adalah   presiden   Indonesia.*
      Joko Widodo TOP COP   president Indonesia
      'Concerning Joko Widodo, he is Indonesia's president.'

In contrast, focalization normally highlights the predicate (at least in the examples relevant for *adalah* insertion). Focalization can be done with either overt *-lah* (64) or with a specific pitch pattern (65), in contrast to the non-focalized form (66) (Halim 1975).

(64) *Makan   nasi-lah   Adit.*
      eat   rice-FOC Adit
      'Eating rice is what Adit does.'

(65) 232f / 211f[15]
      Makan   nasi /   Adit
      eat   rice   Adit.

(66) 233r / 231f
      Adit / makan   nasi.
      Adit   eat   rice

In both cases, *-lah* focalization (68 in contrast to 67) and intonation-focalization (70 in contrast to 69), *adalah* insertion is impossible.

(67) *Presiden Indonesia-lah   Jokowi*
      president Indonesia-FOC Jokowi
      'It is Indonesia's president who Jokowi is.'

(68) **Presiden   Indonesia-lah   adalah   Jokowi.*
       president   Indonesia-FOC COP   Jokowi

(69) *Presiden   Indonesia Jokowi.*
      president Indonesia Jokowi.
      'It is Indonesia's president who Jokowi is.'

(70)(*)*Presiden   Indonesia   adalah   Jokowi*[16].
        president   Indonesia   COP   Jokowi

Although the structure is inverted, the focalized predicate stands before the subject, and the boundary between these two is marked, either overtly with the suffix *-lah* or via prosody with a distinct intonation pattern. Hence, *adalah* insertion is impossible.

5.3 Particles between pause-groups

The third strategy that is not compatible with *adalah* insertion is the occurrence of a particle between the subject and predicate. Here, we find two strategies: an honorific or an emotive particle, usually limited to non-standard registers.

In spoken form, it is polite to insert an honorific in an utterance. Thus, (71) is considered more polite than (72).

---

[15] Halim (1975) describes the pitch contour based on three intonation levels 1, 2 and 3, with 1 being the lowest and 3 the highest pitch level. Additionally, f (falling) and r (rising) describe a pitch tone at the end of the phrase. Thus, 232f means that the intonation starts on the second pitch level, rises to the third and falls to the second with a falling tone at the end of the phrase.

[16] Although the sentence *Presiden Indonesia adalah Jokowi* is grammatical in general, it is impossible with a focalization intonation of 232f/ 211f.

(71) *Selamat pagi, bu.*
    'Good morning, ma'am.'

(72) *Selamat pagi.*
    'Good morning.'

This honorific is expressed by a shortened form of an appropriate kinship term, for example, *(ba)pak* (father), *(i)bu* (mother) or *(ka)kak* (older sibling), or the shortened form of the name, for example, *Sis(ka), (U)din,* or *(Ye)ris(*ca*)*. These honorifics can only be placed at certain positions, namely, at the end of a clause (73) or following the topic. In (74), the subject is the topic, so honorific insertion is possible. In (75), the adverbial *kemarin* (yesterday) is the topic. Hence, the honorific can be inserted after *kemarin* but not after the subject. These positions correlate with the end of the intonation phrase. These intonation phrases are pause-groups. Therefore, honorifics can only be inserted at the end of the pause-group.

(73) *Saya mem-beli buku itu, bu.*
    1SG   AG.VOIC-buy   book   DET   ma'am
    'I bought the book, ma'am.'

(74) *Saya, bu, mem-beli buku itu.*
    1SG   ma'am   AG.VOIC-buy   book   DET
    'I bought the book, ma'am.'

(75) *Kemarin, bu, saya, *bu, mem-beli buku itu.*
    Yesterday   ma'am   1SG   ma'am   AG.VOIC-buy   book   DET
    'Yesterday, I bought the book, ma'am.'

If the honorific occurs after the topic (the first pause-group) and the subject is also the topic (76, 78), *adalah* insertion is less preferable (77, 79).

(76) *Jokowi, bu, presiden Indonesia.*
    Jokowi   ma'am   president   Indonesia.
    'Jokowi, ma'am, is the president of Indonesia.'

(77) ?*Jokowi, bu, adalah presiden Indonesia.*
    Jokowi   ma'am   COP   president   Indonesia.
    'Jokowi, ma'am, is the president of Indonesia.'

