Argument Indexing in Kamang

Walker, K.; Faghiri, P.; van Lier, E.

DOI
10.1075/sl.21077.wal
Publication date
2024
Document Version
Final published version
Published in
Studies in Language

Argument indexing in Kamang

Katherine Walker, Pegah Faghiri and Eva van Lier
University of Amsterdam

Kamang (Alor-Pantar, non-Austronesian/Papuan) has a typologically unusual system of argument indexing, in which the S or P argument can be indexed on the verb by one of several prefix paradigms. Some verbs always show indexing, while others exhibit differential argument indexing (DAI). In DAI, the use of a particular prefix paradigm or zero marking depends on different (combinations of) factors. We investigate the effects of argument role (S, P), independent argument realisation, the animacy and topicality of the indexed argument, and lexical stipulation. We perform a quantitative analysis of these factors for the first time in Kamang discourse, drawing on an annotated corpus of spoken Kamang. A complex picture emerges in which Kamang argument indexing is best viewed not as a single system, but as multiple subsystems for which different factors are relevant in a given context, and which do not operate on all verbs or all indexing strategies equally.

Keywords: Kamang, Alor-Pantar, Papuan, person indexing, corpus data, discourse, differential argument marking, usage-based methods, alternation

1. Introduction

Kamang is a non-Austronesian/Papuan language of east Indonesia, belonging to the Alor-Pantar family. It has a typologically unusual system of argument indexing, in which the S or the P argument can be indexed on the verb by one of several prefix paradigms, while A is never indexed.[1] In (1), for example, two different prefixes – from what we call the /a/-series and the /e/-series – are each shown indexing S and P.

1. S refers to the single argument of an intransitive verb; A refers to the more agent-like argument and P the more patient-like argument of a transitive verb.

https://doi.org/10.1075/sl.21077.wal | Published online: 27 June 2023
Studies in Language 48:2 (2024), pp. 287–350. ISSN 0378-4177 | E-ISSN 1569-9978

(1)

		S indexing		P indexing
/a/-series prefix	a.	*Namaitansi.*	b.	*Leon nataksi.*
		***na**-maitan-si*		*Leon **na**-tak-si*
		1sg./a/-hunger-ipfv		Leon 1sg./a/-see-ipfv
		'I'm hungry.'		'Leon sees me.'
/e/-series prefix	c.	*Nelaitasi.*	d.	*Leon nefaneesi.*
		***ne**-laita-si*		*Leon **ne**-fanee-si*
		1sg./e/-shy-ipfv		Leon 1sg./e/-shoot-ipfv
		'I'm shy.'		'Leon shoots at me.'

(Schapper 2014: 324)[2]

In addition, some verbs always show indexing, while others exhibit what we call differential argument indexing (DAI, cf. Iemmolo's (2011) differential object indexing, DOI).[3] For instance, *tak* 'run' appears with a prefix in (2b) and without one in (2a).

(2) a. *kui tak*
dog run
'The dog runs.'
b. *kui ge-tak*
dog 3./e/-run
'The dog ran off (was forced to run).' (Schapper 2014: 326)

In DAI in Kamang, the use of the various prefix paradigms depends on different (combinations of) factors. Some of these factors – animacy, argument role (S, P), volitionality and telicity – were previously investigated in several Alor-Pantar indexing systems, including in Kamang (Fedden et al. 2013; Fedden et al. 2014). Nevertheless, the exact functioning of differential indexing remained open to further exploration.

2. Glosses and orthography of examples from published work are taken from the source, except for Kamang prefixes, which are glossed in accordance with the conventions introduced in this paper. In Kamang orthography, <'> represents glottal stop and two adjacent vowels indicate a phonemically long vowel. References in square brackets of the form [kamang_XXX] refer to texts in the annotated subcorpus "Multi-CAST Kamang" (Schapper et al. in prep.; see Section 3). Otherwise, square brackets refer to fieldnotes provided by Antoinette Schapper and George Saad.

3. We use DAI for differential argument indexing rather than Iemmolo's (2011) term DOI for differential object indexing, because we are concerned with the indexing of S and P arguments. DOI has also been termed "differential object agreement" (Lazard 2005) and DAI has been referred to as "differential argument realisation" in another Alor-Pantar language, Abui (Kratochvíl 2014).

The present paper is a quantitative corpus study that reveals the role of two discourse factors in Kamang DAI; namely, co-occurrence with independent pronouns and topicality. We also expand on previous findings, providing a more nuanced picture of the factors of argument role and animacy. Finally, we confirm the substantial role of lexical stipulation; that is, synchronically arbitrary associations of verbs and prefixes. These findings are interpreted in the broader context of differential argument marking (DAM), in particular in terms of the difference between differential indexing (marking on the verb) versus differential flagging (case/adpositions).

Kamang – like several other Alor-Pantar languages – is typologically unusual for two reasons: first, because it may index the P argument on the verb, to the exclusion of A (only 6% of 378 languages in Siewierska 2013b), and second because it exhibits a form of active alignment in verbal indexing (only 7% of 380 languages in Siewierska 2013a). Furthermore, there are no languages in the WALS datasets (Dryer & Haspelmath 2013) that combine both of these features, yet they co-occur in a number of Alor-Pantar languages (Holton & Robinson 2017: 147; Klamer 2017: 20; Klamer & Kratochvíl 2018). Moreover, indexing in Alor-Pantar languages is often at least partially differential. While there is much literature on DAM (see, e.g., Seržant & Witzlack-Makarevich 2018), this is mostly about flagging (Haspelmath 2021: 131). Much less research focuses on differential object indexing (exceptions include Iemmolo's work, along with Coghill 2014, Virtanen 2014, Just & Witzlack-Makarevich 2022, Just & Čéplö 2022). Especially rare are studies on differential indexing of arguments other than the object (exceptions include Kratochvíl 2014 on the Alor-Pantar language Abui, and Grossman 2015 on Coptic).

This paper provides a novel angle on the question of indexation in Kamang through our use of quantitative corpus methods. By investigating indexing on Kamang verbs in discourse for the first time, we reveal the impact of two factors not previously considered. First, we show that the discourse factor of topicality has opposite effects on different prefix paradigms. Second, there is a statistical preference for independent pronouns not to co-occur in the clause with prefixed as opposed to zero-marked verbs. While some Alor-Pantar languages categorically prohibit the co-occurrence of pronouns and indexes, no such tendency has been described before for Kamang. In general, we find that lack of prefixation is the default strategy for a wide range of verbs.

Our method also allows us to verify whether the findings reported by Fedden et al. (2013) primarily from video elicitation experiments hold for language in a discourse context.[4] Thus, in addition to providing an enhanced description of (differential) argument indexing in Kamang, we aim to demonstrate the value of applying multiple methodologies in pursuit of answering the same research question.

More generally, we aim to make a microtypological contribution in terms of the diversity of indexing systems in Alor-Pantar languages, and a macrotypological one in terms of a cross-linguistically rare, partially differential S/P indexing system. In particular, we find evidence to support the claim that differential indexing has different functional motivations than differential flagging (Iemmolo 2011). This, in turn, follows from the functional difference between (non-differential) flagging, which performs both a distinguishing and an identifying function (Witzlack-Makarevich & Seržant 2018: 30), and indexing, which has only an identifying function, indexing properties of the referent and tracking it in the discourse (Siewierska & Bakker 2008: 293). As mentioned, we also highlight the considerable role of lexical stipulation: while language description and comparison typically assume lexically specific constructions to be the exception, rather than the rule, they are found throughout Kamang's verbal lexicon.

The paper is structured as follows: In Section 2, we provide the relevant features of Kamang grammar and the argument indexing system. In Section 3 we describe the data used for this paper and the methodology. Section 4 contains an overview of our results for the factors involved in Kamang indexing, including lexical stipulation. In Section 5, we provide a summary of the differential behaviour of each indexing strategy, before rounding off with a conclusion in Section 6.

2. Language background

Kamang (ISO 639-3 code: woi) is spoken by approximately 6,000 people on the island of Alor, eastern Indonesia (Schapper 2014: 286). Like many indigenous languages of the region, it is severely endangered. Published work on Kamang begins in the 1970s (then called Woisika, see Stokhof 1975) and now extends to a preliminary dictionary (Schapper & Manimau 2011) and its inclusion in a video elicitation experiment run with several Alor-Pantar languages (Fedden et al. 2013;

4. Fedden et al. (2013) supplemented the findings from the video elicitation experiment with corpus searches; Fedden et al. (2014) drew on the same experimental data, alongside corpus and elicited data. Some of the findings in the two papers were thus at least partially based on data from naturalistic discourse, but discourse context was not treated as a factor.

see also Fedden et al. 2014, Fedden & Brown 2017). Schapper's (2014) grammar sketch is the fullest treatment of Kamang grammar to date, to which the reader is referred for a more complete picture of the language situation, phonology and morphosyntax. In this section, we summarise only points pertinent to the present study.

The basic constituent order is APV (agent – patient – verb). Occasionally, PAV is attested for topicalised P arguments, but post-verbal arguments are rare. Importantly, any of the basic constituents (A, S or P arguments) can be elided, so that clauses are often realised as AV, PV or simply V. The template for the Kamang verb is given in (3):

(3) SBEN$_{S/A}$-INCORP-INDEX$_{S/P}$-V-ASPECT (based on Schapper 2014: 321)

The prefixes that occupy the INDEX slot on the Kamang verb index person (first, second, third and 'common', used for generic person and reciprocal relations), number for first and second person (singular/plural) and clusivity (inclusive/exclusive for first-person plural) of P or S. There are seven paradigms, given in Table 1. The table provides both the arbitrary Roman numerals used in Fedden et al. (2013) to distinguish the prefix series, and the semantic labels proposed by Schapper (2014). In this paper, we consider only the first three series, due to the relative rarity of the other four, both in our corpus (grouped together as "other" in the results) and in previous research. We name these prefix series after the theme vowel in the singular forms, as shown in Table 2. This convention aids identification without suggesting coherent semantics that do not obtain. We also apply this convention in glossing; for instance, the first-person singular form in the /a/-series, *na*-, is glossed '1SG./a/-'.

Table 1. Kamang prefix paradigms (see Table 12, Schapper 2014: 322)[5]

Fedden et al. 2013	**I**	**II**	**III**	**IV**	**V**	**VI**	**n/a***
Schapper 2014	PATIENTIVE	LOCATIVE	GENITIVE	DATIVE	DIRECTIVE	ASSISTIVE	SELF-BENEFACTIVE (SBEN)
1SG	*na-*	*no-*	*ne-*	*nee-*	*nao-*	*noo-*	*ne'-*
2SG	*a-*	*o-*	*e-*	*ee-*	*ao-*	*oo-*	*e'-*
3	*ga-*	*wo-*	*ge-*	*gee-*	*gao-*	*woo-*	*ge'-*
CMN	*ta-*	*to-*	*te-*	*tee-*	*tao-*	*too-*	*te'-*
1PL.EXCL	*ni-*	*nio-*	*ni-*	*nii-*	*nioo-*	*nioo-*	*ni'-*
1PL.INCL	*si-*	*sio-*	*si-*	*sii-*	*sioo-*	*sioo-*	*si'-*
2PL	*i-*	*io-*	*i-*	*ii-*	*ioo-*	*ioo-*	*i'-*

* Fedden et al. (2013) and Fedden et al. (2014) exclude the self-benefactive (SBEN) series, since it primarily indexes S or A (Schapper 2014: 331) and occupies a different slot in the template (see example 3). However, since a small number of verbs, e.g., *baa* 'say to', do index S or P with SBEN, we include it in the study – it is recorded as "other" in the results section along with the other very infrequent prefix series.

Table 2. Kamang prefix paradigms considered in this paper

	/a/-series	**/o/-series**	**/e/-series**
1SG	*na-*	*no-*	*ne-*
2SG	*a-*	*o-*	*e-*
3	*ga-*	*wo-*	*ge-*
CMN	*ta-*	*to-*	*te-*
1PL.EXCL	*ni-*	*nio-*	*ni-*
1PL.INCL	*si-*	*sio-*	*si-*
2PL	*i-*	*io-*	*i-*

Kamang verbs can be divided into those that exhibit DAI – that is, alternating verbs – and those that do not. The latter – non-alternating verbs – always appear with a particular prefix series, or always occur without a prefix. For example, the

5. Differentiating the /e/-series from 'DATIVE' and 'SBEN' is difficult given the lack of glottal stop in the dialect of one speaker and frequent difficulties in discerning vowel length for all speakers (see Al-Gariri unpublished ms). Prefix glosses are primarily based on native-speaker intuition.

non-alternating verb *tak* 'see' always takes a fixed /a/-series prefix to index the P argument, regardless of discourse context or properties of the argument. This is shown in (4a)–(b), where an animate and an inanimate P are indexed in the same way.

(4) a. *Leon nataksi.*
Leon na-tak-si
Leon 1SG./a/-see-IPFV
'Leon sees me.' (Schapper 2014: 254)

b. *gekere gataksi naa*
ge-kere ga-tak-si naa
3./e/-shirt 3./a/-see-IPFV NEG
'[he] didn't see his shirt' [kamang_piee]

For expository purposes, we refer to the empty slot of verbs without a prefix as 'zero' and gloss it with 'Ø-'. Example (5) shows a non-alternating intransitive verb *mu'tan* 'fall from above' with a fixed zero.

(5) *dum kiding lamibee mu'tan*
dum kiding lami=bee Ø-mu'tan
child small male=also Ø-fall_from_above
'the little boy also fell' [kamang_frog]

We treat all monovalent verbs as having an S argument and all bivalent verbs as having an A and a P argument, regardless of semantic role.[6] This follows the annotation scheme applied to our corpus (Haig & Schnell 2014; see Section 3) and leads to labelling certain arguments 'P' that would not be treated as such elsewhere. For instance, the verb *n* 'give' is a non-alternating verb with a fixed /e/-series prefix that indexes the RECIPIENT role, as in (6). Since the THEME is flagged with the light verb *me* 'LV', we treat the first-person RECIPIENT as P for the purposes of this study.

(6) *…nuaa me nen*
nuaa me ne-n
things LV 1SG./e/-give
'[he] gave me things' [kamang_feri]

In contrast to Examples (4)–(6), alternating verbs are attested with differential indexing, either (i) with zero and one or more prefix series, or (ii) with two or more different prefix series. Example (7) shows the alternating verb *faafa* 'search for', which exhibits DAI based on animacy of the P argument. In (7a), *faafa* is

6. There are no true trivalent verbs (see Klamer & Schapper 2012).

marked with an /a/-series prefix when the P argument is human, and in (7b) is zero-marked when the P is non-human. Note, however, that animacy-based DAI represents a minority pattern in Kamang: it was previously found to be unproductive (Fedden et al. 2014: 67), and only two verbs in our corpus alternate in this way (see Section 4.2.2).

(7) a. *…geduma gafaafa*
ge-dum=a ga-faafa
3./e/-child=SPEC 3./a/-search_for
'…[she] kept looking for the child' [kamang_saiko]
b. *…taweng tebini faafa*
taweng te-bini Ø-faafa
in_turns CMN./e/-lice Ø-search_for
'…[they] search for each other's lice' [Schapper fieldnotes]

For some alternating labile verbs, different indexation strategies correspond with different valencies, as shown in (8). In (8a), *laka* 'hang' is a monovalent verb with zero marking; in (8b) there are two arguments, and *laka* takes an /o/-series prefix that indexes the (location) P argument 'the tree'.[7]

(8) a. *laawang talewuibo laka…*
laawang talewui=bo Ø-laka
bee hive=REL Ø-hang
'the beehive that hangs…'
b. *…me koo bongak wolakasi.*
me koo bong=ak wo-laka-si
come stay tree=DEF 3./o/-hang-IPFV
'…and [the bees] are hanging in the tree.' [kamang_frog]

However, valency change is not always overtly coded in this way. A non-alternating labile verb *oo* 'be born/give birth to' is shown in (9): two adjacent utterances show the monovalent use (in Example 9a) and bivalent use (in Example 9b), each marked with a fixed /a/-series prefix.

(9) a. *watuubo naooa jumat mi*
watuu=bo na-oo=a jumat mi
day=REL 1SG./a/-be_born=SPEC friday in
'The day on which I was born was Friday.'

7. Alternatively, one could analyse this as two different verbs: a monovalent one with fixed zero indexing, and a bivalent one with fixed /o/-series indexing. In this study, however, we consider the two forms to belong to the same verb lemma.

b. *kuul yaa tung suiha dum male nok noukolee nepaa gaoomaih me nala ninnok*
kuul yaa tung su-ih=a dum male nok n-ouko-lee
again reach year three-CSEQ=SPEC child female one 1SG./a/-mother-ASSOC
ne-paa ga-oo-ma-ih me nal=a ninnok
1SG./e/-father 3./a/-give_birth-PFV-CSEQ[8] come 1SG=SPEC 1DU.EXCL
'Until three years later, my parents gave birth to a baby girl so that there were two of us.' [Saad fieldnotes]

Finally, a note is in order about the INCORP(ORATED) prefix slot (see Example 1 above), which hosts one of two, non-inflecting, polyfunctional prefixes *wo-* 'AT' and *mi-* 'IN' (Schapper 2014: 327). These prefixes have valency-increasing and/or semantic effects (Schapper 2020a: 32; for other Alor-Pantar languages see Willemsen 2017). The prefix *wo-* 'AT' is homophonous with the third-person form of the /o/-series, can increase valency like the /o/-series, and is reported not to co-occur with any /o/-series prefix, though it may co-occur with other prefixes (Schapper 2014: 321). Full treatment of INCORP *wo-* 'AT' and INDEX *wo-* '3./o/' is beyond the scope of this paper; here we just note the potential ambiguity of examples in which *wo-* directly precedes the verb root. For instance, the valency of the verb *laka* 'hang' (see Example 8 above) is increased by the prefix *wo-*. With no intervening prefix in the INDEX slot, the form could conceivably be either INCORP *wo-* or INDEX *wo-*. There is no ambiguity when the index slot is filled with another series, illustrated in (10) for the verb *tfa* 'shoot'. In (10a), P is already indexed with an /a/-series prefix; in (10b), the /a/-series prefix occurs together with *wo-*, which here indicates a repeated action (Schapper 2014: 330–331). In this paper, we consider the form *wo-* to be an /o/-series prefix when prefixed directly to the verb root, as in the case of *wo-laka* in (8b). If it precedes another prefix series, as in *wo-ga-tfa* in (10b), we consider it INCORP *wo-* 'AT'.[9]

(10) a. *Na Seb gatfa ipaa.*
na Seb ga-tfa ipaa
1SG.AGT Seb 3./a/-shoot dead
'I shot Seb dead.'

