

***This is the author's accepted manuscript before proof correction**

The (non-)canonical status of the *ka*- passive in Balinese

I Nyoman Udayana, Gede Primahadi Wijaya Rajeg, Ida Ayu Made Puspani

Abstract: In Balinese, a non-canonical passive construction exists alongside the canonical passive. This non-canonical form shares some properties with the canonical passive but differs in key ways, particularly in expressing an unintentional agent. Similar to the English *get* passive, this study explores this construction by compiling a corpus of Balinese passives from storybooks and conducting grammaticality judgment tests. Results reveal that the *ka*- passive is incompatible with agent-oriented adverbials and purposive clauses, and contrasts semantically with the *-a* passive, which marks accomplishments. The *ka*- passive, associated with achieved states in Austronesian languages, often appears as a high-register or polite form, suggesting its (non-) canonical status is gradient due to language change.

Keywords: *ka*- passive, *-a* passive, achievement, accomplishment, language change

ERRATUM

- In the main [README page](#) of the supplementary material repository, I (GPWR) acknowledge the [erratum I made for producing Figure 7](#) in the published manuscript **BUT NOTHING CHANGE FOR THE CORRESPONDING STATISTICAL TEST RESULT**. Therefore, the statistical test for Figure 7 and the overall conclusion of the paper still holds given the data set we analysed.

Second language abstract

Pada bahasa Bali, di samping konstruksi pasif non-kanonis ada juga konstruksi pasif kanonis. Bentuk non-kanonis ini memiliki beberapa kesamaan dengan pasif kanonis tetapi berbeda dalam beberapa hal penting, khususnya dalam mengekspresikan agen yang tidak disengaja. Sama halnya dengan pasif *get* dalam bahasa Inggris, penelitian ini mengeksplorasi konstruksi ini dengan menggunakan data korpus cerita bahasa Bali dan melakukan eksperimen uji penilaian gramatikalitas. Hasilnya menunjukkan bahwa pasif *ka*- dalam bahasa Bali tidak sejalan dengan adverbial berorientasi agen dan klausa *purposive*, dan berbeda secara semantik dengan pasif *-a*, yang menandai *achievement*. Pasif *ka*-, yang dikaitkan dengan keadaan yang dicapai dalam bahasa Austronesia, sering muncul dalam tingkat bahasa sopan. Makalah ini menunjukkan status (non-)kanonisitas bersifat gradien karena perubahan bahasa.

Kata kunci: pasif *ka*-, pasif *-a*, pencapaian, penyelesaian, perubahan bahasa

1. Introduction

Beside the conventional/canonical passive construction, there exists a non-canonical passive structure. The latter is characterized by deviating from the standard passive, as described by Alexiadou (2012). This deviation is exemplified by what is known as the *get* passive in English, as opposed to the *be* passive. Alexiadou suggests that equivalent constructions in other languages may exhibit similar properties to the English *get* passive. Indeed, the *ka*-passive in Balinese demonstrates such properties, as illustrated in (1a)¹.

- (1) a. *Ia ka-tanjung taken I Nyoman*
3 PASS-kick by ART name
(S)he got kicked by I Nyoman'
(Constructed_sentence)
- b. *Ia tanjung-a taken I Nyoman*
3 kick-PASS by ART name
(S)he was hit by I Nyoman'
(Constructed_sentence)

The distinction between (1a) and (1b) lies in the use of different passive markers: the former employs *ka*- while the latter uses *-a*, resulting in distinct meanings. This prompts inquiries into how these semantic variances are achieved and how Balinese incorporates linguistic properties associated with these passive forms. Of utmost importance is whether these forms may share meaning or canonicity status in the gradual development of the language. This paper endeavors to address these questions.

Focusing on the (non-)canonical status of the *ka*-passive in contrast to the *-a* passive, which we posit as the canonical passive in Balinese, this paper is structured as follows. Section 2 reviews prior research on Balinese passives, and Section 3 outlines the study's aims. Section 4 details the data and methodology of the study. Section 5 presents the findings from corpus study (§5.1) and grammaticality rating (§5.2), and Section 6 discusses the implications of the *ka*-passive's (non-)canonical status for language change. Section 7 concludes the paper.

2 Previous studies on Balinese passives

Numerous studies have investigated Balinese passives, including those by Artawa (1994, 2013), Hunter (1988), Arka (1998, 2003), Udayana (2013), and Darmasetiyawan & Ambridge (2022). Arka (2003) identifies two passive forms in Balinese: the *-a* passive and the *ka*-passive, noting that the *-a* form is complex, functioning either as a pronominal clitic or a passive marker (see Section 6.1). Arka (2003) highlights that the *-a* form, though bound, carries a referential index, allowing it to bind reflexive anaphors and function similarly to a free pronominal. He is also the first to associate these passive forms with Balinese high and low speech styles.

¹ Sentence examples are glossed following the Leipzig Glossing Rules. Symbol used in the glosses: 3 third person; abbreviation used in the glosses: ART article, AV agentive-voice verbal prefix; DEM demonstrative; NEG negative; OV objective-voice; PASS passive voice affix; POSS possessive; REL relativiser. The supplementary materials (data and R codes for the quantitative analyses) are available open access on the *Open Science Framework* repository at <https://doi.org/10.17605/OSF.IO/8H2AT> (Udayana et al. 2024).

Building on this, Udayana (2013) argues that Balinese passive forms originate from Javanese. Drawing on Zoetmulder and Poedjawitjana (1993), he proposes that Old Javanese included both the <*in*> passive and the *ka*- passive. The former evolves into the *di*- passive in Modern Javanese and the *-a* passive in Balinese, while the *ka*- passive remained consistent across both languages. Udayana further posits that while the *-a* passive serves as the canonical passive in Balinese, the *ka*- passive does not. Artawa (2013) similarly identifies the *ka*- and *-a* passives in Balinese and, referencing Kersten (1984), suggests that the *ma*- form might also function as a passive marker. However, Udayana (2013) counters this, proposing that *ma*- more accurately marks the middle voice. Recently, Darmasetiyawan and Ambridge (2022) provide experimental evidence for the existing semantic representation of morphological passives in Balinese, supporting the semantic-based account of syntactic representation. Notably, though, none of these studies address the (non-)canonical status of the *ka*- passive, a critical gap that this study aims to explore.

3 Aims

This study is based on the hypothesis that when two languages have two forms of the passive voice, one should correspond to the canonical passive, while the other represents a non-canonical variation. This idea is supported by the analysis of the English passive voice, where the *get* passive shares some, but not all, features of the *be* passive. Consequently, the *get* passive is categorized as a non-canonical type, as opposed to the *be* passive (Reed 2011, Alexiadou 2012, Alexiadou and Schäfer 2013). This phenomenon also applies to Balinese, where the *ka*- passive exhibits similar characteristics to the *get* passive in English. Our study aims to examine the behavior of the *ka*- passive (in comparison to the *-a* passive) in two ways. Firstly, its properties diverge from those of the canonical *-a* passive. Secondly, conversely, the properties of the *ka*- passive may align with those of the *-a* passive. We suggest that the degree of this deviation could be attributed to language change, as canonicity itself is a nuanced and gradient concept (Brown et al. 2013; Arts 2004, 2007; Shibatani 1985).