(78) *Ridwan bang, pacar-nya Dewi.*
    Ridwan   bro   boyfriend-3SG.POSS   Dewi
    'Ridwan, bro, is Dewi's boyfriend.'

(79) ?*Ridwan, bang, adalah pacar-nya Dewi.*
    Ridwan   bro   COP   boyfriend-3SG.POSS   Dewi
    'Ridwan, bro, is Dewi's boyfriend.'

Once again, there is an overt way to mark the boundary between the subject (which is also the topic) and predicate. Therefore, *adalah* is not required for disambiguation. Additionally, the honorific also smoothens intonation and provides the possibility of an additional pause as a phonological cue.

While honorifics occur in formal and non-formal registers, emotive particles compete with *adalah*, and is usually undetectable in the formal register. Prescriptive grammar banned these particles from standard Indonesian. Similar to honorifics, these emotive particles cannot occur randomly; they are limited to the end of a pause-group, after either topic (80), focus (81), or clause-final (82). Other positions were ungrammatical (83, 84).

(80) *Kakak Udin kok udah masak.*
kakak Udin EMPAR ANT cook

(81) *Udah masak kok kakak Udin.*
ANT cook EMPAR kakak Udin

(82) *Kakak Udin (s)udah masak kok*
Kakak[17] Udin ANT cook EMPAR
'Kakak Udin has already cooked (unlike you have thought).'

(83) **Kakak kok Udin udah masak.*
Kakak EMPAR Udin ANT cook

(84) **Kakak Udin udah kok masak.*
Kakak Udin ANT EMPAR cook

When the emotive particles occur after the subject (also the topic), *adalah* insertion is impossible.

(85) *Ridwan kok pacar-nya Dewi.*
Ridwan EMPAR boyfriend-3SG.POSS Dewi.
'Ridwan is Dewi's boyfriend.'

(86) **Ridwan kok adalah pacar-nya Dewi.*
Ridwan EMPAR COP boyfriend-3SG.POSS Dewi.

(87) *Jokowi mah presiden Indonesia.*
Jokowi EMPAR president Indonesia.
'Jokowi is the Indonesian president.'

(88) **Jokowi mah adalah presiden Indonesia.*
Jokowi EMPAR COP president Indonesia.

As illustrated, ambiguity and obscurity are avoided by emotive particles. Therefore, *adalah* insertion is ruled out.

In general, there is a register mismatch between emotive particles available only in informal discourse and *adalah* available more in formal discourse. However, as it is unlikely to prescribe informal varieties, *adalah* insertion should be possible (even in the informal variety) if it would add value to the utterance. Since both disambiguation and smoothening are taken over by the emotive particle, *adalah* insertion is not necessary and thus, does not occur.

5.4 Structural vs. functional competition

In the last three subsections, five different environments were presented where *adalah* insertion was not likely or even impossible. However, with overt T, focalization, and emotive particles, *adalah* insertion is impossible; in the case of honorifics and topicalization, it is unlikely. Thus, how can we distinguish between the cases that are ruled out and those where it is less preferred?

As noted in the previous section, *adalah* insertion occurs most likely in T. Hence, it is in complementary distribution with T-auxiliaries. Such structural competition should completely rule out *adalah* insertion. The focalization case follows a similar analysis. Due to movement operations, *adalah* (if available) would not end up in T in focalization

---

[17] The term refers to an older sibling or to a person slightly older than the speaker. It is a common way to address people in a similar age like older sibling. Therefore, the term is not translated here.

cases, but somewhere in the C-layer. Since it is generally unavailable for movement operations, *adalah* insertion is impossible. Hence, we can conclude that *adalah* insertion is impossible when there is a structural mismatch (overt T, focalization involving movement).

In all other cases, there is 'only' functional competition. There is an overt strategy (particle, topic marker) that can either mark the boundary between the subject and predicate and/or smoothen the intonation contour. In these cases, *adalah* is unlikely, but cannot be ruled out. Emotive particles are exceptions. Here, the general register mismatch makes co-occurrence very unlikely. Thus, co-occurrence is generally to be seen as impossible. Thus, we can conclude that *adalah* insertion is impossible if it cannot be inserted in T, and it is unlikely if it is in functional competition with another overt strategy.