8. The dependency marker *-ih* indicates consequential relations between clauses and is not described as impacting valency (Schapper 2014: 336–337).

9. There is also no ambiguity in the case of first and second person forms of the /o/-series (see Table 2). Because these forms occur only twice in our corpus, it was not possible to exclude ambiguous cases and still include the /o/-series in the analysis.

b. *Na Seb wogatfa ipaa.*
na *Seb wo-ga-tfa* *ipaa*
1SG.AGT Seb AT-3./a/-shoot dead
'I shot Seb again so that he was dead.' (Schapper 2014: 330–331)

3. Data and methods

The present study offers a quantitative corpus analysis of argument indexing in Kamang discourse. This differs from previous research, which compared five Alor-Pantar languages (including Kamang) on the basis of a video stimuli elicitation experiment complemented by insights from corpus data (Fedden et al. 2013; Fedden et al. 2014). The video elicitation stimuli experiment permitted particular verbs to be targeted, controlling for the argument properties animacy and argument role (S or P), and the predicate properties volitionality, telicity and dynamicity. The findings for Kamang were primarily statistical tendencies rather than categorical splits, and much variation remained due to lexical stipulation. In order to investigate some of these statistical and lexical tendencies further, our study targets indexing in a more-or-less natural discourse context.

The corpus for the present study – "Multi-CAST Kamang" (Schapper et al. in prep.) – consists of a selection of spoken-language texts drawn from a larger corpus gathered between 2010 and 2016 by Antoinette Schapper and in 2020 by George Saad. This selection was then annotated with the GRAID (Grammatical Relations and Animacy in Discourse) annotation scheme (Haig & Schnell 2014) in ELAN (2020).[10] GRAID annotations consist of a number of standard tags, which were then modified to the specific structure of Kamang (Schapper et al. in prep.). Additional annotations were added following the RefIND guidelines (Schiborr et al. 2018) and a simplified version of the RefLex annotation scheme was applied (Baumann & Riester 2012; Riester & Baumann 2017).

The annotated texts were selected on the basis of the Multi-CAST (Haig & Schnell 2022) criteria to allow maximum data comparability between corpora; namely, monologic, narrative texts, including personal histories, traditional narratives and elicited narratives. Given the already narrow range of text types and the relatively small size of the corpus, we have not controlled for genre in the analysis.

The annotated corpus consists of ten texts, totalling around one hour and 1,123 clauses. Of these clauses, 828 contain verbal predicates. The texts were produced by three male speakers, aged 44–86. Eight texts are from a single speaker of the Lowland dialect, but were recorded ten years apart. The final two texts are

10. See, for example, Brugman & Russel (2004).

from a speaker of the Upland dialect and a speaker whose dialect is not recorded (for information on dialect, see Schapper 2014; Schapper & Manimau 2011). Annotations in ELAN were then extracted and statistical analyses performed in R (R Core Team 2019).

The annotation schema allows us to perform a quantitative, discourse-based analysis on (1) argument role (Section 4.2.1) and (2) animacy (Section 4.2.2), which are factors previously reported as playing a role in Kamang indexing. It also allows us to consider (3) co-occurrence constraints (with independent pronominal and/or nominal arguments; Section 4.3.1), a factor relevant in neighbouring Alor-Pantar languages but not previously reported for Kamang. Finally, we are able to assess another discourse-related factor, which was reported in Iemmolo's cross-linguistic treatment of DOI: (4) topicality (Section 4.3.2). Note that our corpus annotations did not allow quantitative analysis of volitionality, telicity and affectedness, which were previously found to play a role in particular alternations (Fedden et al. 2013; Fedden et al. 2014; Schapper 2014). A qualitative analysis of these factors is left for future research (see Walker accepted).

In order to achieve a more fine-grained picture of differential indexing in Kamang, we performed multifactorial (logistic) regression modelling on a subset of the data that included only alternating verbs.[11] Multifactorial regression modelling allows us to go beyond studying the effects of different factors in isolation by highlighting their respective contribution. Importantly, some factors may be correlated; that is, the value of a given factor may determine the effect of another to some extent. In addition, this modelling allows us to take verbal bias into account. This is crucial, because the factors we consider capture only some of the variation in the data and account for alternation to different extents, leaving a considerable amount of variation unaccounted for (see Section 4.4). Including verb types, using verbal lemma as a grouping factor and calculating varying intercepts improves the accuracy of the models, suggesting that alternating verbs have different indexing preferences.

Hence, we performed two types of analysis: fixed-effects and mixed-effects modelling. In Section 4.2 and Section 4.3, where we study the general effects of the four factors on differential indexing, we report the results of our fixed-effect analyses. These were performed on the alternating subset of the dataset as a whole, and without considering verb types. We fitted a multinomial model – with a 4-level response variable, taking zero indexing (the most frequent option) as the

11. This classification may not be robust due to the nature of our data; that is, verbs counted as non-alternating may turn out to alternate in a larger (or different) dataset. It is possible that, due to the limited size of the corpus, a fixed indexing strategy in our dataset merely reflects low productivity of a differential strategy; see Section 4.1.1.

baseline – using the nnet package (Venables & Ripley 2002), and four binomial models, each with a binary response variable that coded absence (failure/0) vs. presence (success/1) of a given indexing strategy.

Next, for each strategy we considered a more restricted subset of data, keeping only those verbs whose alternation profile (see Table 7 in Section 4.1.1) contained a match to the given strategy. This subset was used to build fixed-effects models, as well as mixed-effects models with verbal lemma as random intercepts, using the lme4 package (Bates et al. 2015).[12] By comparing the results of different models – which include verb type to different extents – we aim to highlight the role of lexical stipulation in differential indexing (Section 4.4). For each model, we started by building a model that included all main effects and interactions justified by the data and used likelihood ratio tests to eliminate those that did not significantly contribute to the fit. In the following section, we will limit statistical details and only report relevant information on the effect's size and significance. Full model summaries are provided in Appendix A.

4. Results

In this section, we describe the distribution of different indexing strategies in the corpus in terms of the four factors listed in Section 3. We initially consider both fixed and differential strategies, before turning our focus exclusively to the effects of each factor on DAI. Fixed and differential marking strategies are analysed separately because only the latter can be sensitive to context. For instance, if the /a/-series is sensitive to the factor of animacy, only differential uses of the /a/-series can demonstrate the relevance of this factor. However, we also analyse fixed strategies and compare the two, because fixed and differential instances may show the same tendencies. First, because at least some instances only appear fixed due to the limited size of our corpus. And second, because fixed and differential indexing may have the same diachronic source (see Klamer & Kratochvíl 2018: 75 on indexing in Proto-Alor-Pantar). That is, frequency effects may have led to the same prefix series being synchronically fixed on one verb and used differentially on another. Given the synchronic lack of context sensitivity, however, the effects are likely to be weaker or null for fixed occurrences.

12. It is common in corpus linguistics to include lemma as random effect in a mixed-effects regression model (see, e.g., Schäfer 2020).

4.1 Overview of the dataset

Our corpus features 164 verb types, of which 39 (23.8%) are alternating verbs and 125 verbs (76.2%) do not alternate. However, we cannot assume that the latter verbs, especially those with low token counts, do indeed belong to the category of non-alternating verbs. For instance, *suusa* 'be in difficulty' is attested only once in the corpus – with the /o/-series – and is thus classed as non-alternating. Yet, Fedden et al. (2014: 65) report that *suusa* is attested with both the /o/-series and zero.[13] Indeed, given the nature of the data, the "non-alternating" label is less informative for low-frequency verbs and not meaningful at all for verbs that only appear once in our corpus.

Table 3 shows the proportions in our corpus of alternating and non-alternating verbs, with the latter making up more than three quarters of the total.

Table 3. Proportion of alternating and non-alternating verb types (% to 1dp)

Verb types	n=	percent
Alternating	39	23.8%
Non-alternating	125	76.2%
Total	**164**	**100%**

Fedden et al. (2014) use different categories: obligatorily prefixed verbs and non-obligatorily prefixed verbs. The categories in this obligatoriness-based system differ from the alternating/non-alternating categories used here in the treatment of zero marking. The category "non-obligatorily prefixed" includes verbs attested with zero marking, under the assumption that they always have the potential to alternate (Fedden et al. 2014: 59–60). We did not follow this classification to avoid the inconsistency of treating a high-frequency verb attested only with zero marking as non-obligatorily prefixed, while a low-frequency verb with a fixed prefix series is treated as obligatory prefixed. The point of difference is illustrated in Table 4: in Fedden and colleagues' "obligatoriness" classification system, *bun* 'hide' is grouped together with *maa* 'walk' as non-obligatorily prefixed verbs, since both occur in the corpus at least once with zero marking. In the "alternation" classification system, *maa* 'walk' is instead grouped together with *kut* 'stab' as non-alternating verbs, since they always occur only with the same (lack of) prefix.

13. This is the only low-frequency verb in our dataset for which information from previous work suggests a different categorisation than given here.

Table 4. Illustration of different classification systems

Verb	Count with prefix	Count with zero	Classification based on obligatoriness	Classification based on alternation
bun 'hide'	/a/-series: 8 /o/-series: 1	2	non-obligatorily prefixed	alternating
maa 'walk'	–	19	non-obligatorily prefixed	non-alternating
kut 'stab'	/a/-series: 1	–	obligatorily prefixed	non-alternating

Table 5 compares the proportions of obligatorily prefixed/non-obligatorily prefixed in our data to the findings in Fedden et al. (2014). It shows that our corpus has a slightly higher proportion of obligatorily prefixed verbs.

Table 5. Proportion of (non-)obligatorily prefixed verbs in our corpus compared to Fedden et al. 2014 (% to 1dp. Fedden et al. 2014: % are approx.)

	Our corpus		Fedden et al. 2014	
Verb type	**n=**	**percent**	**n=**	**Percent**
Non-obligatorily prefixed verbs	102	62.2%	344	67%
Obligatorily prefixed verbs	62	38.8%	166	33%
Total	**164**		**510**	

4.1.1 Distribution of marking strategies

The distribution of marking strategies for the 125 non-alternating verb types is shown in Table 6. Half of non-alternating verbs are attested with zero marking. For verbs that are always attested with a prefix, the most frequent is the /a/-series (22.4%), followed by the /o/-series (16.8%) and /e/-series (8.8%). The four infrequent prefix series (see Table 1), reported together as "other", make up the remaining 1.6%. For the 39 alternating verb types, 14 alternation profiles are attested. These are listed in order of descending frequency in Table 7, showing that few profiles are attested more than once.

The bottom two rows of Table 7 show total type frequency: a verb type is counted in a given column if it is attested at least once with the respective marking strategy. The figures reveal that zero marking occurs on 82.1% (32/39) of alternating verbs. This shows that the general pattern for alternating verbs is to alternate between zero and one or two prefix series.

Table 6. Marking strategies used with non-alternating verbs (% to 1dp)

	Proportion of all verb types	
Marking strategy	**n=**	**percent**
Other	2	1.6%
/e/-series	11	8.8%
/o/-series	21	16.8%
/a/-series	28	22.4%
Zero	63	50.4%
Total	125	–

Table 7. Total alternating verb types and alternation profiles (% to 1dp). x = strategy is part of the profile

	Strategy					
Profile	**zero**	**/e/**	**/o/**	**/a/**	**other**	**Total verb types**
/e/+zero	x	x				12
/o/+zero	x		x			10
/a/+zero	x			x		5
/o/+zero+other	x		x		x	2
/e/+other		x			x	2
/e+/a/		x		x		1
/o/+/a/+zero	x		x	x		1
/e/+/o/+/zero+other	x	x	x		x	1
zero+other	x				x	1
/o/+/a/			x	x		1
/e/+/o/		x	x			1
other+other					x	1
/o/+other			x		x	1
Total type frequency	32	17	17	8	9	39
%	82.1	43.6	43.6	20.5	28.1	

Table 8 shows the distribution of marking strategies across the 828 verb tokens in the corpus. More than half (54.1%) are attested with no prefix ("zero"), with significantly fewer occurrences of the /a/-series (17.3%), /o/-series (13.0%) and /e/-series (11.8%). Again, all other series, grouped together as "other", are very infrequent. In the following, then, we comment only on the four most frequently used strategies: /e/-series, /a/-series, /o/-series and zero, resulting in a dataset of 797 tokens.

Table 8. Distribution of marking strategies for verb tokens (% to 1dp)

Marking strategy	Proportion of all verb tokens		Proportions of most common marking strategies
	n=	**percent**	**percent**
other	31	3.7%	–
/e/-series	98	11.8%	12.3%
/o/-series	108	13.0%	13.6%
/a/-series	143	17.3%	18.0%
Zero	448	54.1%	56.2%
Total	**828**	**100%**	**100%, *n*=797**

Shown in Table 9, the overall proportion of differential marking is 44.2%. The /e/-series has the greatest proportion of differential occurrences (62.2%), while the /a/-series has the lowest (18.2%).

Table 9. Proportion of differential occurrences (% to 1dp)

Marking strategy	Frequency: differential/total	Proportion of differential marking
	n=	**%**
/e/-series	61/98	62.2%
/o/-series	49/108	45.4%
/a/-series	26/143	18.2%
Zero	220/448	49.1%
Total	**356/797**	**44.2%**

Since only 23.8% of verb types alternate (see Table 3 above) but make up 44.2% of tokens, alternating verbs appear to be more frequent. Indeed, the mean token frequency for alternating verbs is 9.7, compared to 5.7 for non-alternating verbs (excluding verbs attested once only). We should of course be careful about generalising from this observation, given that higher-frequency verbs have a greater chance of exhibiting alternation in our dataset.

Table 10 shows mean token frequencies for verb and marking strategy combinations; that is, how many times verbs in each class are attested with a particular marking strategy. Alternating verbs are most frequently attested with zero marking, with a mean frequency of 6.9, just under double the next highest of 3.6 for differential /e/-series. Hence, not only is zero the most common strategy in alternation profiles (82.1% of alternating verbs have zero as a marking option; bottom row of Table 7), it is also the most frequently used in discourse. Non-alternating verbs do not show the same patterns: rather than zero, the /a/-series has the highest mean token frequency. Although we assume that some of these non-alternating verbs might alternate in a larger corpus, comparing the mean frequencies in Table 10 indicates the important presence of zero in alternation.

Table 10. Mean token frequencies by verb type (alternating/non-alternating) by indexing strategy (n≥1; to 1dp)

Indexing strategy	Alternating *(n≥1)*	Non-alternating *(n≥1)*
/e/-series	3.6	3.5
/o/-series	2.9	2.8
/a/-series	3.3	4.2
Zero	6.9	3.6

4.1.2 Valency classes

Table 11 shows the valency class of non-alternating verb types; that is, whether the marking strategy for a particular verb indexes S, P or both (i.e., they are labile verbs, listed under "S/P" in the table). There is a slight overall preference for P-marking (56.8%), but only the /o/-series marks P exclusively. Other strategies can be used to mark either S or P.

Table 11. Marking strategies and valency class for non-alternating verbs (% to 1dp)

	S		P		S/P		
Marking strategy	**n=**	**percent**	**n=**	**percent**	**n=**	**percent**	**Valency class**
Other	1	50%	1	50%	–	–	S, P
/e/-series	5	45.5%	6	54.5%	–	–	S, P
/o/-series	–	–	21	100%	–	–	P
/a/-series	10	35.7%	15	53.6%	3	10.7%	S, P, S/P
Zero	34	54.0%	28	44.4%	1	1.6%	S, P, S/P
Total	**50**	**40.0%**	**71**	**56.8%**	**4**	**3.2%**	–

Valency classes for alternating verbs are indicated in Table 12. Unlike for non-alternating verbs, there is no preference for P-marking, but instead a light preference for S marking. If alternation were conditioned by valency change, we would expect high numbers of verbs marking both S and P. There are indeed more labile verbs (indicated by Valency "S/P") among alternating verb types (10.3%) than non-alternating types (3.2%, see Table 11). However, this leaves almost 90% of alternating types as not, or not exclusively, valency changing.

Table 12. Indexing profile and valency class for alternating verbs (% to 1dp)

		Valency			
Profile	**Total verb types**	**S**	**P**	**S/P**	**Class**
/e/+zero	12	9	1	2	S, P, S/P
/o/+zero	10	3	5	2	S, P, S/P
/a/+zero	5	1	4	–	S, P
/o/+zero+other	2	1	1	–	S, P
/e/+other	2	1	1	–	S, P
/e+/a/	1	1	–	–	S
/o/+/a/+zero	1	1	–	–	S
/e/+/o/+/zero+other	1	1	–	–	S
zero+other	1	1	–	–	S
/o/+/a/	1	–	1	–	P
/e/+/o/	1	–	1	–	P
other+other	1	–	1	–	P
/o/+other	1	–	1	–	P
Frequency	39	19	16	4	
		48.7%	41.0%	10.3%	

4.2 Argument role and animacy

The argument role of the referent involved in the event described by the verb (S or P) and its animacy are two factors previously investigated in Kamang indexing. In the following we discuss the distribution of these two factors and their interaction, and compare our results with previous findings.