4 Data and methodology

This study integrates corpus and grammatical judgment data (cf. Rajeg et al. 2022 for a similar multimethodological approach to voice alternation in Indonesian). A corpus of *ka*- and *-a* passive constructions was compiled from four Balinese storybooks. Each passive clause was annotated for its verb root, speech style (high, low, or mid-level), presence of agent phrases, and whether agentive entities appeared in the same clause, a prior one, or were unidentified.

To complement the corpus study, a grammaticality judgment questionnaire was distributed online to undergraduate Balinese majors at Udayana University, Mahadewa University, and STKIP Agama Hindu Amlapura. Ninety-nine students responded, but five non-native Balinese speakers from Udayana were excluded, leaving ninety-four native-speaking participants. Participants rated the acceptability of twenty-four constructed Balinese sentences on a 5-point Likert scale. Twenty sentences assessed (non-)canonical status based on factors like (i) compatibility with agent-oriented adverbs, (ii) use in imperative clauses, (iii) opposition of verb meanings with *ka*- and *-a* markings, (iv) duration adverbials, and (v) ability to control implicit agents in purposive clauses. Four additional sentences tested *ka*- and *-a* passives with high

speech level verbs to examine high passive usage. The questionnaire purpose was undisclosed to participants.

5. Results

In this study, we employed two primary data sources to investigate the status of the *ka*- passive: corpus data (§5.1) and experimental, grammaticality judgement data (§5.2). The corpus data provided insights from naturally occurring language usage. The experimental data included tasks where participants provided grammaticality judgments, as well as other controlled measurements. The results from these data sources are presented below.

5.1 Corpus results

The corpus data includes instances of the *-a* and *ka*- passives used in sentences with and without overtly mentioning the agent (§5.1.1), their occurrence in coordinative structures (§5.1.2), and their occurrence with verb roots at different speech levels (§5.1.3).

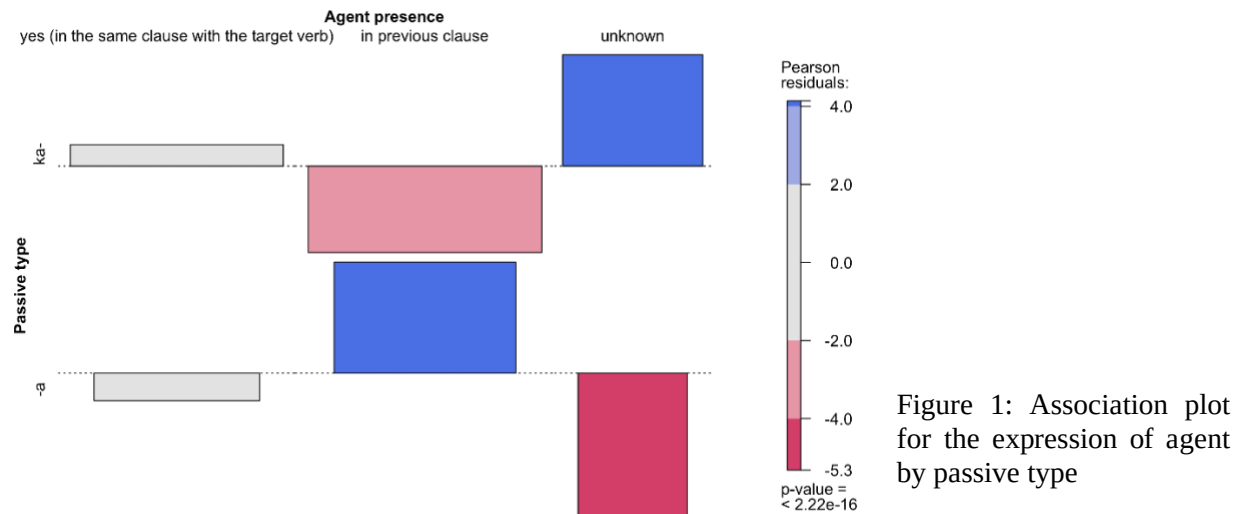
5.1.1 (In)explicit encoding of the agent

This analysis aims at determining the extent to which the Agent argument is expressed explicitly (i) within the same clause as the target *-a* and *ka*- passive verbs, (ii) in the previous clauses, or (iii) unknown (see Table 1).

Table 1: Distribution of the agent argument for *-a* and *ka*- passive in the corpus

Agent presence	<i>-a</i> passive	<i>ka</i> - passive
yes (in the same clause)	57 (33.14%)	116 (40.85%)
in previous clause	115 (66.86%)	93 (32.75%)
Unknown	0	75 (26.41%)

The association plot in Figure 1 shows that *-a* passive more strongly mentions the agent in the previous clause (cf. the rising bluish bars) rather than in the same clause as the *-a* passive target verb. In contrast, the agent for the *ka*- passive is more strongly unknown. The *-a* and *ka*- passives show relatively equal preference in explicitly mentioning agent in the same clause as the target passive verbs. These different preferences are highly significant (with moderate effect size of Cramer's $V = 0.4$) ($X^2=74.43$, $df=2$, $p_{\text{test-for-independence}} < 0.0001$).



Examples (2-4) below show how the agent is expressed for the *-a* and *ka*- passives in the three conditions in Figure 1.

- (2) a. *Ditu ia ka-trima olih bapa=ne*
 there 3 PASS-receive by father=3POSS
 'There he was received by her father'
 (KP: 37)
- b. *Pabesen bapa=ne ento stata ingetang-a pesan*
 advice father=3POSS DEM always remember-PASS very
teken I Gede Sastra Wijaya
 by ART name
 'His father's advice is always well remembered by I Gede Sastra Wijaya'
 (KP: 37)
- (3) *Gelang emas ane 24 karat ane baat-ne makejang*
 bracelet gold REL 24 carat REL heavy-DEF all
100 gram jani suba anggon-a tur maan hadiah
 100 gram now PERF wear-PASS and receive present
 'A 24-carat gold bracelet weighing 100 gram has now been worn
 and she has also received a present'
 (KP: 10)
- (4) *Ka-carita dugas imalu ade kone Sang Resi ane*
 PASS-story since long.ago exist people.say ART priest REL
maparab Bhagawan Kasiapa
 MV.name Name

‘It was told that long time ago there was a priest called Bhagawan Kasiapa’
(IK: 25)

The agent expressions in the *ka-* and *-a* passives in examples (2a) and (2b) are conveyed without any explicit context of intentionality. Due to a nuanced difference between them, they might both be labeled as volitional passives. In example (3), the absence of the agent expressions in the same clause causes the *-a* form to be confused between a pronominal marker and the passive marker itself. In example (4), the co-occurrence of the *ka-* passive and the non-identification of the agent paves the way for the *ka-* passive to indicate an achieved state in Austronesian languages (AN) (see section 6.1).