In general, it can be seen that whenever there is another strategy that sufficiently avoids ambiguity or obscurity, or smoothens the intonation contour, *adalah* is at least unlikely. Therefore, *adalah* has to be considered a last resort. The default case is without *adalah*. However, if there is a chance of ambiguity due to, for example, long nominal phrases or the need for smoothening of the intonation contour, and no other strategy is available, then *adalah* is inserted. With the function of providing an additional cue for prosodic mapping and making the intonation flow, *adalah* is a PF strategy at the prosody interface.

6. Why it is impossible to predict *adalah*?

Is it possible to predict the usage of *adalah*? Most likely, not completely. Two factors make predictions close to impossible. Linguistic factors, namely the length of the constituents and asymmetry in the length of constituents, are relative and not absolute. For the post-syntactic end weight principle in languages like English (Behagel 1909), Wasow (1997) has shown that relative length is a better predictor for the end weight principle than absolute criteria, such as the occurrence of relative clauses or PPs inside the NP. Similar effects should be applied to *adalah*. Thus, there is a good chance to predict the (non-)usage of *adalah* in prototypical examples with either very short or very long phrases. However, it is difficult to find a clear boundary between *adalah* cases and cases without *adalah* in the critical zone in between. This unpredictability is also true for utterance external characteristics. The two most relevant factors are register and situation.

Adalah is much more common in formal registers and, for some speakers, even unacceptable in informal registers. Two factors favored this evolution. First, informal registers are more often used in spoken language than in formal registers. Thus, lacking intonational cues, written registers have a higher need for compensation. One answer is the *adalah* strategy of disambiguation.

Second, strategies such as emotive particles have been banned from formal language. Students must learn that these particles are non-standard and, as such, unacceptable for standard Indonesian. Banning these strategies once again raises the pressure for disambiguation devices such as *adalah*. However, informal registers still have these strategies (intonation, focalization, emotive particles, etc.) in place and do not need *adalah*. These strategies are strong enough that standard Indonesian does not influence colloquial Indonesian in this matter. Although we tend to have more *adalahs* in formal and written register, this observation does not necessarily have to be a pure register phenomenon, as it could be related to the context of the utterance.

One way to avoid the register question is to apply Lindblom's (1990) H&H theory of hyperspeech and hypospeech. Hyperspeech is a well-articulated language normally required when the situation makes communication more complex; for example, noise, or

one participant has only limited proficiency in the language. Hypospeech, in contrast, is a more "floppily" articulated speech used in situations where communication is not impeded. In this sense, the formality of the situation could influence the choice of hyperspeech or hypospeech. This approach is applicable to *adalah*. *Adalah,* as a disambiguate device, is part of hyperspeech, thus, in any context where either communication is impaired (talking to foreigners, noise, etc.) or ambiguity has to be avoided, by all means, for example, at court, *adalah* as part of hyperspeech is chosen more often. In a hypospeech context, *adalah* is more likely to be omitted. The choice of hyperspeech or hypospeech depends on the situation, context, speakers involved, and speaker's style. In the same sense, as some people speak more clearly in general, and other speakers tend to mumble, the individual speaking style can influence *adalah* insertion. Since these factors in general, and the last factor in particular, cannot be controlled, it is not possible to predict the occurrence of *adalah* accurately.

7. Conclusion

Adalah is not verbal but complementary to T-auxiliaries. Therefore, *adalah*-insertion is realized in T. Nevertheless, *adalah* insertion is a post-syntactic operation required by the prosody interface. Here, *adalah* takes over two functions. It marks the boundary between the subject and predicate and smoothens the intonation contour. Since *adalah* is the last resort, *adalah*-insertion is unlikely (or sometimes even impossible) if another strategy has taken over this function. These strategies involve an overt realization of T, overt focalization or topicalization of the subject or predicate, or insertion of a particle between the subject and predicate. In all other cases, *adalah* insertion was possible (at least in theory). In reality, utterance-internal features such as the length of the constituents and utterance-external factors such as register, context, situation, participants, and speaker's choice influence the *adalah*-choice. Therefore, *adalah*-insertion is not entirely predictable, and to a certain extent, the *adalah*-puzzle will most likely remain unsolved.