4.2.1 Argument role

Figure 1 compares the distribution of zero marking versus indexation with a prefix for S and P arguments. There is a clear difference, in that S arguments are more likely to be zero-marked, while P arguments are more likely to be indexed with a prefix. The horizontal dashed line shows the proportions for the dataset as a whole: 43.8% of arguments are indexed with a prefix.

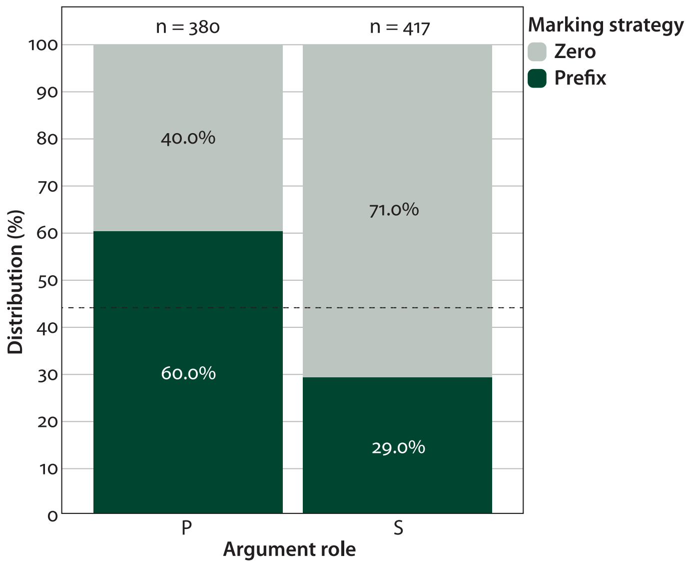


Figure 1. Distribution of zero marking and prefixes by argument role (% to 1dp, $n = 797$)

In comparing the distribution of S and P arguments for each marking strategy separately, we find that different prefix series show different trends. Figure 2 shows the distribution of S and P arguments indexed by the most common marking strategies, with differential instances on the left-hand side and fixed instances on the right. For the dataset as a whole, 47.7% of marking strategies index P (below the dashed line) and 52.3% index S. Only the /o/-series is exclusively used for one argument role (P), while all others can mark both S and P. For differential occurrences, zero marking is more frequently used with S, while the /a/-series more often indexes P. Fixed /a/-series marking, however, is almost 50:50.

The /e/-series is the only strategy that substantially differs between differential and fixed markers. While fixed tokens of the /e/-series index a high proportion of P arguments, differential /e/-series occurs more frequently with S. (Although we note that our multifactorial analyses did not show an effect of argument role for differential /e/-series, which is due to the large effect of animacy and the correlation between argument role and animacy; see Section 4.2.2.) This could hint at a valency-decreasing role of differential /e/-series; however, the high proportion of P indexing in fixed /e/-series markers simply reflects the overrepresentation in the corpus of a single verb, *n* 'give; marry', shown in (11a)–(b). This verb always

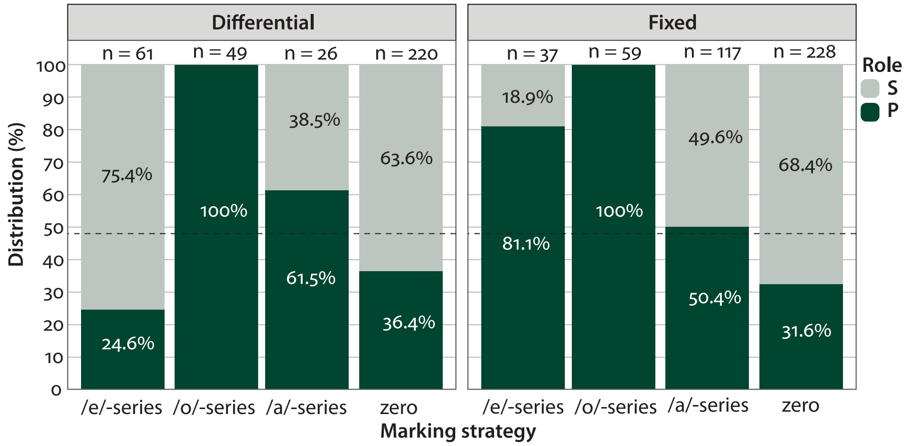


Figure 2. Distribution of indexed argument roles (% to 1dp, *n* = 797)

indexes P (semantically the RECIPIENT; see Section 2) and accounts for 64.9% (18/37) of fixed /e/-series tokens (48.6% for 'give', 16.2% for 'marry'). However, the high proportion of S among differential /e/-series prefixes cannot be explained by a high token count for a single verb.

(11) a. *…nuaa me nen*
nuaa me ne-n
things LV 1SG./e/-give
'[he] gave me things' [kamang_feri]

b. *ining isingok pa maau geeng gamaaung lutei piee gensi?*
i-ning isingnok=apa maau geeng ga-maaung lutei piee
2PL./A/-CLF:HUM seven=this who 3.FOC 3./a/-want young crocodile
ge-n-si
3./e/-marry-IPFV
"You seven! Who wants to marry this young crocodile?" [kamang_piee]

For the /o/-series, the fact that only P is indexed suggests that it may serve a valency-increasing function. This is sometimes the case (e.g., *laka* 'hang', *wo-laka* 'hang on something' in Example 8 above), but there are also examples in which differential /o/-series prefixes do not increase valency. For instance, in (12), *lakai* 'take off' is bivalent in both (12a) and (12b), but has zero marking in the former and an /o/-series prefix in the latter.

(12) a. *alma ang gera wosol gefalaka lakai.*
alma ang gera wosol ge-falak=a Ø-lakai
human DIST.NKNWN 3.CONTR begin 3./e/-cloth=SPEC Ø-take_off
'the person straightaway took off his cloth.' [kamang_sun]

b. *Lami ang sue gekul pang wolakaisak…*
lami ang sue ge-kul pang wo-lakai-sa=ak
man DIST.NKN arrive.IPFV 3./e/-skin DIST.KNWN 3./o/-take_off-COMP=DEF
'The man came and took off his skin…' [kamang_piee]

4.2.2 Animacy

Animacy is frequently a factor – though rarely the only factor – in differential object marking cross-linguistically, but arises much less frequently in differential subject marking (Fauconnier 2011). Animacy is a relevant factor in the indexing systems of other Alor-Pantar languages; for instance, in Western Pantar certain verbs allow only animates to be expressed by an independent pronoun and a prefix in the same clause (Holton 2014: 77). Animacy is also an important factor in the indexing system of Teiwa, where prefixes only rarely index inanimates (Klamer 2010: Sec. 3.3.2). In previous work on Kamang, animacy is relevant to a "handful" of verbs that typically index animate P arguments with a prefix and inanimates with zero (Fedden et al. 2014: 67). In our corpus, two verbs show this pattern (where animate P is indexed with the /a/-series): *buh* 'lift; cradle' and *faafa* 'look for'. As described in the following, however, animacy interacts with argument role rather than being the sole factor that conditions alternation.

There is high correlation between argument role and animacy in our data. As illustrated in Figure 3, P is associated with inanimate referents, in line with observations on other languages (see Brickell & Schnell 2017). S, on the other hand, is strongly correlated with animate referents. As such, we report the frequency distribution for animacy for S and P separately.

Since we fitted the multifactorial models to the whole dataset (i.e., not split by argument role), we do not report this analysis in this section. Relevant here is that, first, both argument role and animacy contribute significantly to the fit, but they have different effects for different strategies (see Appendix A). This can be observed in the frequency data shown in Figure 2 above for argument role, and in Figure 5 and Figure 6 below for animacy. Second, their interaction is not significant. This is visible in Figure 4, which shows that the effect of animacy on differential indexing strategies does not depend on argument role. That is, animate arguments are more likely than inanimates to be indexed with a prefix regardless of the role. More precisely, the animacy-based preference (for prefixing versus zero marking) is the same for P and S (i.e., there is no significant difference ($\chi2$ (1) = 0.17809, p = 0.67)).

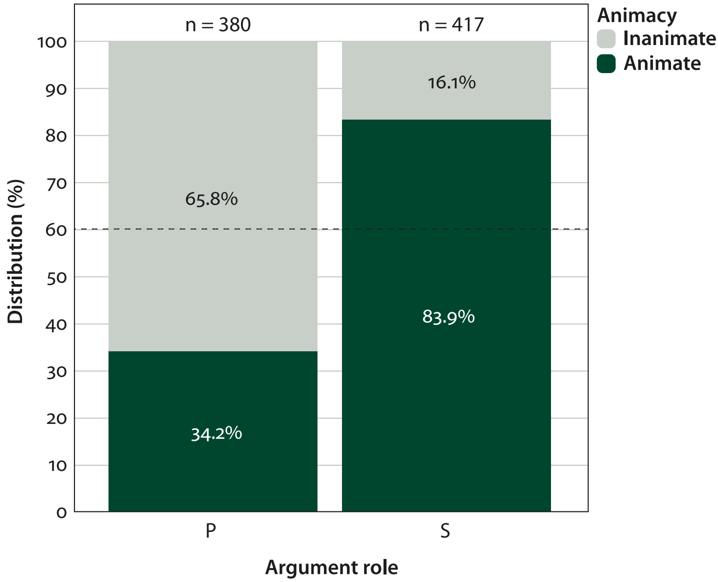


Figure 3. Distribution of animate arguments by argument role (% to 1dp, $n=797$)

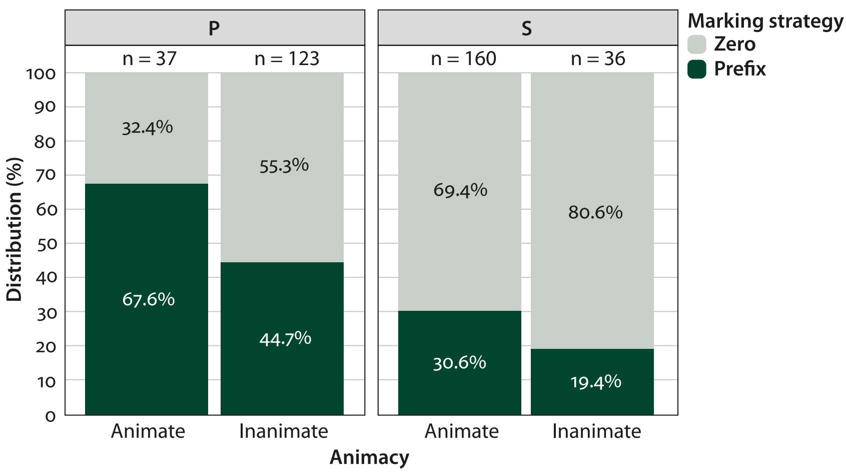


Figure 4. Distribution of differential zero marking and prefixes by animacy and argument role (% to 1dp, $n=356$)

For P arguments, the results from the corpus are shown in Figure 5.[14] The dashed line indicates the overall proportion of animate P in the corpus. The /e/- and /a/-series show a preference for animate arguments, with differential /a/-series showing the strongest preference. Zero marking – differential and fixed – shows a preference for inanimates, as does the /o/-series. In fact, differential /o/-series has the lowest proportions of animates of any strategy. The /o/-series is the only strategy to show a statistically significant difference ($\chi 2$ (1) = 8.6656, $p < 0.01$) between fixed and differential occurrences. Only 6.1% of differential occurrences of the /o/-series index animate referents, compared with 30.5% of fixed occurrences.

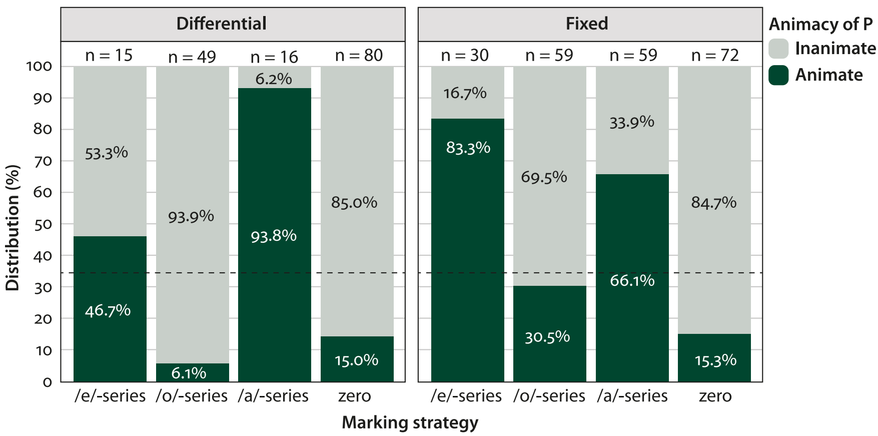


Figure 5. Distribution of P arguments by animacy (% to 1dp, $n = 380$)

In a frequency-based explanation of asymmetric (i.e., overt vs. zero; Witzlack-Makarevich & Seržant 2018) DAM, unusual combinations of role and animacy are expected to be overtly marked (see, e.g., Haspelmath 2006). In Kamang, this means we would expect animate P to be indexed with a prefix. The expected results are found for two prefixes, /e/-series and /a/-series, which index animate P with above-average frequency. The /o/-series, however, shows the opposite behaviour. In Kamang, what appears to be purely animacy-conditioned DAI of P is observed for a small number of alternating verbs, with a split in argument indexing between (higher) animates and lower animates/inanimates (e.g., *faafa* 'search for', shown in Example 7 above). All such examples concern the /a/-series prefix.

14. More fine grained animacy distinctions (inanimate – non-human animate – human – speech-act participant) were tested but not found to be statistically significant.

Table 13 compares the animacy of P arguments for the overt prefixes (fixed and differential) with proportions in Fedden et al. (2013). The preferences are the same, albeit less strong in our data: animate for /a/- and /e/-series; inanimate for the /o/-series.

Table 13. Comparison of % of animate P arguments indexed by marking strategy (corpus % to 1dp)

	% animate Ps	
	Fedden et al. 2013	**Corpus**
/a/-series	79%	72.0%
/o/-series	favours inanimates	19.4%
/e/-series	83%	71.1%

In Table 14, we compare the proportions of P arguments indexed with any prefix (fixed and differential). The proportion of animate P arguments indexed with a prefix is higher in Fedden et al. (2013) than in the present study (2.7 percentage-point difference; statistically significant: $\chi 2\ (1) = 5.6227$, $p < 0.05$). The difference is much greater for inanimate P arguments: 70% of inanimate arguments were indexed in the previous study, while in the corpus this figure is 48.4%. This results in a lower overall rate of prefixation of P arguments, dropping by 18 percentage points from 78% in Fedden et al. to 60% in the present study. It may be due to chance that the 12 verbs targeted for elicitation with inanimate P arguments in the previous study happen to be those that are frequently indexed. In the present study, 81 verbs (256 tokens) are attested with inanimate P arguments, which may give a more accurate view of indexation of inanimate P arguments.

Table 14. Comparison of % of P arguments indexed by a prefix (corpus % to 1dp)

	% indexed by prefix	
	Fedden et al. 2013	**Corpus**
All P	78%	60.0%
Animate P	85%	82.3%
Inanimate P	70%	48.4%

We now turn to the indexing of S arguments. The results in Figure 6 show that, unlike for P arguments, there is little in the way of clear animacy preferences among the different strategies. The /a/-series and differential /e/-series have slightly higher proportions of animates compared to zero. Fixed /e/-series, though numbers are low, has the largest proportion of inanimates.

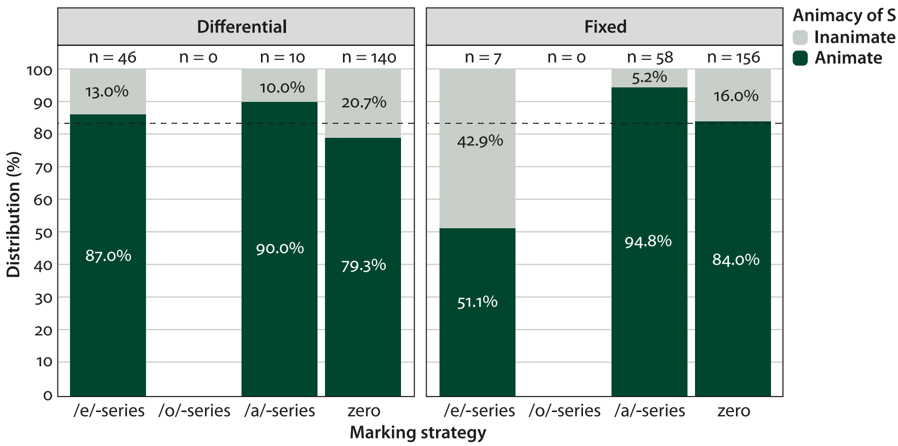


Figure 6. Distribution of S arguments by animacy (% to 1dp, $n = 417$; /o/-series absent due to indexing P only)

It is not clear what an unexpected combination of role and animacy should be for S: in one study, S was found to have an intermediate position between A (correlation with animacy) and P (correlation with inanimacy) (Brickell & Schnell 2017). The strong correlation between S and animates shown in Figure 3 suggests that the unexpected combination would be inanimate S; that is, inanimate S should be more likely to be overtly indexed. Our data shows the opposite, with inanimate S less frequently prefixed than animate S (19.4% vs. 30.6%). As mentioned, for differential indexing the same trend is observed for P arguments (44.7% vs. 67.6%). Hence, the effect of animacy on prefix versus zero marking does not depend on argument role. We thus confirm previous analysis that an explanation based on the expected combination of role and animacy cannot account for the rate of prefixation versus zero for S arguments (Fedden et al. 2013: 59). We also confirm the finding "if an S is indexed, it is animate" (Fedden et al. 2013: 58); however, even if S is not indexed with an overt prefix, it is highly likely to be animate as well. At this point, then, we cannot generalise about the effect of animacy of the S argument on indexation across all prefixes.

4.3 Discourse factors

This subsection discusses two discourse-related factors: co-occurrence with independent (nominal or pronominal) arguments (Section 4.3.1) and topicality (Section 4.3.2).