5.1.2 The *-a* and *ka-* passives in coordination structure

This analysis aims to measure the frequency of co-occurrence of the same and mixed passive types in coordinate structure. The same co-occurrence will be the co-occurrence of *-a* or *ka-* passives. Meanwhile, the mixed co-occurrence means the combination of *-a* and *ka-* passive in a coordinate structure in the same clause/sentence. In total, there are 27 attestations of coordinate structure for the two passive types, the distribution of which is shown in Table 2.

Table 2: Distribution of the coordinated passive types in a clause

Coordinated passive type	Frequency	%
<i>-a</i> with <i>-a</i>	11	40.7
<i>ka-</i> with <i>ka-</i>	10	37.0
mixed (<i>ka-</i> with <i>-a</i> , vice versa)	6	22.2

Chi-Square test for Goodness-of-Fit for the Frequency column in Table 2 reveals no significant difference between the coordinated passive type ($X^2=1.5556$, $df=2$, $p_{\text{Goodness-of-fit}}=0.4594$). Nevertheless, the overall picture is that there is still a significantly higher proportion of coordination between the same passive types combined (21 cases) compared to the mixed coordination (6 cases) ($X^2=8.33$, $df=1$, $p_{\text{Goodness-of-fit}} < 0.05$).

Preferably, two passive clauses are coordinated using the same passive formative (i.e., *ka-* and *ka-* or *-a* and *-a*). However, based on the data, there are 6 cases where two passives are coordinated with different formatives, suggesting that the *ka-* passive converges from its initial non-canonical status to a newly acquired canonical status.

5.1.3 Speech level of the root and the passive types

This analysis aims to determine if there is association between the speech level of the root (i.e., high, low, neutral [or *mider* in Balinese]) and the passive type it is combined with. The motivation for this analysis is that the *-a* passive, historically considered to be derived from the third person pronoun in low speech style, is associated with low passive. In contrast, the *ka-* passive, which is not influenced by this phenomenon, is regarded as belonging to the classification of high passive (Arka 2003).

The root in the passive database from the corpus was annotated for one of the three speech level categories (high, low, neutral). Then, for a given passive type, the quantitative analysis captures the type-frequency (i.e., the number of unique form) of the root from a given speech level. The next steps are (i) cross-tabulating these two frequencies of co-occurrences for a given speech level of the root with the *-a* and *ka-* passives (cf. Table 3), and (ii) running the

statistical, Chi-Square Test to determine the degree of association between the passive types and the root speech level, which is visualized as an association plot in Figure 2.

Table 3: Type frequency of roots from three speech levels with *-a* and *ka-* passives.

Root Speech Level	<i>-a</i> passive	<i>ka-</i> passive
Low	81 (76.42%)	60 (40.82%)
Neutral (i.e., <i>mider</i>) ²	13 (12.26%)	31 (21.09%)
High	12 (11.32%)	56 (38.1%)

Overall, we found a highly significant association ($X^2=33.19$, $df=2$, $p_{\text{test-for-independence}} < 0.0001$) with moderate effect (Cramer's $V = 0.36$) between the root speech level and the passive type. The specific association and its degree are visualized in Figure 2. The *ka-* passive is strongly associated with root from the high speech level (cf. the bluish bar, indicating strong positive residuals), meanwhile the *-a* passive is strongly associated with root from the low speech level.

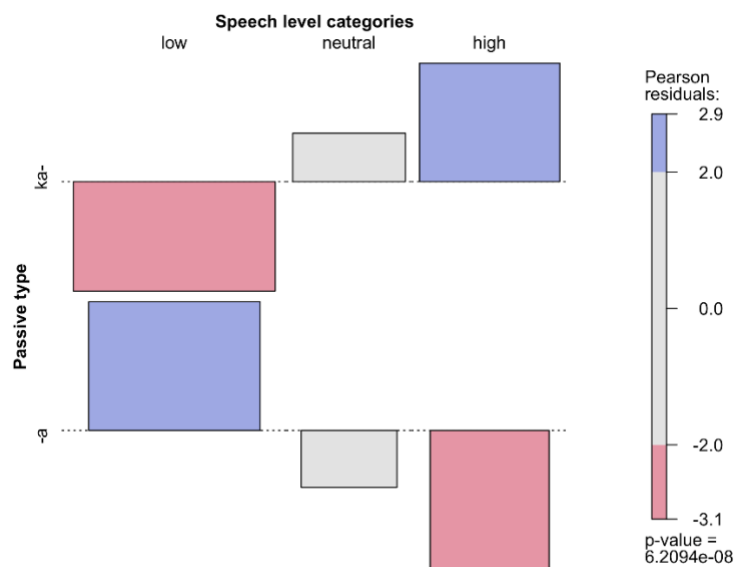


Figure 2: The association plot between the *-a* and *ka-* passives and the root speech level

Based on the corpus sample, the expectation that the *ka-* passive is associated with high speech level roots is confirmed (cf. Arka 2003). This is because the *ka-* passive, considered a high passive, requires that the selected verb root must be of high style. Consequently, verb roots from the low speech style have to be avoided.

5.2 Grammaticality rating results

The experimental data examines participants' ratings concerning the non-canonical status of the *ka-* passive in environments where it can and cannot be used. These environments include negative imperative structures (§5.2.1), subordinative purposive clauses (§5.2.2), adverbial of duration (§5.2.3), the speech level of the verb root (§5.2.4), the occurrence of the *ka-* passive with agent-oriented adverbials (§5.2.5), and contrastive coordination structures (§5.2.6).

² *Mider* is associated with a lexicon, whether native or non-native, that is considered neutral regarding its categorization into high speech or low speech.

5.2.1 Negative imperative structure

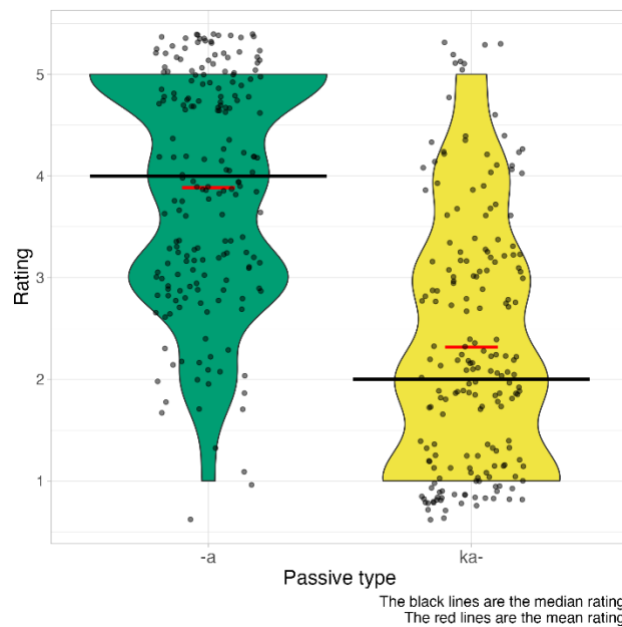


Figure 3: The rating of *-a* and *ka-* passives in negative imperative structures