Abbreviations

1	first person	3	third person
AG.VOIC	agent voice	ANT	anterior
AP	adjective phrase	CAUS.NOM	cause nominalizer
CL.HUM	classifier human	COMP	complementizer
CONT	Continuative	COP	copula
DET	Determiner	EMPAR	emotive particle
FOC	Focus	FUT	future
NEG	Negation	NOM	nominalizer
NP	noun phrase	PERF	perfect
PL	Plural	POSS	possessive pronoun
PP	preposition phrase	PROG	progressive
QUEST	question marker	REL	relative pronoun
SG	Singular	TOP	topic
VP	verb phrase		

References

Arka, I Wayan. 2013a. On the typology and syntax of TAM in Indonesian. In John Bowden (ed.), *Tense, aspect, mood and evidentiality in languages of Indonesia*. 23–40. Pusat Kajian Budaya dan Bahasa. Jakarta: Atma Jaya Catholic University of Indonesia.

Arka, I Wayan. 2013b. Nonverbal predicates in Austronesian and Papuan languages: An LFG perspective. *Proceedings of the 6th International Seminar on Austronesian and non-Austronesian Languages and Literatures.* 30–46.

Behaghel, Otto. 1909. Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25. 110–142.

Bowers, John. 1993. The syntax of predication. *Linguistic Inquiry* 24 (4). 591–656.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective.* Oxford: Oxford University Press.

Cumming, Susanna. 1991. *Functional change: The case of Malay constituent order.* The Hague: Mouton de Gruyter.

Eid, Mushira. 1983. The copula function of pronouns. *Lingua* 59 (2–3). 197–207.

Guilfoyle, Eithne, Henrietta Hung & Lisa Travis. 1992. Spec of IP and Spec of VP: Two subjects in Austronesian languages. *Natural Language & Linguistic Theory* 10 (3) 375–414.

Halim, Amran. 1975. *Intonation in relation to syntax in Bahasa Indonesia.* Jakarta: Djambatan.

Harley, Heidi Britton. 1995. *Subjects, events, and licensing.* Cambridge, MA: MIT dissertation.

Hopper, Paul J. 1972. Verbless stative sentences in Indonesian. In John W. M. Verhaar (ed.), *The verb 'be' and its synonyms: Part V.* 115–152. Berlin: Springer.

Kaswanti Purwo, Bambang. 1984. The categorical system in contemporary Indonesian: Pronouns. *Towards a description of contemporary Indonesian: preliminary studies Part 2 (NUSA 19).* 55–73.

Kratzer, Angelika. 1996. Severing the external argument from its verb. In Johan Rooryck & Laurie Ann Zaring (eds.), *Phrase structure and the lexicon.* 109–137. Heidelberg: Springer.

Liberman, Mark Yoffe. 1975. *The Intonational System of English.* Cambridge, MA: MIT dissertation.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the H&H theory. In William J. Hardcastle & Alain Marchal (eds.), *Speech production and speech modelling.* 403–439. Heidelberg: Springer.

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania working papers in linguistics* 4 (2). 201–225.

Moeljadi, David, Francis Bond & Luís Morgado da Costa. 2016. Basic copula clauses in Indonesian. *Proceeding of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar.* 442–456.

Musgrave, Simon. 2001. Non-Subject Arguments in Indonesian. Ph.D. thesis. Melbourne, VIC: University of Melbourne.

Mustaffa, Amir Rashad bin. 2018. Silent syntactic structures in Malay: Copular clauses. *Journal of Modern Languages* 28. 20–43.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology.* Dordrecht: Foris.

Pierrehumbert, Janet. Breckenridge. 1980. The phonology and phonetics of English intonation. Ph.D. thesis. Cambridge, MA: MIT.

Pustet, Regina. 2003. *Copulas: Universals in the categorization of the lexicon*. Oxford: Oxford University Press.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar: Handbook in generative syntax*. 281–337. Heidelberg: Springer.

Sneddon, James Neil. 1996. *Indonesian: A comprehensive grammar*. London: Routledge.

Sneddon, James Neil, K. Alexander Adelaar, Dwi N. Djenar & Michael Ewing. 2010. *Indonesian: A comprehensive grammar*. 2nd edn. London: Routledge.

Tadmor, Uri. 2007. Grammatical borrowing in Indonesian. In Yaron Matras, Jeanette Sakel (eds.), *Grammatical borrowing in cross-linguistic perspective*. 301–328. The Hague: Mouton de Gruyter.

Wasow, Thomas. 1997. Remarks on grammatical weight. *Language variation and change* 9 (1). 81–105.