4.3.1 Co-occurrence constraints with independent arguments

Arguments may be realised as a lexical NP, an independent pronoun or null (i.e., they are not overtly expressed). In some Alor-Pantar languages, arguments may not be expressed as both an independent pronoun and a prefix in a single clause (e.g., Abui (Saad 2020), Adang (Robinson & Haan 2014),[15] Blagar (Steinhauer 2014), Kula (Williams 2017)). The indexing prefix is permitted to co-occur with a co-referential NP, however. In other Alor-Pantar languages, by contrast, no such co-occurrence restriction on independent pronouns and indexing prefixes exists (e.g., Kui (Windshuttel & Shiohara 2017), Bukalabang (Steinhauer 2020), Nedebang (Schapper 2020b), Teiwa (Klamer 2010)). In Western Pantar (Holton 2014), there is an animacy split for certain verbs such that independent pronouns can only co-occur with indexes if the referent is animate. Neither Fedden and colleagues (Fedden et al. 2013; Fedden et al. 2014) nor Schapper (2014) report co-occurrence restrictions for Kamang.

Like other Alor-Pantar languages, Kamang has a range of independent pronoun forms, which encode such features as agentivity/control, focus, quantity and possession, among others (see Schapper 2014: 315). In the current analysis, we consider all paradigms equally as "pronouns", since some are used very rarely and numbers would be too low to analyse. Potential effects of particular pronoun series on indexing remains a matter for future research. Here, we investigate whether there is a correlation between the formal realisation of an argument in the clause – NP, pronoun, null – and if and how it is indexed on the verb. Excluded are cases in which the three options are not available: clausal arguments, interrogative pronouns and syntactically mandated null expression, as in the case of gapped relative clauses. This leaves 733 tokens in the subset, of which 43.4% are differential (318/733).

Figure 7 shows the proportions of different argument-realisation types across fixed and differential occurrences of each marking strategy. The dashed lines indicate the corpus averages: below the lower dashed line for NPs (41.5%), between the lines for pronouns ("Pro", 19.2%) and above the upper line for null (39.3%).

15. As pointed out by a reviewer, this holds for "object" pronouns; there is a potential exception in the case of third-person plural non-object forms (see Haan 2001: 49; Robinson & Haan 2014: 260–267).

Unlike in some other Alor-Pantar languages in which independent pronouns may not co-occur with pronominal marking, Kamang has no absolute restriction, since pronouns co-occur with all prefix series. Nevertheless, zero marking does indeed have the greatest proportion of independent arguments realised as pronouns.

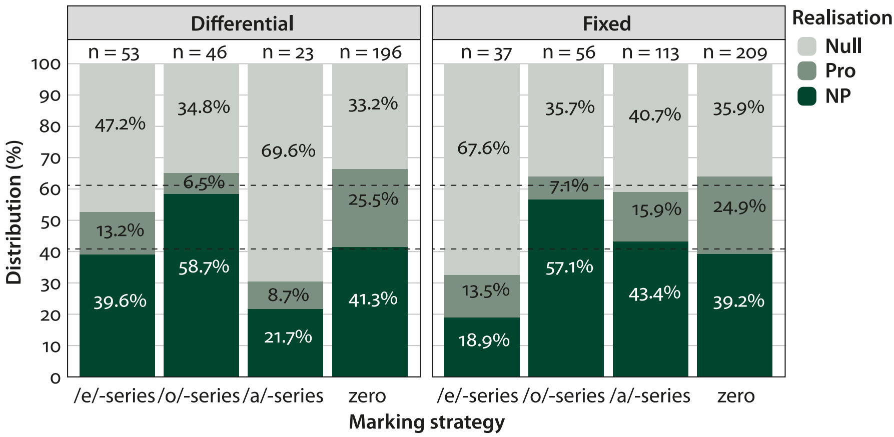


Figure 7. Distribution of independent argument realisation by marking strategy: NP vs. pronoun vs. null (% to 1dp, $n = 733$)

Focusing on DAI, we observe that the share of independent pronouns decreases for all prefixes compared to zero. These differences are significant: according to the results of our multinomial model, argument realisation is a significant factor in predicting prefix choice ($\chi 2$ (6) = 18.558, $p < 0.01$) with a negative effect for pronouns – compared to null – for all prefix series (significant for /e/-series (Est. = −0.75, $p < 0.05$) and /a/-series (Est. = −1.19, $p < 0.05$), but not for the /o/-series).[16] From the binominal model, we find, as expected, a positive effect for zero marking (Est. = 0.72, $p < 0.01$).

Figure 8 compares the fixed and differential occurrences of prefixes versus zero marking for the different forms of the independent argument. It clearly shows the general trend for pronouns to co-occur less frequently with prefixes compared with zero marking. However, as mentioned, there is no categorical rule against such co-occurrence. And, since there is no effect of co-occurring with a

16. Considering the results of our binomial model, it is plausible that for the /o/-series the important effect of animacy (Est. = −1.40, $p < 0.001$) covers the effect of argument realisation. Note that due to the size of the dataset our models have low statistical power. If we remove animacy, argument realisation does significantly contribute to the fit and similarly shows a significant negative effect for pronouns (Est. = −0.86, $p < 0.05$).

lexical NP, it seems unlikely that prefixes are purely a tracking device used in order to mention referents that are not overtly expressed elsewhere in the clause. Further research should investigate whether there are information-structural constraints on the co-occurrence of prefixes and independent pronouns.

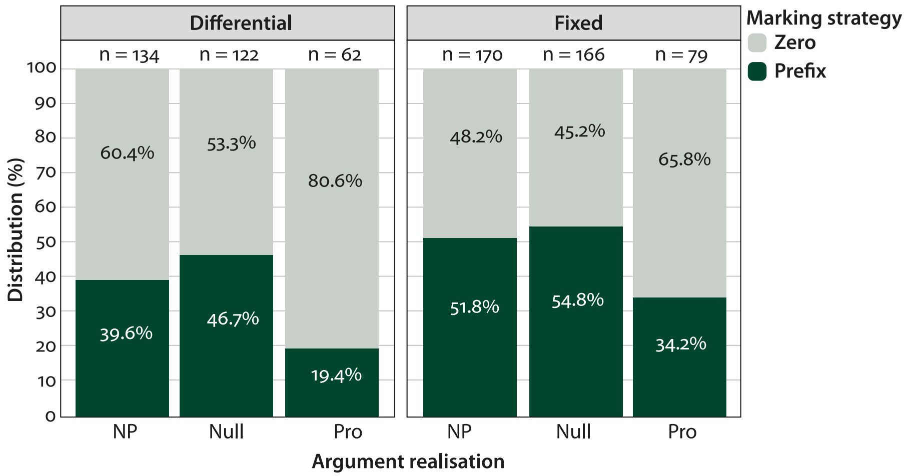


Figure 8. Distribution of independent argument realisation by marking strategy (prefix vs. zero): NP vs. pronoun vs. null (% to 1dp, $n=733$)

4.3.2 Topicality

Several proxy measures of topicality mentioned in the literature were taken, two of which contributed significantly to the fit of the model.[17] First, "referential distance" (Givón 1983) measures the number of clauses between a referent and its previous mention. The shorter the distance to last mention, the more topical the referent. Following Givón (1983), we assigned distance scores of up to 20 (i.e., distances of 20 clauses or more between the referent and its previous mention were recorded as the maximum distance of 20). Second, "topic persistence" concerns the number of times a referent is mentioned in the following clauses. We follow previous authors (Givón 1994; Payne 1994) in assigning a topic persistence score of 0–10. Zero means that the referent is not mentioned again in the next 10 clauses; 10 is the maximum score (= greatest topicality) and means that a referent is mentioned in each of the ten following clauses. Given the high correlation

17. Further measures related to topicality that were not found to have a significant effect in the model were ACTIVATION STATE – whether a referent is given or new (see Chafe 1976; Gundel et al. 1993; Prince 1981) – and GENERALIZABILITY, which concerns whether referents are "specific individuals or instances of a category or group" (Ewing 2005: 144).

between referential distance and topic persistence ($F(1, 245) = 21.13$, $p < 0.001$; Est. -0.13, $p < 0.001$), as expected, we report only on topic persistence, which gave better results in our models.

Figure 9 compares the topic persistence scores for S and P arguments across the corpus. Note that it is a continuous variable in the model, but, for clear visualisation, it is categorical in the figures. The three categories are 0 (low topicality), 1–2, and 3+ (high topicality) (cf. Givón 1994).[18] Figure 9 reveals a clear difference between the topicality profiles of P and S: S arguments are more than twice as likely to be highly topical (with a score of 3+) than P arguments.

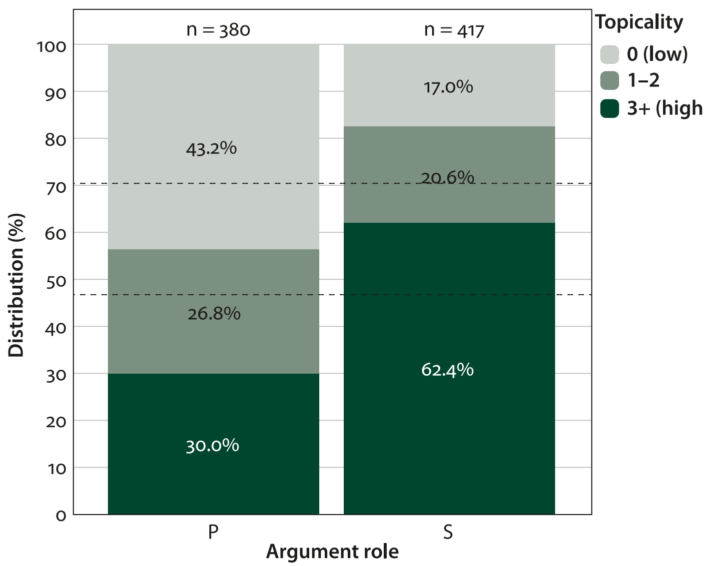


Figure 9. Distribution of topic persistence categories by argument role (% to 1dp, $n = 797$)

In Figure 10, the topicality of arguments indexed by each strategy is shown for S arguments. Figure 11 shows the proportions for P arguments.

18. Dryer (1994) and Quick (2005) also compare relative distance of A and P, but the size of the corpus and number of different marking strategies precludes us from usefully analysing this measure.

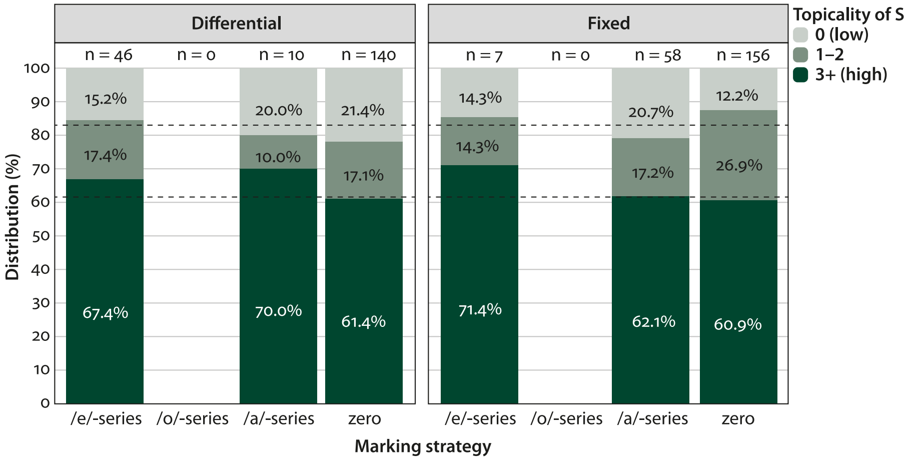


Figure 10. Distribution of topic persistence categories by marking strategy for S arguments (% to 1dp, $n = 417$)

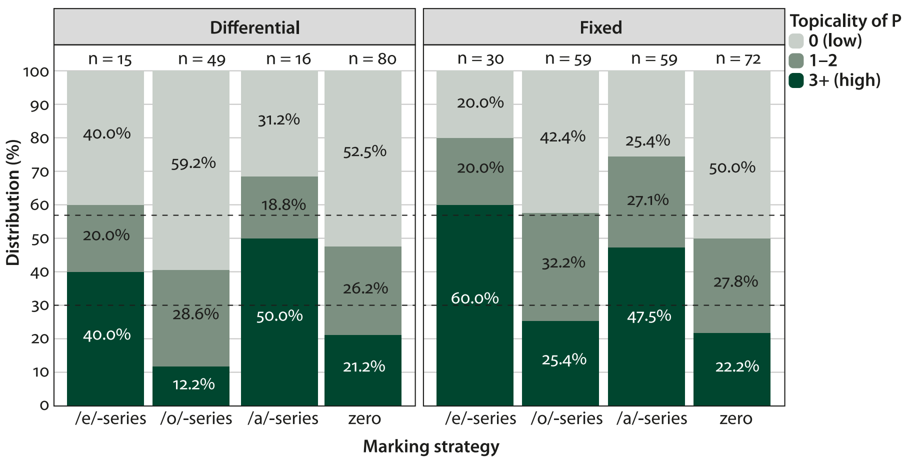


Figure 11. Distribution of topic persistence categories by marking strategy for P arguments (% to 1dp, $n = 380$)

Focusing on differential tokens, we found a negative significant effect of topic persistence for the /o/-series (Est. 0.25, $p < 0.05$). That is, an /o/-series index is more likely to occur with low-topicality (P) arguments.[19] This result is to be expected given that the /o/-series is more likely to index inanimates, which are

19. Our multinomial model diagnostics confirm that topic persistence significantly contributes to the fit ($\chi2(3) = 10.49$, $p < 0.05$) and give a better fit than referential distance (see Appendix A for details).

less likely to be topical, as shown in Figure 12. However, the effect of topic persistence is significant even with animacy included in the model (significant negative effect, Est. −1.40, $p < 0.001$, see Section 4.2.2). The association between differential /o/-series prefixes and low-topicality P arguments is a significant finding considering that overt marking in differential object indexing is claimed to be associated with highly topical arguments cross-linguistically (Iemmolo 2011: 134).

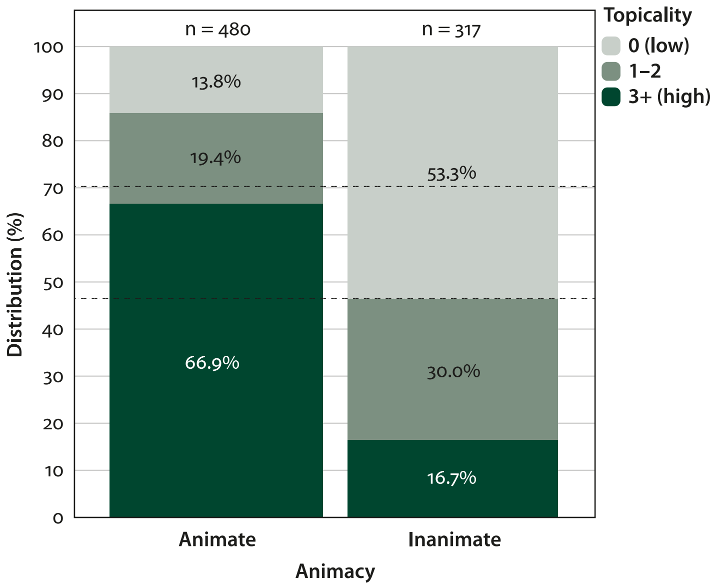


Figure 12. Distribution of topic persistence categories by animacy (% to 1dp, $n = 797$)

In contrast to the /o/-series, the /e/- and /a/-series appear to follow the cross-linguistically attested association between overt differential indexation and more-topical P arguments. That is, where the /o/-series is more frequently attested with low-topicality P referents, the /e/- and /a/-series more frequently index highly topical P referents; see Figure 11. However, the effect of topic persistence is statistically significant (Est. +0.23, $p < 0.05$) for the /a/-series, but not the /e/-series.

4.4 Lexical stipulation

As previous research has shown, a large part of variation in Kamang indexing is due to lexical stipulation. In the previous two sections, we outlined statistically significant tendencies concerning the effect of four factors in DAI in Kamang. However, not all alternation can be accounted for by these factors. In fact, the models reported in the previous two sections have quite weak predictive power:

the highest accuracy rate obtained by our multinomial model is 66.0%. This is very low compared to the 61.6% obtained for the null model, a basic model that always predicts the most probable outcome. This most probable outcome is the most frequent one; here, zero indexing is predicated as opposed to any of the prefix series.

Concerning the binomial models, the fit is substantially improved by simply limiting the dataset to verbs whose alternation profile matches the indexing strategy under investigation. Furthermore, fitting mixed-effect models, where the outcome is calculated for each verb, yields considerably better accuracy. For the /o/-series, for instance, the predictive power of our primary fixed-effect model is the same as for the null model. Yet, the former includes animacy and topic persistence – both have significant effects (animacy: Est = −1.40, $p < 0.001$; topic persistence: Est. = −0.25, $p < 0.05$) and contribute significantly to the fit. Limiting the data to include only verbs that have the /o/-series in their alternation profile makes a considerable difference, improving the predictive power. The best results come from a mixed-effect model fitted to the same dataset (see Appendix A).[20] These differences show that verbs have specific alternation preferences, confirming the observation that lexical stipulation plays an important role in Kamang differential indexing.

Also in line with previous research, for the most part the associations between particular prefix series and particular verbs – whether alternating or not – do not produce coherent semantic classes (Fedden et al. 2014: 60–64). Starting with verbs that choose the (differential or fixed) /o/-series, we do not find that they describe similar events, nor that they take semantically similar kinds of P arguments. Consider, for instance, *aakai* 'trap; trick', *baa* 'make', *o* 'follow', *tee* 'cover'. However, we are able to identify two semantic subgroups: (1) verbs of speech, including *um* 'discuss; teach', *kawaai* 'discuss; speak to', *wal* 'sing', *iti* 'call out'; and (2) verbs involving manipulation of some kind – carrying, gathering or moving something or someone – such as *pok* 'bundle', *pakah* 'carry weight', *ra* 'carry', *ileh* 'drag'.