The violin plot shows the distribution of the ratings, including the mean (red lines) and the median (black horizontal lines); the dots indicate the data points/responses. The Wilcoxon rank-sum test reveals a highly significant difference in the mean ratings of the two passives in negative imperative constructions ($W = 29452$, $p < 0.0001$). Specifically, the *-a* passive (in green) ($\bar{x} = 3.88$, $s = 1.12$; $Med = 4$, $IQR = 2$) is rated, on average, significantly higher than the *ka-* passive (in yellow) ($\bar{x} = 2.31$, $s = 1.21$; $Med = 2$, $IQR=2$) in negative imperative constructions; this suggests that the *-a* passive seems to be more acceptable/conventional to be negated in the construction than the *ka-* passive.

5.2.2 Subordinative Purposive Clause

As in the negative imperative construction, the Wilcoxon rank-sum test indicates a highly significant difference ($W = 25978$, $p < 0.0001$) in the mean ratings of the two passives when appearing in the context of subordinative purposive clause.

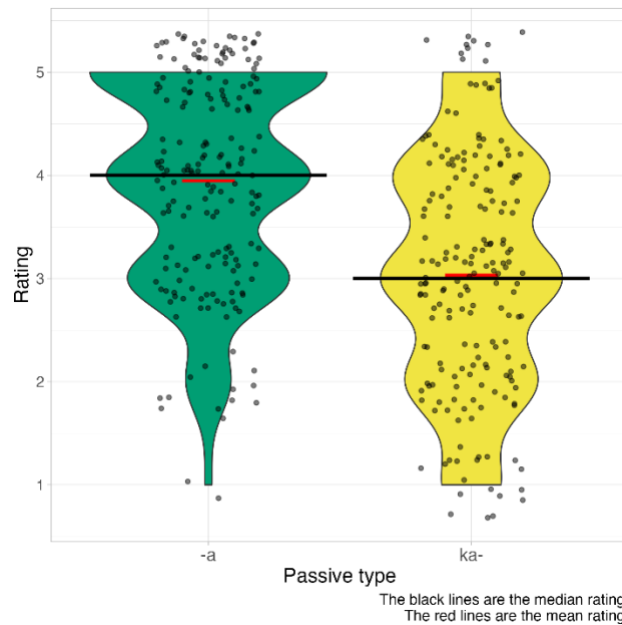


Figure 4: The rating of *-a* and *ka-* passives in subordinative purposive clauses

In particular, the *-a* passive ($\bar{x} = 3.94$, $s = 0.99$; $Med = 4$, $IQR = 2$) is, on average, rated significantly higher than the *ka-* passive ($\bar{x} = 3.03$, $s = 1.13$; $Med = 3$, $IQR = 2$) when they appear in subordinate clauses (i.e., purposive clauses).

5.2.3 Adverbial of duration

The Wilcoxon rank sum test indicates a highly significant difference ($W=55398$, $p < 0.0001$) in the mean rating of the two passives (Figure 5) when co-occurring with the adverb of duration *makelonè* ‘for the duration of’.

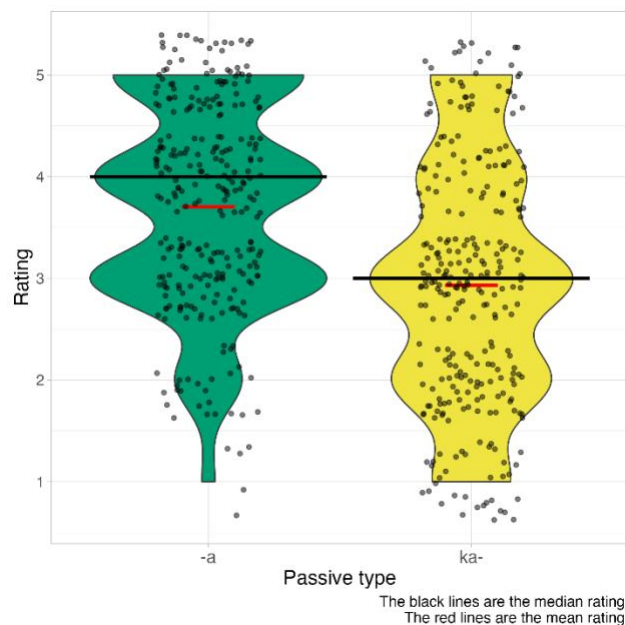


Figure 5: The rating of *-a-* and *ka-* passives co-occurring with an adverbial of duration

As in the previous two properties, the *-a* passive ($\bar{x} = 3.70$, $s = 1$; $Med = 4$, $IQR = 2$) is, on average, rated more highly (thus likely more acceptable) than *ka-* passive ($\bar{x} = 2.93$, $s = 1.19$; $Med = 4$, $IQR = 2$) to occur in a clause with the given adverb of duration.

5.2.4 Speech level of the root

This stimulus, using passives with verb roots in high and low speech styles, is intended to complement the corpus analysis in Section 5.1.3. The experimental data specifically examines whether verbs of high speech level/style agree with the *-a* passive and/or with the *ka-* passive.

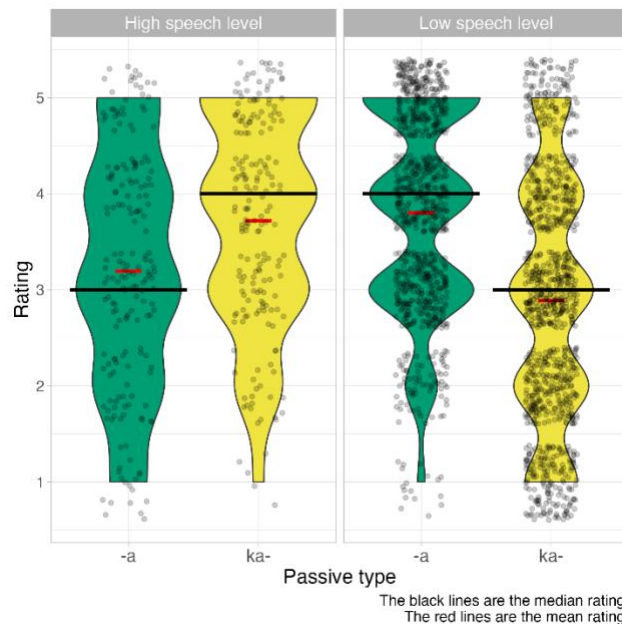


Figure 6: The rating of *-a* and *ka-* passives with a given speech level of verb root

Wilcoxon tests with Holm's corrected p -values indicate that the *-a* and *ka-* passives significantly differ in their use with verb roots of different speech levels ($p < 0.0001$ within both the high speech-level [left panel] and low speech-level [right panel] root groups). The violin plots indicate a contrasting preference. On the left panel in Figure 6, we observe that the *ka-* passive has a significantly higher rating compared to the *-a* passive to co-occur with high speech level roots. On the contrary, on the right panel, the *-a* passive is rated significantly higher compared to the *ka-* passive to co-occur with roots of low speech level.

5.2.5 Agent-oriented adverbial

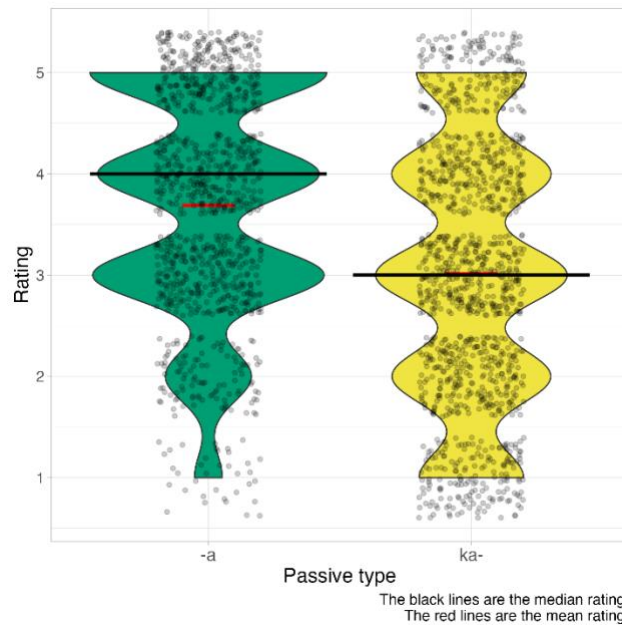


Figure 7: The rating of *-a* and *ka*- passives modified with an adverb of intentionality

The Wilcoxon rank-sum test indicates that the *-a* passive is considered significantly more conventional/acceptable than the *ka*- passive (hence receiving a higher rating; see Figure 7 above) when co-occurring with a verb root in a clause containing the agent-oriented adverbial *nyelapang* ‘intentionally’ ($W = 27262$, $p < 0.0001$).

5.2.6 Contrastive Coordination Structure

The aim of this rating is to offer a preliminary test regarding the semantic differentiation between the *-a* and *ka*- passives in a contrastive coordination structure with *kewala* ‘but’. The theoretical idea is that the *-a* passive and *ka*- passive exhibit nuances: the latter indicates ‘unintentionality’ (§5.2.5). If this is the case, then the rated sentence (see example 5 below) contrasting the two passives should trigger higher rating scores than just the mid-level score of the rating scale (here, we use the value of three [on the five-scale values] as the mid-level baseline).

- (5) *Ia tusing ka-lempag kewala lempag-a taken I Made*
 3 NEG PASS-hit but hit-PASS by ART name
 ‘(S)he did not get hit but (s)he was (intentionally) hit by I Made’
 Constructed_sentence

The higher rating will be interpreted as highlighting the acceptability of the sentence in approximating the contrast between the two passives. As this is a preliminary analysis, we focused on contrasting the passives with two highly agentive verb roots, namely *lempag* ‘hit/smash’ and *tanjung* ‘kick’, as test cases. Figure 8 below summarizes the distribution of the ratings for these two verbs.

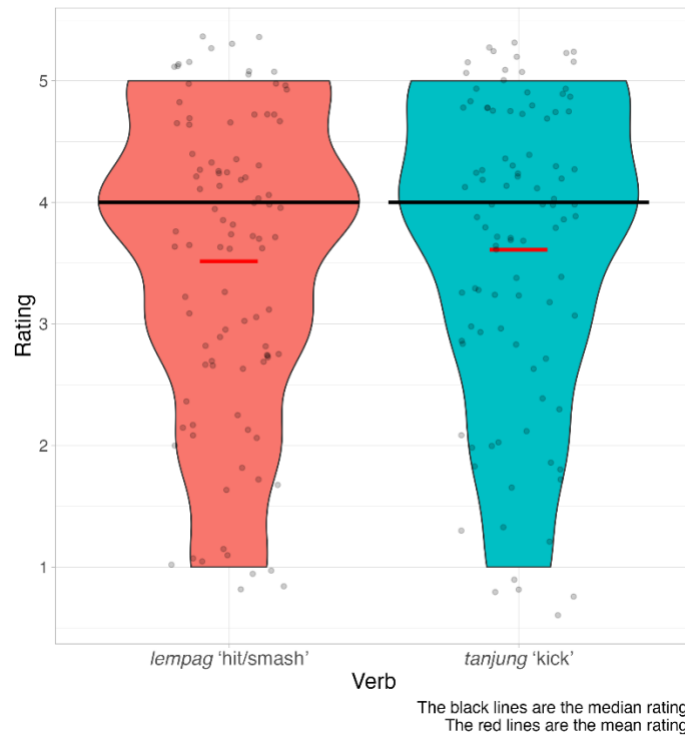


Figure 8: The rating *-a* and *ka*- passives in contrastive coordinate structures with the two selected verbs

The Shapiro-Wilk test for normality (Gries 2009: 150) for each verb, as they are contrasted in the two passive patterns, indicates that the ratings significantly deviate from a normal distribution (for *lempag* 'hit/smash': $W = 0.87764$, $p < 0.0001$; for *tanjung* 'kick': $W = 0.86577$, $p < 0.0001$). In other words, the distribution of the rating scores is skewed and not centralized in the mid-level rating. The density plot in Figure 9 below further shows the skewness of the ratings towards higher scores (i.e., higher than the central three-point).

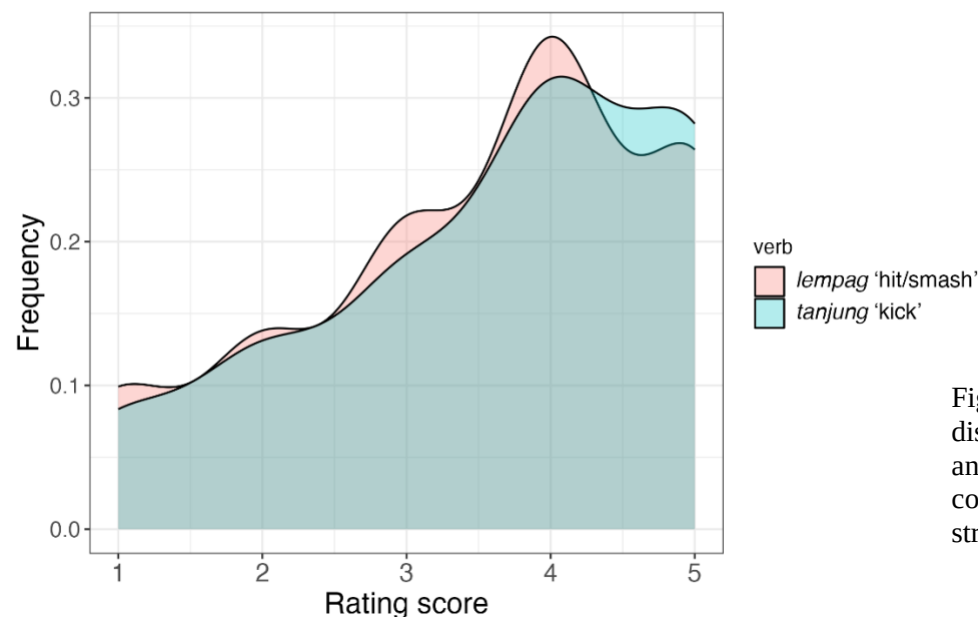


Figure 9: Density plot for the distribution of ratings for *-a* and *ka*- passives in contrastive coordinating structure