For /a/-series verbs, too, it is difficult to propose coherent semantics. The /a/-series indexes the P argument of some highly transitive verbs, such as *kut* 'stab', *rot* 'cut off', but also of a number of low-transitivity verbs, including *tau* 'meet' and *suma* 'compare; try'. Some semantic coherence can be found for monovalent

20. Note that the effect size and significance change in mixed-models, and some factors are no longer significant when verbal lemma is taken into account. This is not surprising since it is plausible that with verb bias covering a large part of the variation, the effects of other factors are smaller in our data and hence harder to detect, given the weak statistical power due to low numbers in our dataset.

verbs for which the /a/-series indexes S. One group includes several emotional-state/experiencer verbs: *maaung* 'want (intr.)', *mai* 'not want', *pa* 'have fun', *lai* 'be glad', *maitang* 'be hungry'. Another group includes the change of state verbs *oo* 'be born; give birth to' (labile), *bo'ra* 'die' and *saakda* 'die of old age'.

The class of non-alternating /e/-series verbs is small, and is dominated in our corpus by instances of *n* 'give; marry'. Other verbs that take fixed or differential /e/-series are experiencer verbs *maringki* 'be startled', *biee* 'be afraid' and directionals: *silang* 'go down', *te* 'go up', *yaangme* 'come down'. As with all other classes, there are verbs with quite different semantics; for instance, *taa* 'sleep' and *fanee* 'work; strike' both take /e/-series prefixes.

Just as verbs grouped by prefix are semantically heterogeneous, verbs with similar semantics may belong to different prefix groups. For instance, *sooi* 'ask' is marked with the /a/-series to index the addressee, while *eisin* 'ask' is never prefixed. And *silang* 'descend; fall' can take differential /e/-series marking to index S, while *mu'tan* 'fall from above' is never prefixed. The full list of verbs is provided in Appendix B.

5. Summary: Differential indexing in Kamang

Section 4 reveals that Kamang argument indexing is not a single, unified system. Rather, the various conditioning factors impact each prefix series in a different way. Table 15 displays the results for each prefix series as it occurs with alternating verbs; that is, their behaviour in DAI. Zero marking is the baseline – almost all alternating verbs can take differential zero marking. Note, however, that there are numerous verbs in the corpus that appear with zero only (see Appendix B).

Table 15. Summary of factors per differential prefix series

Differential prefix series	Preferred animacy	Preferred argument role	Preferred topicality	Co-occurrence
/a/-series	Animate	P (weak preference)	high	Disfavour pronoun
/o/-series	Inanimate	P	low	Disfavour pronoun
/e/-series	Animate	S	high	Disfavour pronoun

In the following subsections, we summarise the profile of each indexing strategy separately, and provide a comparison with earlier research where relevant.

5.1 /o/-series

We confirmed previous findings (Fedden et al. 2013) that, unlike the other prefixes, the /o/-series generally indexes inanimate arguments. It also indexes arguments with low topicality and prefers not to co-occur with an independent pronoun. It was the second most frequent prefix in our corpus after the /a/-series, not the most frequent as Fedden and colleagues found. Fedden et al. (2014) report that differential /o/-series prefixes can be used to index S in order to indicate that it is more affected, but our data could not confirm this due to the fact that, in the current corpus, it indexes P arguments only.

Differential /o/-series is also used as a valency-increasing device. It adds a P argument, which may result in a verb having a different meaning, as in (13a), or adds a location to a positional verb, as in (13b). Since the INCORP prefix *wo-* also has an applicative function, the question remains as to whether INCORP prefix *wo-* and INDEX *wo-* can be distinguished (see Section 2).

(13) S = Zero P = /o/-series

a. *Ø-kawaila* *wo-kawaila*
'stumble' 'surround something'

b. *Ø-nih* *wo-nih*
'sit' 'sit on something'

5.2 /a/-series

For the /a/-series – the most frequent prefix for fixed and differential occurrences combined – we confirmed previous findings in that it favours animates and can index P and S arguments. The preferred argument is a topical, animate P, which is not realised as an independent pronoun. We also find that the /a/-series is only very infrequently differential: as shown by Fedden et al. (2013), there are a small number of verbs that exhibit animacy-based DAI. In our corpus, one additional function is as an applicative that adds an argument: *taang* 'descend' becomes /a/-*taang* 'descend from someone'.

The /a/-series in Kamang reflects – formally and functionally – the reconstructed P prefix in Proto-Alor-Pantar (Klamer & Kratochvíl 2018). Proto-Alor-Pantar likely had an animacy-based alignment system, in which animate referents were indexed with a P prefix and inanimates were not indexed (Klamer & Kratochvíl 2018: 75–76). Remnants of the older system can even be found in predominantly syntactic alignment system such as Kaera (Pantar), where a small number of S arguments may still be encoded like P with a prefix, even if the S is agentive and controlled (Klamer 2014: 136).

5.3 /e/-series

As in previous research, we find that the /e/-series favours animate arguments. For P arguments, however, the proportion of animates, 71.1%, is somewhat lower than the 83% reported in Fedden et al. (2013), a figure based on a limited number of verbs. It is overall the least frequent of the three prefix series studied here, but the most frequent differential series. Fedden and colleagues (Fedden et al. 2013; Fedden et al. 2014) report that the /e/-series frequently indexes P and can mark S, where S-indexation on motion and posture verbs "indicates that the motion is forced by an external cause [...] or a posture is held for a long time" (Fedden et al. 2014: 65).[21] We find that fixed /e/-series prefixes generally index P (primarily the recipient of *n* 'give; marry'), but differential occurrences favour S. That is, the /e/-series participates in DAI of animate S arguments. Like the /a/-series, the /e/-series favours more topical arguments and rarely co-occurs with independent pronouns. However, it is unlike the other prefixes in that alternation is not conditioned by animacy, but alternates within the animate category. Whether this is due to properties of the predicate/event as suggested by previous authors is addressed in Walker (accepted).

6. Conclusion

Our study presents a detailed description of a cross-linguistically rare and under-described type of indexing system; namely, partially differential S/P indexing. It shows that the various strategies used in Kamang react differently to different factors. As Iemmolo (2011) has noted, differential indexing does not seem to adhere to the same functional tendencies as differential flagging. In particular, it is not the case that the more complex marking strategy (prefixation instead of zero marking, in our case) is used to mark the least frequent combination of argument role and argument properties. In Kamang, we rather see the opposite, especially for differential /o/- and /e/-series: the /o/-series marks the frequent combination of an inanimate, non-topical P, whereas the /e/-series prefers to mark the frequent combination of an animate, topical S. Only the very restricted animacy-based alternation involving differential /a/-series – the infrequent combination of animate P is

21. There is a discrepancy in the source, which states that /e/-series prefixes do not appear on alternating verbs (Fedden et al. 2014: 64), whilst including an example of differential /e/-series on (*ge-*)*tak* 'run (away)' (Fedden et al. 2014: 66). This may be an overgeneralisation from the finding, reported in Schapper (2014: 256), that the /e/-series cannot be used in applicative alternations (i.e., to add a P argument to an otherwise monovalent verb).

indexed – can be considered in line with the pattern predicted by the frequency-based account of differential marking.

The fact that DAI in Kamang mostly does not involve marking infrequent role-reference associations ties in with earlier literature concerning the general function of indexation as opposed to flagging. As Witzlack-Makarevich & Seržant (2018: 30) point out, flagging has been associated in the literature with both a discriminatory function (to distinguish A and P arguments from each other) and with an identifying function (to highlight a particular property of an argument). Indexing, in contrast, is claimed to have only the latter function. This is in line with the primary functions of each type of marking in general: whereas flagging is first and foremost a way to indicate relations between a verb and its dependents, indexation is a means of tracking referents through discourse. The reference-tracking function of indexing in Kamang is supported by our findings on co-occurrence constraints, which show that indexation and free pronouns tend not to occur together. Yet, we also show that tracking cannot be the only function of Kamang DAI, since full NP arguments do co-occur with indexation, at least in the case of /e/-series and /o/-series prefixes.

The /a/-series is the oldest verbal prefix – a prefix series with theme vowel /a/ is reconstructed for Proto-Alor-Pantar, which indexed animates (Klamer & Kratochvíl 2018: 74). This helps to explain some of the synchronic behaviour of the /a/-series in Kamang: it is most often fixed and indexes animate referents, as in Proto-Alor-Pantar. Additionally, in the few verbs that alternate, the alternation is generally animacy-based DAI. Animacy-based systems are claimed to be more likely to lead to arbitrary lexical classes than systems based on factors such as affectedness and volitionality (Fedden & Brown 2017: 435). The high degree of lexical stipulation currently found in Kamang fits neatly with this assertion.

The description of the Kamang indexing system presented here results from applying different methodologies in pursuit of the same research question. The next step is to conduct further qualitative work into factors that could not be coded for due to not being obligatorily marked or well understood in Kamang, or appearing only very rarely in the corpus. Above all, further investigation is needed into the effects of volitionality and telicity (Fedden et al. 2013; Fedden et al. 2014), affectedness (Schapper 2014), and interaction with aspect (Walker accepted). Additional factors include modality, identified as a relevant factor in Western Pantar (Holton 2010: 109), focus, identified as a factor in Kabola (Stokhof 1987: 638), and specificity and definiteness, identified as factors in Abui (Kratochvíl 2007).

In sum, Kamang argument indexing is best viewed not as a single system sensitive to one or even a range of competing factors in a given context, but as multiple subsystems for which different factors are relevant, and which do not operate on all verbs or all indexing strategies equally.

Funding

The research and part of the data collection for this paper were carried out within the project *Exceptions Rule! Lexical conditions on grammatical structure* funded by the NWO (Dutch Research Council, file number VI.Vidi.195.008) at the University of Amsterdam.
Open Access publication of this article was funded through a Transformative Agreement with University of Amsterdam.

Acknowledgements

We are grateful to Antoinette Schapper for sharing her Kamang corpus with us, and to George Saad for the Kamang data he collected and his consultation during the annotation process. Our thanks also go to Clemens Mayer for his help with corpus annotation and to those who provided helpful feedback and suggestions on earlier versions of this paper: Sebastian Fedden, Kees Hengeveld, Sonja Riesberg and two anonymous reviewers.

Abbreviations

1, 2, 3	first, second, third person	INCL	inclusive
AT	'at'	HUM	human
/e/, /o/, etc.	/e/-series prefix, /o/-series, etc.	INCORP	incorporation
AGT	agentive	IPFV	imperfective
ASSOC	associative	KNWN	knowledge (i.e., accessible/ proximal/visible/well-known to speaker and/or addressee)
CLF	classifier		
CMN	common person		
COMP	complementiser	LV	light verb
CONTR	contrastive	NEG	negative
CSEQ	consequential relations	NKNWN	non-knowledge
DEF	definite	PFV	perfective
DIST	distal	PL	plural
dp	decimal place	REL	relative
DU	dual	SBEN	self-benefactive
EXCL	exclusive	SPEC	specific
FOC	focus	SG	singular

References

Al-Gariri, Husam. unpublished ms. *Vowel length contrast in prefixes of Kamang (dialectal variation)*. Universiteit van Amsterdam.

doi Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Baumann, Stefan & Arndt Riester. 2012. Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In Gorka Elordieta & Pilar Prieto (eds.), *Prosody and Meaning*, 119–162. Berlin: De Gruyter Mouton.

Brickell, Timothy C. & Stefan Schnell. 2017. Do grammatical relations reflect information status? Reassessing Preferred Argument Structure theory against discourse data from Tondano. *Linguistic Typology* 21(1). 177–208.

Brugman, Hennie. & Albert. Russel. 2004. Annotating Multi-media/Multi-modal resources with ELAN. In Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa & Raquel Silva (eds.), *Proceedings of the 4th International Conference on Language Resources and Language Evaluation* (LREC 2004), 2065–2068. Paris: European Language Resources Association.

Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Charles N. Li (ed.), *Subject and topic*, 27–55. New York: Academic Press.

Coghill, Eleanor. 2014. Differential object marking in Neo-Aramaic. *Linguistics* 52(2). 335–364.

Dryer, Matthew S. 1994. The Discourse Function of the Kutenai Inverse. In Talmy Givón (ed.), *Voice and inversion*, 65–99. Amsterdam: John Benjamins.

Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *WALS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://wals.info/

ELAN, (version 6.0) [computer software]. 2020. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Available at https://archive.mpi.nl/tla/elan (last access 16 January 2023).

Ewing, Michael C. 2005. *Grammar and inference in conversation: Identifying clause structure in spoken Javanese*. Amsterdam: John Benjamins.

Fauconnier, Stefanie. 2011. Differential agent marking and animacy. *Lingua* 121(3). 533–547.

Fedden, Sebastian & Dunstan Brown. 2017. Participant marking: Corpus study and video elicitation. In Marian Klamer (ed.), *Alor-Pantar languages (Second Edition): History and typology* (Studies in Diversity Linguistics), 403–446. Berlin: Language Science Press.

Fedden, Sebastian, Dunstan Brown, Greville Corbett, Gary Holton, Marian Klamer, Laura C. Robinson & Antoinette Schapper. 2013. Conditions on pronominal marking in the Alor-Pantar languages. *Linguistics* 51(1). 33–74.

Fedden, Sebastian, Dunstan Brown, František Kratochvíl, Laura C. Robinson & Antoinette Schapper. 2014. Variation in pronominal indexing: Lexical stipulation vs. referential properties in Alor-Pantar languages. *Studies in Language* 38(1). 44–79.

Givón, Talmy. 1983. Topic continuity in discourse: A quantitative cross-language study. In Talmy Givón (ed.), *Topic continuity in discourse: An introduction*, 1–42. Amsterdam: John Benjamins.

Givón, Talmy (ed.). 1994. *Voice and inversion*. Amsterdam: John Benjamins.

Grossman, Eitan. 2015. No case before the verb, obligatory case after the verb in Coptic. In Eitan Grossman, Martin Haspelmath & Tonio Sebastian Richter (eds.), *Egyptian-Coptic linguistics in typological perspective*, 203–226. Berlin: De Gruyter Mouton.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.

Haan, Johnson Welem. 2001. *The grammar of Adang: A Papuan language spoken on the Island of Alor East Nusa Tenggara – Indonesia*. Sydney: University of Sydney.

Haig, Geoffrey & Stefan Schnell. 2014. Annotations using GRAID (Grammatical Relations and Animacy in Discourse). version 7.0. Available at https://multicast.aspra.uni-bamberg.de/ (last access 16 January 2023).

Haig, Geoffrey & Stefan Schnell (eds.). 2022. Multi-CAST: Multilingual corpus of annotated spoken texts. Version 2211. Bamberg: University of Bamberg. Available at https://multicast.aspra.uni-bamberg.de/ (last access 16 January 2023).

Haspelmath, Martin. 2006. Against markedness (and what to replace it with). *Journal of Linguistics* 42(1). 25–70.

Haspelmath, Martin. 2021. Role-reference associations and the explanation of argument coding splits. *Linguistics* 59(1). 123–174.

Holton, Gary. 2010. Person marking, verb classes, and the notion of alignment in Western Pantar (Lamma). In Michael C. Ewing & Marian Klamer (eds.), *Typological and areal analyses: Contributions from East Nusantara*, 99–119. Canberra: Pacific Linguistics.

Holton, Gary. 2014. Western Pantar. In Antoinette Schapper (ed.), *The Papuan languages of Timor, Alor and Pantar: Volume 1: Sketch Grammars* (Pacific Linguistics 644), 23–96. Berlin: De Gruyter Mouton.

Holton, Gary & Laura C. Robinson. 2017. The linguistic position of the Timor-Alor-Pantar languages. In Marian Klamer (ed.), *The Alor-Pantar languages: History and typology* (Studies in Diversity Linguistics 3), 147–191. 2nd edn. Berlin: Language Science Press.

Iemmolo, Giorgio. 2011. Towards a typological study of differential object marking and differential object indexation. Pavia: University of Pavia PhD dissertation.

Just, Erika & Slavomír Čéplö. 2022. Differential object indexing in Maltese – a corpus based pilot study. In Przemysław Turek & Julia Nintemann (eds.), *Maltese: Contemporary changes and historical innovations*, 105–132. Berlin: De Gruyter Mouton.

Just, Erika & Alena Witzlack-Makarevich. 2022. A corpus-based analysis of P indexing in Ruuli (Bantu, JE103). *South African Journal of African Languages* 42(2). 234–242.

Klamer, Marian. 2010. *A grammar of Teiwa* (Mouton Grammar Library 49). Berlin: De Gruyter Mouton.

Klamer, Marian. 2014. Kaera. In Antoinette Schapper (ed.), *The Papuan languages of Timor, Alor and Pantar*: Volume 1: *Sketch grammars* (Pacific Linguistics 644), 97–146. Berlin: De Gruyter Mouton.

Klamer, Marian. 2017. The Alor-Pantar languages: Linguistic context, history and typology. In Marian Klamer (ed.), *The Alor-Pantar languages: History and typology*, 1–49. 2nd edn. Berlin: Language Science Press.

Klamer, Marian & František Kratochvíl. 2018. The evolution of differential object marking in Alor-Pantar languages. In Ilja A. Seržant & Alena Witzlack-Makarevich (eds.), *Diachrony of differential argument marking*, 69–95. Berlin: Language Science Press.

Klamer, Marian & Antoinette Schapper. 2012. 'Give' Constructions in the Papuan Languages of Timor Alor Pantar. *Linguistic Discovery* 10(3). 174–207.

Kratochvíl, František. 2007. A grammar of Abui: A Papuan language of Alor. Utrecht: LOT PhD Thesis.

Kratochvíl, František. 2014. Differential argument realization in Abui. *Linguistics* 52(2). 543–602.

doi Lazard, Gilbert. 2005. What are we typologists doing? In David S. Rood, Adam Hodges & Zygmunt Frajzyngier (eds.), *Linguistic Diversity and Language Theories* (Studies in Language Companion Series), 1–23. Amsterdam; Philadelphia: John Benjamins.

doi Payne, Thomas E. 1994. The pragmatics of voice in a Philippine language. In Talmy Givón (ed.), *Voice and inversion*, 318–64. Amsterdam: John Benjamins.

Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, 223–255. New York: Academic Press.

Quick, Phil. 2005. Topic continuity, voice and word order in Pendau. In I. Wayan Arka & Malcolm Ross (eds.), *The many faces of Austronesian voice systems: Some new empirical studies*, 221–242. Canberra: Pacific Linguistics.

R Core Team. 2019. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at https://www.R-project.org/ (last access 16 January 2023).

Riester, Arndt & Stefan Baumann. 2017. The RefLex scheme – Annotation guidelines. In *SinSpeC. Working Papers of the SFB* 732, vol. 14, 1–30. University of Stuttgart.

doi Robinson, Laura C. & John W. Haan. 2014. Adang. In Antoinette Schapper (ed.), *The Papuan languages of Timor, Alor and Pantar: Volume 1: Sketch grammars* (Pacific Linguistics 644), 221–284. Berlin: De Gruyter Mouton.

doi Saad, George. 2020. Abui. In Antoinette Schapper (ed.), *The Papuan languages of Timor, Alor and Pantar*, Volume 3: *Sketch grammars* (Pacific Linguistics 660), 267–348. Berlin: De Gruyter Mouton.

Schäfer, Roland. 2020. *Mixed-Effects Regression Modeling. A practical handbook of corpus linguistics*. Cham: Springer.

doi Schapper, Antoinette. 2014. Kamang. In Antoinette Schapper (ed.), *The Papuan languages of Timor, Alor and Pantar: Volume 1: Sketch grammars* (Pacific Linguistics 644), 285–350. Berlin: De Gruyter Mouton.

doi Schapper, Antoinette. 2020a. Introduction. In Antoinette Schapper (ed.), *The Papuan languages of Timor, Alor and Pantar*, Volume 3: *Sketch grammars* (Pacific Linguistics 660), 1–52. Berlin: De Gruyter Mouton.

doi Schapper, Antoinette. 2020b. Nedebang. In Antoinette Schapper (ed.), *The Papuan languages of Timor, Alor and Pantar*, Volume 3: *Sketch grammars* (Pacific Linguistics 660), 53–138. Berlin: De Gruyter Mouton.

Schapper, Antoinette & Marten Manimau. 2011. *Kamus Pengantar Bahasa Kamang-Indonesia-Inggris* (Introductory Kamang – Indonesian – English Dictionary). Kupang, Indonesia: UBB-GMIT.

Schapper, Antoinette, George Saad, Katherine Walker & Clemens Mayer. in prep. Multi-CAST Kamang. In Geoffrey Haig & Stefan Schnell (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. Version 2211. Bamberg: University of Bamberg.

Schiborr, Nils N., Stefan Schnell & Hanna Thiele. 2018. *RefIND – Referent Indexing in Natural-language Discourse: Annotation guidelines* (v1.1). Bamberg; Melbourne: University of Bamberg/University of Melbourne.

Seržant, Ilja A. & Alena Witzlack-Makarevich (eds.). 2018. *Diachrony of differential argument marking* (Studies in Diversity Linguistics 19). Berlin: Language Science Press.

Siewierska, Anna. 2013a. Verbal person marking. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at https://wals.info/chapter/102 (last access 16 January 2023).

Siewierska, Anna. 2013b. Alignment of verbal person marking. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at https://wals.info/chapter/100 (last access 16 January 2023).

doi Siewierska, Anna & Dik Bakker. 2008. Case and alternative strategies: Word order and agreement marking. In Andrej L. Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 290–303. Oxford: Oxford University Press.

doi Steinhauer, Hein. 2014. Blagar. In Antoinette Schapper (ed.), *The Papuan languages of Timor, Alor and Pantar*: Volume 1: *Sketch grammars* (Pacific Linguistics 644), 147–220. Berlin: De Gruyter Mouton.

doi Steinhauer, Hein. 2020. Bukalabang. In Antoinette Schapper (ed.), *The Papuan Languages of Timor, Alor and Pantar*, Volume 3: *Sketch grammars* (Pacific Linguistics 660), 139–188. Berlin: De Gruyter Mouton.

Stokhof, W. A. L. 1975. *Preliminary notes on the Alor and Pantar languages (East Indonesia)* (Pacific Linguistics B 43). Canberra: Linguistic Circle of Canberra.

Stokhof, W. A. L. 1987. A short Kabola text (Alor, East Indonesia). In Donald C. Laycock & Werner Winter (eds.), *A world of language: Papers presented to Professor S. A. Wurm on his 65th birthday*, 631–648. Canberra: Pacific Linguistics.

doi Venables, William N. & Brian D. Ripley. 2002. *Modern Applied Statistics with S*. Fourth. New York: Springer.

doi Virtanen, Susanna. 2014. Pragmatic direct object Marking in Eastern Mansi. *Linguistics* 52(2). 391–413.

Walker, Katherine. accepted. Differential indexing in Kamang: A viewpoint alternation. *Linguistics Vanguard* (Special Issue).

Willemsen, Jeroen. 2017. Predicative Augmentation Applicatives. *Linguistica* ONLINE 19. 1–22. Available at http://www.phil.muni.cz/linguistica/art/issues/issue-019.pdf (last access 16 January 2023).

Williams, Nicholas J. 2017. Kula. In Antoinette Schapper (ed.), *The Papuan languages of Timor, Alor and Pantar*, Volume 2: *Sketch grammars* (Pacific Linguistics 655), 185–266. Berlin: De Gruyter Mouton.

Windshuttel, Glenn & Asako Shiohara. 2017. Kui. In Antoinette Schapper (ed.), *The Papuan languages of Timor, Alor and Pantar*, Volume 2: *Sketch grammars* (Pacific Linguistics 655), 109–184. Berlin: De Gruyter Mouton.

Witzlack-Makarevich, Alena & Ilja A. Seržant. 2018. Differential argument marking: Patterns of variation. In Ilja A. Seržant & Alena Witzlack-Makarevich (eds.), *Diachrony of differential argument marking*, 1–40. Berlin: Language Science Press.

Appendix A. Statistical modelling

I. Introduction

This appendix provides the details of our multifactorial (logistic) regression modelling, the results of which were presented in Section 4.2 to 4.4. We start with the dataset, variable definition and coding, and then provide full summaries of the models along with details on model selection and model quality.

As mentioned in the text, all the analyses were done in R. The code and data can be found at the following link: https://osf.io/pvmbw/?view_only=43f05df2f2954951a5945ea4ac8fad5f

II. Dataset and variable coding

The goal of our modelling was to investigate the effect of a set of functional factors (animacy, argument role, argument realisation and topicality) in differential indexing. More precisely, we wanted to estimate the contribution of each factor in determining the choice of indexing strategy in the case of alternating verbs. Hence, the data is limited to the alternating subset of the corpus, including only the following four major marking strategies: /e/-series, /o/-series, /a/-series and zero.

A. Predicting variables

Choosing between Topic persistence (TP) and Referential distance (RD) as measures of topicality:

These measures are highly correlated:

```
lm(formula = ArgTP1 ~ ArgRD1, data = AltDS)

Residuals:
    Min       1Q   Median       3Q      Max
-4.0809  -1.9809  -0.7943  1.9191   6.0921

Coefficients:
              Estimate   Std. Error   t value   Pr(>|t|)
(Intercept)    4.25391      0.19617    21.685    < 2e-16  ***
ArgRD1        -0.17298      0.01817    -9.521    < 2e-16  ***
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.614 on 316 degrees of freedom

Multiple R-squared: 0.2229, Adjusted R-squared: 0.2204

F-statistic: 90.64 on 1 and 316 DF, p-value: < 2.2e-16
```

We compared goodness-of-fit to select the best predictor:

```
multinom(formula = choice ~ ArgTP1, data = AltDS)
Coefficients:
          (Intercept)          ArgTP1
1          -1.5369436     0.064471241
2          -0.5920302    -0.489911647
3          -2.1737311     0.009409013
Std. Errors:
          (Intercept)          ArgTP1
1           0.2455880      0.05117558
2           0.2045975      0.10819499
3           0.3317169      0.07424120
Residual Deviance: 637.1292
AIC: 649.1292
-> R2 = 0.12
```

```
multinom(formula = choice ~ ArgRD1, data = AltDS)
Coefficients:
          (Intercept)          ArgRD1
1          -1.150103      -0.02595372
2          -2.146889       0.07385478
3          -1.851602      -0.05405053
Std. Errors:
          (Intercept)          ArgRD1
1           0.1969863      0.02150812
2           0.2722856      0.01933959
3           0.2707848      0.03579558
Residual Deviance: 655.1943
AIC: 667.19
-> R2 = 0.80
```

The predicting factors investigated in our analysis were defined as follows:

– Argument role as a binary variable (S vs. P) → ArgType
– Animacy as a binary variable (animate vs. inanimate) → ArgAnim
– Argument realisation as a nominal variable with three levels (NP, Pro and Null) → ArgRlz
– Topic persistence as a discrete variable with values spanning from zero to 10 → ArgTP1

Our nominal predicting variables were sum-coded (or centred), as follows:

```
ArgType:  ArgAnime:   ArgRlz:
    [,1]       [,1]        [,1]  [,2]
S      1  Anim     1   Pro     1     0
P     -1  Inan    -1    NP     0     1
                      Null    -1    -1
```

Given these definitions, the final dataset included 318 data points. Note that the ratios reported in the main text are calculated for the whole dataset (i.e., the total of 797 tokens, of which 365 are alternating verbs), whereas the coefficient estimates are drawn from the multifactorial modelling results presented in this appendix.

B. Response variable

In the multinomial model, the response variable (named “choice”) codes the indexing strategy on 4 levels, with “zero” as the baseline:

```
0  →  zero
1  →  /e/-series (named “GEN” in the data frame)
2  →  /o/-series (named “LOC” in the data frame)
3  →  /a/-series (named “PAT” in the data frame)
```

In the binomial models, fitted for each strategy separately, the response variable is always a binary variable coding presence (=1/success) vs. absence (=0/failure) of the given strategy. Note that the prefixes were named as “GEN”, “LOC” and “PAT” (instead of /e/-series, /o/-series and /a/-series, respectively) for coding convenience.

III. Models

To study the effect of our four factors, we fitted a multinomial model (Section A) and a series of four binary fixed-effects models (Section B) to the whole dataset (named “AltDS”), without taking verb lemmas into account. These are the models from which we reported the coefficient estimates for different effects in Section 4.2 and Section 4.3 of the main text.
Next, aiming to estimate the role of lexical stipulation (Section 4.4 in the main text), we first fitted a second series of fixed-effect binary models (Section C) to subsets of the data limited to the verbs whose alternation profile includes the prefix under investigation; that is, is coded as the response variable. Finally, we fitted a series of mixed-effects models with the verbal lemma as random intercept (Section D). We compare the quality of the fit of our three series of binary models in Section E.

A. Multinomial model – fitted to the whole dataset (Sections 4.2 & 4.3)

```
318 tokens

Distribution:
       /e/  /o/  /a/    Z
   0     0    0    0  196
   1    53    0    0    0
   2     0   46    0    0
   3     0    0   23    0
```

Model selection:

We report below only the results of the most relevant model comparisons using Likelihood ratio tests of Multinomial Models.

	Model	Resid. df	Resid. Dev	Test	Df	LR stat.	Pr(Chi)
1	ArgAnim + ArgTP1 + ArgType	942	516.0940				
2	ArgType + ArgAnim + ArgRlz + ArgTP1	936	497.5383	1 vs 2	6	18.5557	0.004983645
1	ArgAnim + ArgRlz + ArgType	939	501.1369				
2	ArgType + ArgAnim + ArgRlz + ArgTP1	936	497.5383	1 vs 2	3	3.598616	0.3081953
1	ArgAnim + ArgRlz	942	567.9497				
2	ArgAnim + ArgRlz + ArgTP1	939	557.4552	1 vs 2	3	10.49449	0.01479835
1	ArgType + ArgRlz	942	563.8842				
2	ArgType + ArgRlz + ArgTP1	939	553.2466	1 vs 2	3	10.63759	0.013856
1	ArgType + ArgRlz + ArgTP1	939	553.2466				
2	ArgRlz + ArgType * ArgTP1	936	546.5991	1 vs 2	3	6.647497	0.08402358
1	ArgRlz + ArgType * ArgTP1	936	546.5991				
2	ArgAnim + ArgRlz + ArgType * ArgTP1	933	495.5851	1 vs 2	3	51.01402	4.858558e-11
1	ArgType + ArgAnim + ArgRlz + ArgTP1	936	497.5383				
2	ArgType * ArgAnim + ArgRlz + ArgTP1	933	497.4341	1 vs 2	3	0.1041331	0.9913369

Summary of results

Call:

```
multinom(formula = choice ~ ArgAnim + ArgRlz + ArgType * ArgTP1, data = AltDS)
```

Coefficients:

	(Intercept)	ArgAnim1	ArgRlz1	ArgRlz2	ArgType1	ArgTP1	ArgType1:ArgTP1
1	-1.654055	0.7767583	-0.7503987	0.3079579	-0.1007462	0.01662436	-0.06865582
2	-8.995567	-0.3790691	-0.3005077	-0.1344887	-8.5379892	-0.30970969	-0.11383577
3	-10.221565	8.9382960	-1.1934081	-0.1738070	-1.1483932	-0.05561162	-0.12619068

Std. Errors:

	(Intercept)	ArgAnim1	ArgRlz1	ArgRlz2	ArgType1	ArgTP1	ArgType1:ArgTP1
1	0.2799792	0.2656414	0.3037171	0.2482772	0.3073998	0.07494490	0.07225559
2	0.2038896	0.3638613	0.4987661	0.3179002	0.2038896	0.06783363	0.06783366
3	0.2477353	0.2477355	0.5614181	0.4629413	0.4508107	0.09978019	0.10202008

Residual Deviance: 495.5851

AIC: 537.5851

Goodness-of-fit

Null Model:

```
MN_M0 <-multinom(choice~1, data=AltDS)
```

R^2

```
r.squaredLR(MN_M1, MN_M0)
[1] 0.4371003
attr(,"adj.r.squared")
[1] 0.4958427
```

Observed/predicted contingency tables:

	0	1	2	3
0	185	0	10	1
1	47	0	1	5
2	31	0	14	1
3	12	0	0	11

-> Accuracy = 66.04

	0	1	2	3
0	196	0	0	0
1	53	0	0	0
2	46	0	0	0
3	23	0	0	0

-> Accuracy = 61.64

```
P-values (using z-test):
                      Estimate  Std. Error   z value   Pr(>|z|)
     1:(Intercept)      -1.654055    0.279979   -5.9078  3.468e-09  ***
     1:ArgAnim1          0.776758    0.265641    2.9241   0.003455  **
     1:ArgRlz1          -0.750399    0.303717   -2.4707   0.013484  *
     1:ArgRlz2           0.307958    0.248277    1.2404   0.214835
     1:ArgType1         -0.100746    0.307400   -0.3277   0.743111
     1:ArgTP1            0.016624    0.074945    0.2218   0.824453
     1:ArgType1:ArgTP1  -0.068656    0.072256   -0.9502   0.342021
     2:(Intercept)      -8.995567    0.203890  -44.1198  < 2.2e-16  ***
     2:ArgAnim1         -0.379069    0.363861   -1.0418   0.297507
     2:ArgRlz1          -0.300508    0.498766   -0.6025   0.546840
     2:ArgRlz2          -0.134489    0.317900   -0.4231   0.672256
     2:ArgType1         -8.537989    0.203890  -41.8755  < 2.2e-16  ***
     2:ArgTP1           -0.309710    0.067834   -4.5657  4.978e-06  ***
     2:ArgType1:ArgTP1  -0.113836    0.067834   -1.6782   0.093316   .
     3:(Intercept)     -10.221565    0.247735  -41.2600  < 2.2e-16  ***
     3:ArgAnim1          8.938296    0.247736   36.0800  < 2.2e-16  ***
     3:ArgRlz1          -1.193408    0.561418   -2.1257   0.033528  *
     3:ArgRlz2          -0.173807    0.462941   -0.3754   0.707333
     3:ArgType1         -1.148393    0.450811   -2.5474   0.010853  *
     3:ArgTP1           -0.055612    0.099780   -0.5573   0.577294
     3:ArgType1:ArgTP1  -0.126191    0.102020   -1.2369   0.216117
```

B. Binary fixed-effects models – without considering the verbs (Sections 4.2 & 4.3)

1. *Zero*

318 tokens

```
Distribution:
     1  196
     0  122

Model selection:
     Z ~ ArgType * ArgTP1 + ArgAnim + ArgRlz
                   Df  Deviance     AIC      LRT   Pr(>Chi)
     <none>                374.10  388.10
     ArgAnim          1    385.26  397.26  11.1621  0.0008348  ***
     ArgRlz           2    387.40  397.40  13.3005  0.0012937  **
     ArgType:ArgTP1   1    374.10  386.10   0.0002  0.9888247

     Z ~ ArgType + ArgAnim + ArgRlz + ArgTP1
             Df  Deviance     AIC      LRT   Pr(>Chi)
     <none>        374.10  386.10
     ArgType   1    404.16  414.16  30.0675  4.173e-08  ***
     ArgAnim   1    385.58  395.58  11.4863  0.0007011  ***
     ArgRlz    2    387.66  395.66  13.5670  0.0011323  **
     ArgTP1    1    375.56  385.56   1.4596  0.2269898

     Z ~ ArgType + ArgAnim + ArgRlz
             Df  Deviance     AIC     LRT   Pr(>Chi)
     <none>        375.56  385.56
     ArgType   1    409.27  417.27  33.717  6.375e-09  ***
     ArgAnim   1    385.58  393.58  10.029   0.001541  **
     ArgRlz    2    388.32  394.32  12.765   0.001691  **

Summary of results:
Call:
     glm(formula = Z ~ ArgType + ArgAnim + ArgRlz, family = "binomial", data = AltDS)

Deviance Residuals:
         Min       1Q  Median      3Q     Max
     -2.1042  -1.2564  0.5477  0.9732  1.5941
```

```
Coefficients:
                 Estimate  Std. Error  z value  Pr(>|z|)
      (Intercept)   0.7039     0.1509    4.664  3.11e-06  ***
      ArgType1      0.9572     0.1813    5.279  1.30e-07  ***
      ArgAnim1     -0.5625     0.1879   -2.993   0.00276   **
      ArgRlz1       0.7228     0.2517    2.871   0.00409   **
      ArgRlz2      -0.1254     0.1974   -0.635   0.52550
      ---
      Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
          Null  deviance:  423.46  on  317  degrees  of  freedom
      Residual  deviance:  375.56  on  313  degrees  of  freedom
      AIC: 385.56
```