Comparing the mean ratings for the two verbs, no significant difference is found ($W = 4301.5$, $p = 0.5652$), as indicated by the overlapping median lines in Figure 8 for the two verbs. These results suggest that, for the two verbs we preliminarily used, the two passives are still distinguishable in the contrastive structure for each verb.

6 Discussion

6.1 Non-canonicity status of the *ka*- passive

This subsection explores the non-canonical status of the *ka*- passive in Balinese, supported by results from grammatical judgment tests that show a preference for viewing it as non-canonical. We attribute this to the influence of Old Javanese (OJ) on Balinese (Clyne 1989, 1994). The *ka*- passive in Balinese originates from OJ, which used two passives: the *ka*- passive (indicating an achieved state) and the *<in>* passive (denoting an action passive) (Blust & Trussel 2020; Zoetmolder & Poedjawijatna 1993). The *ka*- passive, part of the Proto Malayo-Polynesian (PMP) and Proto-Austronesian (PAN) language families, aligns with non-agentive passive constructions, while the *-a* passive, which replaces *<in>* in Balinese, corresponds to agentive passive usage.

The non-agentive aspect of the *ka*- passive, underlying its non-canonical status, is evident in imperatives. Imperatives typically involve agentive actions with an indirect agentive role for the addressee (Isac 2015; Rupp 2003; Kaufmann 2012; Wurff 2007; Eikhenvald 2010). Consequently, the *ka*- passive is less compatible with imperatives, unlike the *-a* passive, which is more suitable for such contexts, as shown in (6). The grammaticality rating data presented in §5.2.1 provides evidence for the incompatibility of *ka*- in imperative context (as it received a significantly lower rating than the *-a* passive in this context).

- (6) a. *Da* *Made* *tanjung-a*
 NEG name kick-PASS
 ‘Do not let made be kicked’
 (Constructed_sentence)
- b. * *Da* *Made* *ka-tanjung*
 NEG name PASS-kick
 ‘Do not let Made be kicked’
 (Constructed_sentence)

Secondly, the non-agentive characteristic of the *ka*- passive implies that this type of passive clause cannot be embedded within a purposive clause, as shown by its lower acceptability rating in §5.2.2. In simpler terms, the agent associated with the active clause or the main clause cannot control the action denoted by the purposive clause (Alexiadou 2012, Alexiadou and Schäfer 2013). Thus, in (7a), the individual desiring to obtain money cannot be linked to the implicit agent. Instead, the desire must be attributed to an entity unrelated to the agent, as observed in the grammaticality of (7b). However, (7c) is acceptable because, in the *-a* passive, the implied agent controls the event in the purpose clause.

- (7) a. *Anake_i ento ka-lempag [PRO_{*i/*j} pang maan pipis]*
 child-DEF DEM PASS-hit- To get money
 'The child was hit to get some money'
 (Constructed_sentence)
- b. *Anak-e ento ka-lempag [pang ia maan pipis]*
 child-DEF DEM PASS-hit so.that 3 get money
 'The child was hit so that he could get some money'
 (Constructed_sentence)
- c. *Anake_i ento lempag-a [PRO_{*i/j} pang maan pipis]*
 child-DEF DEM hit-PASS To get money
 'The child was hit to get some money'
 (Constructed_sentence)

Thirdly, the *ka*- form, which denotes an achieved state, cannot be associated with an eventuality embedded with an adverbial of duration (cf. §5.2.3 whereby *ka*- receives a significantly lower rating for this adverb of duration variable than the *a*- passive). On the other hand, the *-a* passive can be, as it signifies accomplishment (Alexiadou 2012, Alexiadou and Schäfer 2013).

- (8) a. * *Luu-ne ka-sampatang makelo-ne limang menit*
 rubbish-DEF PASS-sweep long-DEF five minute
 'The rubbish got swept for five minutes'
 (Constructed_sentence)
- b. *Luu-ne sampatang-a makelo-ne limang menit*
 rubbish-DEF sweep-PASS long-DEF five minute
 'The rubbish was swept for five minutes'
 (Constructed_sentence)

Fourthly, as a non-agentive passive, the *ka*- passive does not pair well with agent-oriented adverbials, such as that indicating intentionality (cf. §5.2.5). Lastly, because the *ka*- passive differs in meaning from the *-a* passive, they can be contrasted using the coordinating conjunction *sakewala* 'but' (see example (5)).

6.2 The changed status of the *ka*- passive reflects its role in the gradual development of Balinese

In Section 6.1, we discussed the historically non-canonical status of the *ka*- passive, tracing its origins to OJ. Though *ka*- retains some non-canonical properties today (as evidenced by the results of grammaticality ratings), Balinese speakers show varied interpretations of its usage. This section explores its shift toward a more canonical status, driven by internal linguistic developments. While Balinese has borrowed extensively from OJ (Shadeg 1977, Udayana 2015), it has also developed its own lexicon. OJ and Old Balinese initially lacked a speech style system, but both languages have since evolved to include one.

At the grammatical level, a key shift is Balinese's replacement of the OJ agentive marker <in> (Tinggen 1987) with the *di-* passive, while retaining the *ka-* passive. In Modern Javanese, the *ka-* passive is largely non-canonical or adversative (Davies 1995). However, in Balinese, as we have found in this study, the *ka-* passive is attested in certain canonical contexts (such as its use with overt oblique agent as in [2a]); this appearance of *ka-* with canonical context could possibly be influenced by the introduction of the *-a* passive. Arka (2003) observes that short passives, which lack an explicit agent phrase, often carry ambiguity. The *-a* suffix can signal either a third-person pronominal reading or serve as a passive marker. This dual interpretation is seen in example (9). Example (9a) is the original attestation from the corpus we used; we do not gloss it here because it can have multiple interpretation shown in examples (9b-c), which are constructed based on (9a). The dual interpretation of *-a* is shown where it appears as a passive marker (9b) or as an agent of third-person pronominal clitic =*a* in Objective Voice construction (9c). The ambiguous reading of *-a* in (9a) disappears when it is expressed as an Agentive Voice (AV) construction in (9d). This flexibility is aligned with the symmetrical voice system in Balinese and other Austronesian languages (Arka 2003; Himmelmann & Riesberg 2013; Riesberg 2014; Riesberg & Primus 2015).