2. */e/-series*

```
318 tokens

Distribution:
      1   53
      0  265

Model selection:
      GEN ~ ArgAnim + ArgRlz + ArgType * ArgTP1
                      Df  Deviance     AIC     LRT  Pr(>Chi)
      <none>                266.59  280.59
      ArgAnim          1    272.76  284.76  6.1782   0.01293  *
      ArgRlz           2    271.71  281.71  5.1199   0.07731  .
      ArgType:ArgTP1   1    267.40  279.40  0.8125   0.36737

      GEN ~ ArgType + ArgAnim + ArgRlz + ArgTP1
               Df  Deviance     AIC     LRT  Pr(>Chi)
      <none>         267.40  279.40
      ArgType   1    267.91  277.91  0.5099  0.475203
      ArgAnim   1    274.70  284.70  7.3052  0.006876  **
      ArgRlz    2    272.35  280.35  4.9529  0.084043   .
      ArgTP1    1    267.40  277.40  0.0006  0.980812

      GEN ~ ArgAnim + ArgRlz
               Df  Deviance     AIC      LRT   Pr(>Chi)
      <none>         267.94  275.94
      ArgAnim   1    283.82  289.82  15.8786  6.754e-05  ***
      ArgRlz    2    272.69  276.69   4.7539    0.09283    .

Summary of results:

Call:

      glm(formula = GEN ~ ArgAnim + ArgRlz, data = AltDS)

Deviance Residuals:
          Min       1Q   Median       3Q     Max
      -0.8159  -0.7684  -0.4185  -0.3913  2.2838

Coefficients:
                 Estimate  Std. Error  z value  Pr(>|z|)
      (Intercept)  -2.0307     0.2138   -9.496    <2e-16  ***
      ArgAnim1      0.7312     0.1990    3.675  0.000238  ***
      ArgRlz1      -0.6012     0.2966   -2.027  0.042680  *
      ArgRlz2       0.3705     0.2399    1.545  0.122446
      ---
      Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
          Null  deviance:  286.56  on  317  degrees  of  freedom
      Residual  deviance:  267.94  on  314  degrees  of  freedom
      AIC: 275.94
```

3. /o/-series

318 tokens

Distribution:

```
1   46
0  272
```

Model selection:

NB. ArgType is not relevant here, as the data only include P arguments.

```
LOC ~ ArgAnim * ArgRlz + ArgAnim * ArgTP1
                Df  Deviance     AIC      LRT  Pr(>Chi)
<none>                185.12  201.12
ArgAnim:ArgRlz   2    188.18  200.18  3.05702    0.2169
ArgAnim:ArgTP1   1    185.68  199.68  0.56181    0.4535

LOC ~ ArgRlz + ArgAnim * ArgTP1
                Df  Deviance     AIC      LRT  Pr(>Chi)
<none>                188.18  200.18
ArgRlz           2    190.42  198.42  2.24341    0.3257
ArgAnim:ArgTP1   1    188.83  198.83  0.65218    0.4193

LOC ~ ArgAnim + ArgRlz + ArgTP1
         Df  Deviance     AIC      LRT   Pr(>Chi)
<none>         188.83  198.83
ArgAnim   1    217.60  225.60  28.7708  8.147e-08  ***
ArgRlz    2    190.91  196.91   2.0802    0.35341
ArgTP1    1    193.70  201.70   4.8661    0.02739  *

LOC ~ ArgRlz
        Df  Deviance     AIC     LRT  Pr(>Chi)
<none>        253.49  259.49
ArgRlz   2    262.87  264.87  9.3801  0.009186  **

LOC ~ ArgRlz + ArgTP1
        Df  Deviance     AIC     LRT   Pr(>Chi)
<none>        217.60  225.60
ArgRlz   2    223.26  227.26   5.656    0.05912   .
ArgTP1   1    253.49  259.49  35.891  2.087e-09  ***
```

Summary of results:

Call:

```
glm(formula = LOC ~ ArgAnim + ArgTP1, family = "binomial", data = AltDS)
```

Deviance Residuals:

```
    Min       1Q   Median       3Q     Max
-0.9827  -0.5769  -0.1668  -0.1055  2.9262
```

Coefficients:

```
Estimate           Std.    Error   z value    Pr(>|z|)
(Intercept)      -1.8794   0.3661    -5.134    2.84e-07  ***
ArgAnim1         -1.4025   0.3240    -4.328    1.50e-05  ***
ArgTP1           -0.2464   0.1199    -2.054      0.0399  *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for binomial family taken to be 1)

```
    Null  deviance:  262.87  on  317  degrees  of  freedom
Residual  deviance:  190.91  on  315  degrees  of  freedom
AIC: 196.9
```

→ Removing animacy and including argument realization

Call:

```
glm(formula = LOC ~ ArgRlz + ArgTP1, family = "binomial", data = AltDS)
```

```
Deviance Residuals:

    Min       1Q   Median       3Q      Max
-0.8978  -0.7083  -0.3115  -0.1357   2.3612

Coefficients:

               Estimate  Std. Error  z value  Pr(>|z|)
(Intercept)     -1.1625      0.2601   -4.470  7.83e-06  ***
ArgRlz1         -0.8594      0.4209   -2.042    0.0411    *
ArgRlz2          0.4620      0.2709    1.706    0.0880    .
ArgTP1          -0.4898      0.1069   -4.580  4.65e-06  ***
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
(Dispersion parameter for binomial family taken to be 1)
        Null  deviance:  262.87  on  317  degrees  of  freedom
    Residual  deviance:  217.60  on  314  degrees  of  freedom
    AIC: 225.6
```

4. */a/-series*

```
318 tokens

Distribution:
     1   23
     0  295

Model selection:
     PAT ~ ArgAnim + ArgRlz + ArgType * ArgTP1
                    Df  Deviance     AIC     LRT   Pr(>Chi)
     <none>               100.30  114.30
     ArgAnim         1    139.23  151.23  38.935  4.381e-10  ***
     ArgRlz          2    111.06  121.06  10.769   0.004586  **
     ArgType:ArgTP1  1    101.66  113.66   1.367   0.242268

     PAT ~ ArgRlz + ArgType * ArgTP1
                    Df  Deviance     AIC      LRT   Pr(>Chi)
     <none>               139.23  151.23
     ArgRlz          2    154.02  162.02  14.7879  0.0006149  ***
     ArgType:ArgTP1  1    145.77  155.77   6.5351  0.0105769  *

     Model 1: PAT ~ ArgAnim + ArgRlz + ArgType * ArgTP1
     Model 2: PAT ~ ArgRlz + ArgType * ArgTP1
        Resid.     Df  Resid. Dev       Df   Deviance  Pr(>Chi)
     1     311  100.30
     2     312  139.23          -1  -38.935  4.381e-10  ***

     Model 1: PAT ~ ArgTP1 * ArgType + Overt
     Model 2: PAT ~ ArgRlz + ArgType * ArgTP1
        Resid. Df  Resid. Dev  Df  Deviance  Pr(>Chi)
     1        313      139.27
     2        312      139.23   1  0.037213     0.847

     PAT ~ ArgTP1 * ArgType + Overt
                    Df  Deviance     AIC      LRT   Pr(>Chi)
     <none>               139.27  149.27
     Overt           1    154.02  162.02  14.7507  0.0001227  ***
     ArgTP1:ArgType  1    145.88  153.88   6.6145  0.0101153  *

Summary of results:

Call:

     glm(formula = PAT ~ ArgRlz + ArgType * ArgTP1, family = “binomial”, data = AltDS)

Deviance Residuals:
         Min       1Q   Median       3Q      Max
     -1.1382  -0.4074  -0.2231  -0.1947   2.9566
```

```
Coefficients:
                      Estimate Std.    Error  z value  Pr(>|z|)
      (Intercept)          -3.10742  0.43583   -7.130     1e-12  ***
      ArgRlz1              -0.52030  0.54018   -0.963    0.3355
      ArgRlz2              -0.69539  0.42485   -1.637    0.1017
      ArgType1             -0.05350  0.37967   -0.141    0.8879
      ArgTP1                0.12435  0.08503    1.462    0.1436
      ArgType1:ArgTP1      -0.22472  0.08827   -2.546    0.0109  *
      ---
      Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
          Null  deviance:  165.12  on  317  degrees  of  freedom
      Residual  deviance:  139.23  on  312  degrees  of  freedom
      AIC:151.23
```

→ With animacy included:

```
      Call:

            glm(formula = PAT ~ ArgAnim + ArgRlz + ArgType * ArgTP1, family = "binomial", data = AltDS)


Deviance Residuals:
           Min        1Q    Median        3Q      Max
      -1.49882  -0.32946  -0.11467  -0.00004  3.01394

Coefficients:
                       Estimate       Std.   Error  z value  Pr(>|z|)
      (Intercept)     -11.72125  728.34793  -0.016   0.9872
      ArgAnim1          9.95361  728.34785   0.014   0.9891
      ArgRlz1          -0.96356    0.54895  -1.755   0.0792         .
      ArgRlz2          -0.24204    0.45390  -0.533   0.5939
      ArgType1         -0.98320    0.42878  -2.293   0.0218  *
      ArgTP1           -0.05081    0.09428  -0.539   0.5899
      ArgType1:ArgTP1  -0.11255    0.09677  -1.163   0.2448
      ---
      Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
          Null  deviance:  165.12  on  317  degrees  of  freedom
      Residual  deviance:  100.30  on  311  degrees  of  freedom
      AIC:114.3
```

→ Replacing ArgRlz by a binary variable: Overt(NP+Pro) = +1 vs. Null = −1

```
Call:

      glm(formula = PAT ~ Overt + ArgType * ArgTP1, family = "binomial", data = AltDS)

Deviance Residuals:
          Min       1Q   Median       3Q     Max
      -1.1369  -0.4089  -0.2200  -0.1907  2.9378

Coefficients:
                      Estimate     Std.   Error   z value  Pr(>|z|)
      (Intercept)     -1.89016  0.40679  -4.646  3.38e-06  ***
      Overt1          -1.86227  0.52305  -3.560   0.00037  ***
      ArgType1        -0.04295  0.37555  -0.114   0.90895
      ArgTP1           0.12446  0.08509   1.463   0.14359
      ArgType1:ArgTP1 -0.22576  0.08811  -2.562   0.01040  *
      ---
      Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
          Null  deviance:  165.12  on  317  degrees  of  freedom
      Residual  deviance:  139.27  on  313  degrees  of  freedom
      AIC:149.27
```

```
→ With animacy included:
Call:
     glm(formula = PAT ~ ArgTP1 * ArgType + Overt + ArgAnim, family = "binomial", data = AltDS)

Deviance Residuals:
          Min        1Q    Median        3Q      Max
     -1.51292  -0.32861  -0.13168  -0.00004  2.86970

Coefficients:
                      Estimate       Std.   Error  z value  Pr(>|z|)
     (Intercept)     -10.47810  721.93744  -0.015  0.98842
     ArgTP1           -0.04691    0.09374  -0.500  0.61682
     ArgType1         -0.99751    0.42697  -2.336  0.01948  *
     Overt1           -1.73028    0.58635  -2.951  0.00317  **
     ArgAnim1          9.90777  721.93737   0.014  0.98905
     ArgTP1:ArgType1  -0.11370    0.09640  -1.180  0.23818
     ---
     Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
         Null  deviance:  165.12  on  317  degrees  of  freedom
     Residual  deviance:  100.94  on  312  degrees  of  freedom
     AIC: 112.94
```

C. Binary fixed-effects models – fitted to matching verbs only (Section 4.4)

1. *Zero*

Same dataset as above.

2. */e/-series*

```
163 tokens
Distribution:
     1   53
     0  110

Model selection:
     GEN ~ ArgType + ArgAnim + ArgRlz + ArgTP1
              Df  Deviance     AIC     LRT  Pr(>Chi)
     <none>         192.46  204.46
     ArgType   1    199.98  209.98  7.5229  0.006092  **
     ArgAnim   1    194.86  204.86  2.4068  0.120812
     ArgRlz    2    199.07  207.07  6.6176  0.036560  *
     ArgTP1    1    192.46  202.46  0.0083  0.927296

     GEN ~ ArgType + ArgRlz
              Df  Deviance     AIC     LRT  Pr(>Chi)
     <none>         195.59  203.59
     ArgType   1    200.25  206.25  4.6540   0.03098  *
     ArgRlz    2    201.41  205.41  5.8115   0.05471  .

Summary of results:
Call:
     glm(formula = GEN ~ ArgType + ArgRlz, family = "binomial", data = kdvge)

Deviance Residuals:
         Min       1Q   Median      3Q     Max
     -1.2910  -0.9137  -0.8498  1.1154  1.9561
```

```
Coefficients:
                  Estimate    Std.    Error   z value   Pr(>|z|)
     (Intercept)    -0.6011   0.2190   -2.745   0.00606  **
     ArgType1       -0.4604   0.2128   -2.163   0.03053  *
     ArgRlz1         0.2883   0.2508    1.149   0.25041
     ArgRlz2        -0.6921   0.3101   -2.232   0.02564  *
     ---
     Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
         Null  deviance:  205.61  on  162  degrees  of  freedom
     Residual  deviance:  195.60  on  159  degrees  of  freedom
     AIC: 203.6
```

3. /o/-series

```
139 tokens
Distribution:
     1  46
     0  93

Model selection:
     LOC ~ ArgAnim * ArgTP1 + ArgRlz
                    Df  Deviance    AIC      LRT  Pr(>Chi)
     <none>               132.95  144.95
     ArgRlz          2    133.84  141.84  0.88811    0.6414
     ArgAnim:ArgTP1  1    133.89  143.89  0.94188    0.3318

     LOC ~ ArgAnim + ArgTP1 + ArgRlz
             Df  Deviance    AIC      LRT   Pr(>Chi)
     <none>        133.89  143.89
     ArgAnim  1    145.44  153.44  11.5538  0.0006761  ***
     ArgTP1   1    137.41  145.41   3.5199  0.0606343    .
     ArgRlz   2    134.81  140.81   0.9204  0.6311663

     LOC ~ ArgAnim * ArgTP1
                    Df  Deviance    AIC      LRT  Pr(>Chi)
     <none>               133.84  141.84
     ArgAnim:ArgTP1  1    134.81  140.81  0.97414    0.3236

     LOC ~ ArgAnim + ArgTP1
             Df  Deviance    AIC     LRT   Pr(>Chi)
     <none>        134.81  140.81
     ArgAnim  1    150.26  154.26  15.449  8.475e-05  ***
     ArgTP1   1    138.35  142.35   3.540     0.0599    .