- (9) a. *I Puwuh uber-a kanti Ke padange ngerembun totonan*
(PMB: 4)
- b. *I Puwuh uber-a kanti Ke padang-e ngerembun totonan*
ART Name chase-PASS until To grass-DEF lush DEM
'I Puwuh was chased into the lush field'
- c. *I Puwuh uber=a kanti Ke padage ngerembun totonan*
ART Name OV.chase=3 until To grass.DEF lush DEM
'(S)he chased I Puwuh into the lush field'
- d. *Ia ng-uber I Puwuh kanti Ke padang-e ngerembun totonan*
3 AV-chase ART name until To grass-DEF lush DEM
'(s)he chased I Puwuh into the lush field'

At first glance, using the *-a* formative as a passive marker in Balinese seems straightforward. Originating from the third-person pronoun *ia* '(s)he' in the low speech style (contrasting with *ida* for high style), it is now lexicalized as a passive marker that appears style-neutral. However, in practice, the Balinese speech style system complicates its use. Balinese politeness, unlike the honorific markers in languages like Japanese and Korean, operates through an asymmetrical speech style system (Arka 2005). This system functions as follows:

Scenario (1): A commoner addressing someone of higher social status must use high-register vocabulary. For passive expressions, they choose between *ka-* and *-a* formatives. Selecting the correct formative is essential to convey respect.

Scenario (2): Using *-a* might imply lower respect due to its association with low-register forms. To avoid any perceived disrespect, speakers opt for *ka-* in high-register contexts.

Thus, the *ka-* passive serves as both low and high register, with speakers favoring it for high-status addressees. Our data confirm a strong preference for *ka-* over *-a* in high-register contexts (cf. §5.1.3 for the preferred high-register root attached to *ka-* in the corpus study). Additionally, the *ka-* passive generally aligns with a volitional (canonical) reading, whether the verb root itself is high or low register. Its flexibility also allows it to coordinate with *-a* in passive constructions, as seen in example (10) (see also §5.1.2).

- (10) a. *Di suba-ne jam 09.00 WITA makejang pelamar ka-kumpulan*
 at after-DEF time 09.00 WITA all applicant PASS-gather
tur tunden-a ka lantai Dua
 and ask-PASS to floor Two
 ‘After 9.00 WITA (Central Indonesian Time) all applicants were gathered and
 asked to enter the second floor’
 (KP: 18)

Sentence (10) illustrates the evolving usage of passive formatives in Balinese, indicating adaptability in language structure.

In sum, the dynamics of Balinese passives indicate that the *ka-* passive is gradually shifting toward canonical status (as it is attested in several canonical contexts), even though the *a-* passive is still preferred in many canonical contexts (such as higher rating for properties indicating agentivity, a marker of canonicity of passive, such as higher frequency of overt agent [§5.1.1], higher rating for adverb of intention [indicating volitionality] [§5.2.5], higher rating for negative imperative structure [§5.2.1]). This shift is shaped by subtle distinctions, such as the modification by agent-oriented adverbials. When decontextualized, the *ka-* passive can function as a volitional, canonical passive, akin to the *-a* passive. We suggest that volitionality aligns with high speech levels in Balinese, where politeness is conveyed through intentional acts, thus reducing unintentional expressions. This supports the universal view that canonicity exists on a gradient. Table 4 summarises the dynamicity of the two passives along the dimensions of canonicity and other hypothesized property (e.g., association with high- or low-registers) in relation to the results of grammaticality rating and corpus studies.

Table 4: The dynamics of the *-a* and *ka-* passives along the canonicity and speech-level dimensions

Dimension	Grammaticality rating results	Corpus Results
Canonicity: Volitionality/Agentivity	<i>-a</i> is preferred than <i>ka-</i> in adverbial of intention rating (§5.2.5)	<i>-a</i> is more frequently used with explicit agent than <i>ka-</i> (§5.1.1)
Canonicity: Negative Imperative	<i>-a</i> is preferred than <i>ka-</i> (§5.2.1)	-
Canonicity: Purposive clause	<i>-a</i> is preferred than <i>ka-</i> (§5.2.2)	-
Canonicity: Adverbial of duration	<i>-a</i> is preferred than <i>ka-</i> (§5.2.3)	-
Canonicity: Passive coordination in coordinative clause (semantic differentiation with respect to volitionality)	<i>-a</i> and <i>ka-</i> is semantically differentiated in contrastive coordinative with <i>kewala</i> 'but' (§5.2.6)	same passive type (e.g., <i>-a</i> and <i>-a</i> or <i>ka-</i> and <i>ka-</i>) is significantly more frequent than mixed type (§5.1.2)
Hypothesised property: Speech-level of the verb root	<i>ka-</i> is preferred for high-register root than <i>-a</i> (§5.2.4)	<i>ka-</i> has significantly higher type frequency than <i>-a</i> for

		high-register root (§5.1.3)
--	--	-----------------------------

7. Conclusion

Balinese passives have historical roots in OJ, which featured two types: the *<in>* passive, marking an accomplished state, and the *ka*- passive, indicating an achieved state. In Balinese, the *<in>* passive evolved into the *-a* passive, while the *ka*- passive was retained. The *ka*- passive also maintains its original functions from OJ denoting an achieved state.

Balinese uses distinct lexicons to mark speech register levels. In this study, we found a preference of *ka*- passive with high-level root while the *a*- passive prefers low-level root, as shown in Figure 2. The positive association of *ka*- passive with high-level root supports Arka's (2003) assumption, suggesting that *ka*- is prominently used for high-register. Over time, the *-a* passive has become the canonical form, with the *ka*- passive as non-canonical and more context-sensitive, especially when used with agent-oriented adverbials. In neutral contexts, however, the *ka*- passive may be perceived as canonical. Our study, based on data from four storybooks found a few cases of this canonical usage of the *ka*- passive, even though *ka*- is still predominantly non-canonical compared to *-a*, given *ka*-s tendency to receive lower rating than *-a* in the grammaticality judgement for several dimensions of canonicity (see Table 4). We recommend further research with a broader dataset to better capture the *ka*- passive's evolving role.

Acknowledgments

We sincerely thank the anonymous reviewer for the valuable suggestions to improve this article. Our gratitude also goes to the Balinese and Old Javanese students of Udayana University, and to the Balinese Department students of Mahadewa University and STKIP Agama Hindu Negeri Amlapura for their participation in the grammatical judgment study. The second author (GPWR) is grateful for the postdoctoral fellowship at the University of Oxford, UK, funded by the Arts and Humanities Research Council (AHRC) ([AH/W007290/1](https://doi.org/10.1017/9781017007290/1)), that partially supports his contribution to co-authoring this article.

References

- Alexiadou, Artemis. 2012. Noncanonical passives revisited: Parameters of nonactive voice. *Linguistics*, 50 (6) 1079-1110. <https://doi.org/10.1515/ling-2012-0036>
- Alexiadou, Artemis & Florian Schäfer (eds.) 2013. *Non-canonical passives*. Amsterdam: John Benjamins Publishing Company.
- Arka, I Wayan. 1998. *From morphosyntax to pragmatics in Balinese*. Sydney, Australia: University of Sydney Ph.D. dissertation.
- Arka, I Wayan. 2003. *Balinese morphosyntax: A lexical-functional approach*. Canberra: Pacific Linguistics, School of Pacific and Asian Studies, the Australian National University.
- Arka, I Wayan. 2005. Speech levels, Social Predicates and Pragmatic Structure in Balinese: A Lexical Approach in *Pragmatics*, 15, 2-3: 151-168
- Artawa, Ketut. 1994. *Ergativity and Balinese Syntax*. Melbourne, Australia: La Trobe University Ph.D. dissertation.
- Artawa, Ketut. 2013. The basic verb construction in Balinese in Alexander Adelaar (ed), Voice variation in Austronesian Languages of Indonesia. *NUSA*, 54, 5-27.
- Aarts, Bas. 2004. Modelling linguistic gradience. *Studies in Language*, 28: 1-49.
- _____. 2007. *Syntactic gradience*. Oxford: Oxford University Press.

- Blust, Robert & Stephen Trussel. 2010. The Austronesian comparative dictionary: Web edition. Updated 21 June 2020. <http://www.trussel2.com/acd/> (accessed 2 May 2024).
- Brown, Dunstan, Marina Chumakina, and Greville G. Corbett, eds. 2013. *Canonical morphology and syntax*. Oxford: Oxford University Press.
- Clynes, Adrian. 1989. *Speech Style in Javanese and Balinese: A Comparative Study*. Canberra: Australian National University Master Thesis.
- Clynes, Adrian. 1994. Old Javanese influence in Balinese: Balinese speech styles in Dutton, Tom and Darrell T. Tyron (ed.). *Language contact in the Austronesian world*, 140-179. Berlin: Mouton de Gruyter.
- Darmasetiyawan, I Made Sena & Ben Ambridge. 2022. Syntactic Representations Contain Semantic Information: Evidence from Balinese Passives. (Ed.) Dermot Lynott. *Collabra: Psychology* 8(1). 33133. <https://doi.org/10.1525/collabra.33133>.
- Davies, William D. 1995. Javanese adversatives, passives, and mapping theory. *J. Linguistics* 31. 15-51. DOI: <https://doi.org/10.1017/S0022226700000554>.
- Gries, Stefan Th. 2009. *Quantitative corpus linguistics with R: A practical introduction*. London: Routledge.
- Himmelman, Nikolaus P. & Sonja Riesberg. 2013. Symmetrical voice and applicative alternations: Evidence from Totoli. *Oceanic Linguistics*. 52 (2). 396-422. DOI: [10.1353/ol.2013.0021](https://doi.org/10.1353/ol.2013.0021)
- Hunter, Thomas M. 1988. *The Balinese language: Historical background and contemporary state*. Michigan: Michigan University Ph.D. dissertation.
- Isac, Danilea. 2015. *The morphosyntax of imperatives*. Oxford: Oxford University Press.
- Kersten, J SVD. 1984. *Bahasa Bali: Tata bahasa dan kamus bahasa lumrah* [Balinese: Grammar and dictionary of common expressions]. Ende: Nusa Indah.
- Rajeg, I Made, Gede Primahadi Wijaya Rajeg & I Wayan Arka. 2022. Corpus linguistic and experimental studies on the meaning-preserving hypothesis in Indonesian voice alternations. *Linguistics Vanguard* 8(1). 367–382. <https://doi.org/10.1515/lingvan-2020-0104>.
- Reed, Lisa A. 2011. Get-passives. *The Linguistic Review* (28). 41-78. DOI: 10.1515/tlir.2011.022.
- Riesberg, Sonja. 2014. *Symmetrical voice and linking in Western Austronesian languages*. Berlin: Mouton De Gruyter.
- Riesberg, Sonja & Beatrice Primus. 2015. Agent Prominence in Symmetrical Voice Language. *STUF - Language Typology and Universals* 68 (4). 551-564. <https://doi.org/10.1515/stuf-2015-0023>
- Rupp, Laura. 2003. *The syntax of imperatives in English and Germanic: Word order variation in minimalist framework*. New York: Palgrave Macmillan.
- Sedeng, I Nyoman. 2007. *Morfosintaksis Bahasa Bali Dialek Sembiran* [Morphosyntax of Sembiran Dialect Balinese] Denpasar: Ph.D Dissertation. Udayana University.
- Shadeg, N S.V.D. MA. 1977. *A basic Balinese vocabulary*. Denpasar: Dharma Bhakti.
- Shibatani, Masayoshi. 1985. Passives and related constructions: A prototype analysis. *Language*, 61: 821-848.
- Tinggen, I Nengah. 1987. *Tatabasa Bali (Wredi): Istilah-istilah Bali-Indonesia* [Balinese grammar (wredi): Balinese-Indonesian terms]. Singaraja: Sekolah Pendidikan Guru Negeri.
- Udayana, I Nyoman. 2013. *Voice and reflexives in Balinese*. Austin: University of Texas at Austin Ph.D. Dissertation.

- Udayana, I Nyoman. 2015. Speech levels in Balinese. *Proceeding of the International on Language and Arts (ISLA)* – 4, 287-293. Padang: Universitas Negeri Padang.
- Udayana, I Nyoman, Gede Primahadi Wijaya Rajeg & Ida Ayu Made Puspani. 2024. Supplementary materials for “The (non-)canonical status of the *ka*- passive in Balinese.” Open Science Framework (OSF). <https://doi.org/10.17605/OSF.IO/8H2AT>.
- Wurff, Wim van der. 2007. *Imperative clauses in generative grammar*. Amsterdam: John Benjamins Publishing Company.
- Zoetmulder, P. J & Poedjawijatna. 1993. *Bahasa Parwa: Tatabahasa Jawa Kuno bentuk kata* [Parwa Language: Grammar of Old Javanese word forms]. Yogyakarta: Gajah Mada University Press.

Source materials

Abbreviations for source materials taken from the story corpus

- a. MB = Men Brayut
- b. KP = Keris Penganten
- c. PMB = Penyu Metaluh di Biase
- d. IK = I Kelesih

- Antara, Pande Putu Alit. 2021. *Keris Penganten*. Tabanan: Pustaka Ekspresi.
- Dalu, I Buyut. 2009a. *Penyu Metaluh di Biase*. Denpasar: CV Kayumas Agung
- Dalu, I Buyut. 2009b. *I Kelesih*. Denpasar: CV Kayu Mas Agung
- Dalu, I Buyut. 2013. *Men Brayut*. Denpasar: CV Kayu Mas Agung