Summary of results:
Call:
     glm(formula = LOC ~ ArgAnim + ArgTP1, family = "binomial", data = kdvlo)

Deviance Residuals:
         Min       1Q   Median      3Q     Max
     -1.3017  -0.9107  -0.2786  1.0580  2.4651

Coefficients:
                  Estimate    Std.    Error    z value   Pr(>|z|)
     (Intercept)    -0.8706   0.4027   -2.162   0.030618  *
     ArgAnim1       -1.1581   0.3453   -3.354   0.000797  ***
     ArgTP1         -0.2401   0.1342   -1.789   0.073594         .
     ---
     Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
```

```
    Null  deviance:  176.49  on  138  degrees  of  freedom
Residual  deviance:  134.81  on  136  degrees  of  freedom
AIC:140.81
```

4. */a/-series*

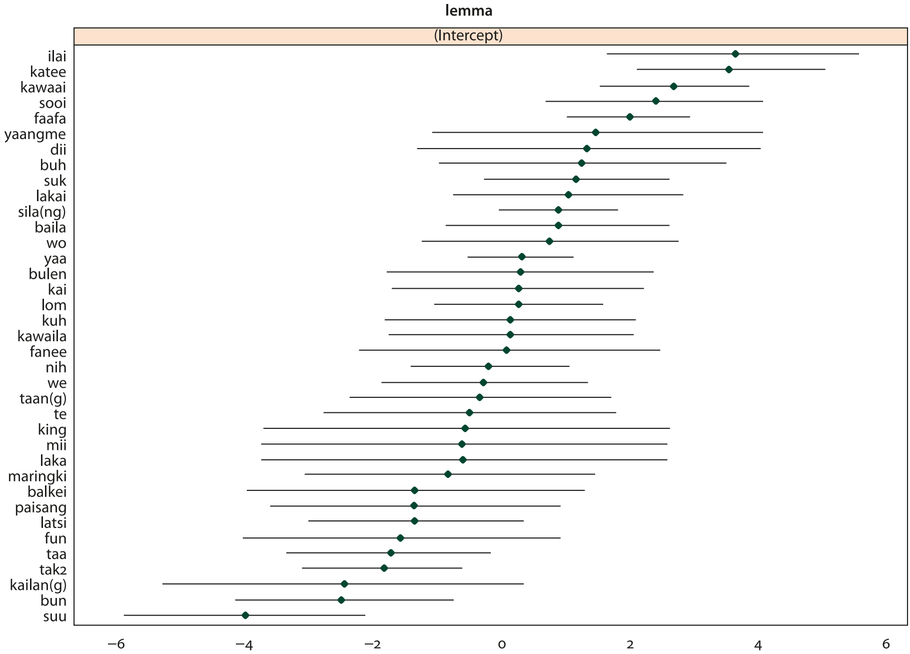
```
60 tokens

Distribution:
     1  23
     0  37

Model selection:
     PAT ~ ArgType * ArgTP1 + ArgRlz + ArgAnim
                     Df  Deviance     AIC      LRT   Pr(>Chi)
     <none>                22.008  36.008
     ArgRlz           2    23.931  33.931   1.9231     0.3823
     ArgAnim          1    53.101  65.101  31.0937  2.459e-08  ***
     ArgType:ArgTP1   1    22.937  34.937   0.9295     0.3350

     PAT ~ ArgType + ArgAnim + ArgTP1 + ArgRlz
              Df  Deviance     AIC      LRT   Pr(>Chi)
     <none>         22.937  34.937
     ArgType   1    23.471  33.471   0.5336     0.4651
     ArgAnim   1    53.923  63.923  30.9860  2.599e-08  ***
     ArgTP1    1    23.097  33.097   0.1598     0.6894
     ArgRlz    2    25.151  33.151   2.2142     0.3305

     PAT ~ ArgType * ArgTP1 + ArgAnim + Overt
                     Df  Deviance     AIC      LRT   Pr(>Chi)
     <none>                22.329  34.329
     ArgAnim          1    53.392  63.392  31.0633  2.497e-08  ***
     Overt            1    23.931  33.931   1.6020     0.2056
     ArgType:ArgTP1   1    23.550  33.550   1.2207     0.2692

     PAT ~ ArgType * ArgTP1 + Overt
                     Df  Deviance     AIC      LRT   Pr(>Chi)
     <none>                53.392  63.392
     Overt            1    72.258  80.258  18.8656  1.403e-05  ***
     ArgType:ArgTP1   1    53.969  61.969   0.5766     0.4476

     Model 1: PAT ~ Overt
     Model 2: PAT ~ Overt + ArgTP1
        Resid. Df  Resid. Dev  Df  Deviance  Pr(>Chi)
     1         58      62.089
     2         57      54.446   1    7.6427    0.0057  **

     Model 1: PAT ~ Overt * ArgTP1
     Model 2: PAT ~ Overt + ArgTP1
        Resid. Df  Resid. Dev  Df  Deviance  Pr(>Chi)
     1         56      54.045
     2         57      54.446  -1  -0.40078    0.5267

Summary of results:
Call:
     glm(formula = PAT ~ Overt + ArgTP1, family = “binomial”, data = kdvp0)

Deviance Residuals:
         Min       1Q   Median      3Q     Max
     -1.9917  -0.6049  -0.3433  0.6531  2.0643

Coefficients:
                     Estimate     Std.    Error   z value   Pr(>|z|)
     (Intercept)       0.2409   0.5555    0.434    0.6646
     Overt1           -3.0424   0.7795   -3.903   9.5e-05  ***
     ArgTP1            0.3986   0.1628    2.449    0.0143  *
     ---
     Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 79.881 on 59 degrees of freedom

Residual deviance: 54.446 on 57 degrees of freedom

AIC: 60.446
```

D. Binary mixed-effects models – fitted to matching verbs only (Section 4.4)

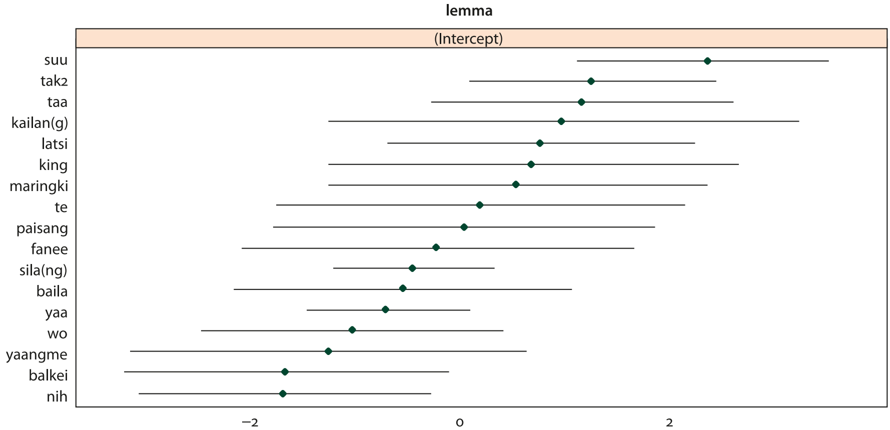
1. *Zero*

```
Summary of results:

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']

 Family: binomial ( logit )

Formula: Z ~ ArgType + ArgAnim + ArgRlz + (1 | lemma)

   Data: AltDS
      AIC    BIC  logLik  deviance  df.resid
    327.7  350.2  -157.8     315.7       312

Scaled residuals:
       Min       1Q  Median      3Q     Max
   -4.4867  -0.5172  0.2249  0.4466  9.1454

Random effects:
     Groups Name         Variance  Std.Dev.
     lemma (Intercept)   4.041     2.01
    Number of obs: 318,  groups:   lemma, 37

Fixed effects:
                 Estimate   Std. Error   z value   Pr(>|z|)
    (Intercept)    0.39779      0.41304     0.963    0.33550
    ArgType1       2.02207      0.35086     5.763   8.25e-09  ***
    ArgAnim1      -0.66599      0.25797    -2.582    0.00983  **
    ArgRlz1        0.99825      0.34394     2.902    0.00370  **
    ArgRlz2       -0.07045      0.26731    -0.264    0.79212
    ---
    Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
              (Intr)  ArgTy1  ArgAn1  ArgRl1
    ArgType1   0.034
    ArgAnim1  -0.103  -0.497
    ArgRlz1    0.256   0.223  -0.225
    ArgRlz2   -0.215   0.026   0.258  -0.646
```



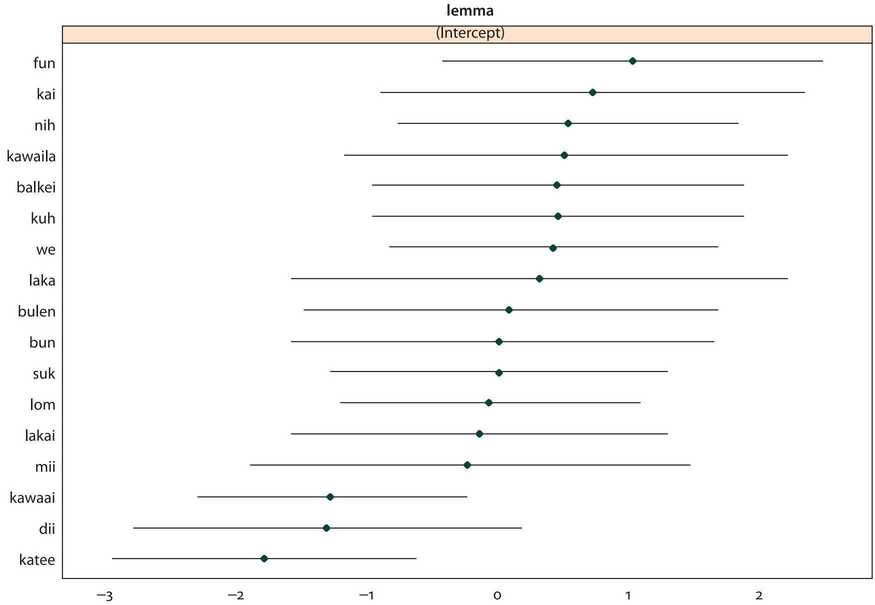
2. */e/-series*

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial ( logit )
Formula: GEN ~ ArgType + ArgRlz + (1 | lemma)
    Data: kdvge
     AIC     BIC  logLik  deviance  df.resid
   186.6   202.1   -88.3     176.6       158

Scaled residuals:
        Min       1Q   Median      3Q     Max
   -1.5274  -0.5725  -0.3501  0.6547  6.9489

Random effects:
       Groups Name          Variance   Std.Dev.
       lemma (Intercept)    1.959      1.4
      Number of obs: 163,   groups:    lemma, 17

Fixed effects:
                     Estimate   Std. Error   z value   Pr(>|z|)
      (Intercept)     -0.2934       0.4588    -0.640    0.52247
      ArgType1        -0.9057       0.3336    -2.715    0.00663  **
      ArgRlz1          0.2180       0.3128     0.697    0.48597
      ArgRlz2         -0.9981       0.3996    -2.498    0.01249  *
      ---
      Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
                (Intr)  ArgTy1  ArgRl1
      ArgType1  -0.302
      ArgRlz1   -0.163   0.070
      ArgRlz2    0.133   0.252  -0.604
```



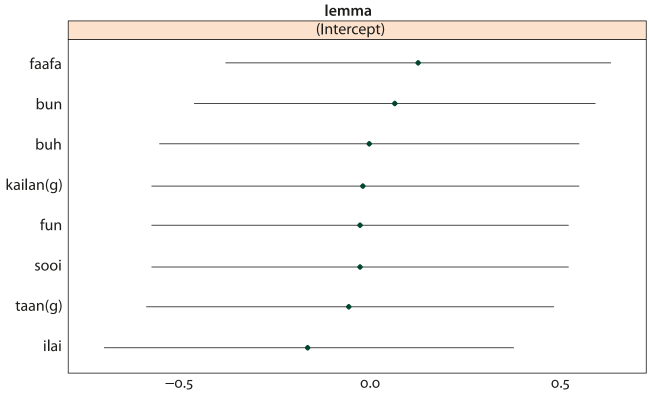
3. */o/-series*

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial ( logit )
Formula: LOC ~ ArgAnim + ArgTP1 + (1 | lemma)
   Data: kdvl0
      AIC     BIC   logLik  deviance  df.resid
    133.0   144.7    -62.5     125.0       135

Scaled residuals:
        Min       1Q   Median      3Q     Max
    -1.7328  -0.3953  -0.1760  0.5771  7.5150

Random effects:
      Groups Name         Variance  Std.Dev.
      lemma (Intercept)   1.17      1.082
    Number of obs: 139,   groups:   lemma, 17

Fixed effects:
                  Estimate  Std. Error  z value  Pr(>|z|)
    (Intercept)    -0.8109      0.5538   -1.464  0.143148
    ArgAnim1       -1.4448      0.4137   -3.492  0.000479  ***
    ArgTP1         -0.2493      0.1682   -1.482  0.138398
    ---
    Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
                (Intr)  ArgAn1
    ArgAnim1     0.648
    ArgTP1      -0.589  -0.417
```



4. */a/-series*

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial ( logit )
Formula: PAT ~ Overt + ArgTP1 + (1 | lemma)
    Data: kdvp0
      AIC    BIC  logLik  deviance  df.resid
     62.4   70.8   -27.2      54.4        56

Scaled residuals:
        Min       1Q   Median      3Q     Max
    -2.4146  -0.4228  -0.2504  0.5000  2.5958

Random effects:
      Groups Name        Variance  Std.Dev.
      lemma (Intercept)  0.08003   0.2829
     Number of obs: 60,  groups:   lemma, 8

Fixed effects:
                  Estimate   Std. Error   z value   Pr(>|z|)
     (Intercept)    0.2197       0.5841     0.376   0.706874
     Overt1        -3.0578       0.7943    -3.850   0.000118  ***
     ArgTP1         0.3959       0.1643     2.409   0.015995  *
     ---
     Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
             (Intr)  Overt1
     Overt1  -0.347
     ArgTP1  -0.387  -0.471
```



E. Goodness-of-fit of binomial models

1. *Zero*

```
R2 values:
	FE Model
	r.squaredLR(BM1_Z)
	[1] 0.1398526
	attr(,"adj.r.squared")
	[1] 0.1900282
	ME Model
	r.squaredGLMM(BM2_Z)
	                    R2m       R2c
	theoretical 0.3216779 0.6955911
	delta       0.3016059 0.65218

Null Model:
	BM0_Z=glm(Z~1, family="binomial" , data=AltDS)

Observed/predicted contingency tables:
	Null Model
	      1
	0   122
	1   196

	FE Model
	    0    1
	0  41   81
	1  24  172

	ME Model
	    0    1
	0  99   23
	1  19  177
```

2. */e/-series*

```
R2 values:
```

```
FE Model 1
r.squaredLR(BM1_GEN)
[1] 0.05685856
attr(,"adj.r.squared")
[1] 0.09573979
FE Model 2
r.squaredLR(BM2_GEN)
[1] 0.05956831
attr(,"adj.r.squared")
[1] 0.083110
ME Model
r.squaredGLMM(BM3_GEN)
                       R2m        R2c
     theoretical  0.1485881  0.4663572
     delta        0.1296784  0.4070072
```

```
Null Models:
     BM0_GEN=glm(GEN~1, family="binomial" , data=AltDS)
     BM02_GEN=glm(GEN~1, family="binomial" , data=kdvg0)
```

```
Observed/predicted contingency tables:
     Null Model 1
              0
          0  265
          1   53

     FE Model 1
              0
          0  265
          1   53

     Null Model 2
              0
          0  110
          1   53

     FE1 Mode 2
              0    1
          0  100   10
          1   40   13

     ME Model
             0    1
          0  99   11
          1  18   35
```

3. /o/-series

```
R2 Values:
     FE Model 1
     r.squaredLR(BM1_LOC)
     [1] 0.202516
     attr(,"adj.r.squared")
     [1] 0.3600376
```

FE Model 2

r.squaredLR(BM2_LOC)

[1] 0.2590459

attr(,"adj.r.squared")

[1] 0.3602474

ME Model

r.squaredGLMM(BM3_LOC)

```
                   R2m       R2c
theoretical  0.4508898  0.5949573
delta        0.4079808  0.5383381
```

Null Models:

BM0_LOC=glm(LOC~1, family="binomial" , data=AltDS)

BM02_LOC=glm(LOC~1, family="binomial" , data=kdvl0)

Observed/predicted contingency tables:

Null Model 1

```
     0
0  272
1   46
```

FE Model 1

```
     0
0  272
1   46
```

Null Model 2

```
    0
0  93
1  46
```

FE1 Mode 2

```
    0   1
0  61  32
1  11   3
```

ME Model

```
    0   1
0  81  12
1   9  37
```

4. /a/-series

R2 Values:

FE Model 1

r.squaredLR(BM1_PAT)

[1] 0.07806917

attr(,"adj.r.squared")

[1] 0.1927519

FE Model 2

r.squaredLR(BM2_PAT)

[1] 0.3455202

attr(,"adj.r.squared")

[1] 0.4695346

```
ME Model
r.squaredGLMM(BM3_PAT)
                    R2m       R2c
theoretical  0.4716706  0.4842174
delta        0.4054077  0.4161919
```

```
Null Models:
BM0_PAT=glm(PAT~1, family="binomial" , data=AltDS)
BM02_PAT=glm(PAT~1, family="binomial" , data=kdvp0)
```

```
Observed/predicted contingency tables:
Null Model 1
       0
  0  295
  1   23

FE Model 1
       0  1
  0  295  0
  1   22  1

Null Model 2
      0
  0  37
  1  23

FE1 Mode 2
      0   1
  0  31   6
  1   6  17

ME Model
      0   1
  0  30   7
  1   6  17
```

Appendix B. Verb frequencies

I. Alternating verbs

			Frequency per marking strategy						
Verb	**Gloss**	**Alternation profile**	**zero**	**/a/-series**	**/o/-series**	**/e/-series**	**other: 'DAT'**	**other: 'SBEN'**	**Total**
yaa	reach	/e/+zero	40	–	–	8	–	–	48
silang	descend; descend into	/e/+zero	19	–	–	15	–	–	34
wo	exist	/e/+zero	12	–	–	1	–	–	13
tak	run; get up and run	/e/+zero	3	–	–	8	–	–	11
yaangme	come down	/e/+zero	9	–	–	1	–	–	10
latsi	stand; stand up	/e/+zero	3	–	–	3	–	–	6

I. *(continued)*

			Frequency per marking strategy						
Verb	Gloss	Alternation profile	zero	/a/-series	/o/-series	/e/-series	other: 'DAT'	other: 'SBEN'	Total
taa	sleep	/e/+zero	2	–	–	4	–	–	6
baila	buy	/e/+zero	2	–	–	2	–	–	4
paisang	bright	/e/+zero	2	–	–	1	–	–	3
maringki	startle	/e/+zero	1	–	–	2	–	–	3
fanee	work	/e/+zero	2	–	–	1	–	–	3
te	go up straight; raise up	/e/+zero	1	–	–	1	–	–	2
kailang	sorry; angry	/e/+/a/	–	1	–	1	–	–	2
balkei	sell	/e/+/o/	–	–	4	1	–	–	5
nih	sit	/e/+/o/+zero+other	8	–	4	1	–	2	15
king	thought	/e/+other	–	–	–	2	1	–	3
suu	go ahead	/e/+other	–	–	–	9	–	2	11
lom	say	/o/+zero	19	–	4	–	–	–	23
kawaai	say something; speak	/o/+zero	16	–	3	–	–	–	19
katee	eat	/o/+zero	15	–	1	–	–	–	16
suk	think	/o/+zero	4	–	6	–	–	–	10
kuh	close up	/o/+zero	2	–	4	–	–	–	6
kawaila	circle around; stumble	/o/+zero	2	–	2	–	–	–	4
lakai	shave off	/o/+zero	2	–	2	–	–	–	4
kai	cheer	/o/+zero	2	–	2	–	–	–	4
bulen	write on; carve on	/o/+zero	1	–	2	–	–	–	3
laka	hang; hang on	/o/+zero	1	–	1	–	–	–	2
dii	lie; agree on	/o/+zero+other	8	–	1	–	–	1	10
we	give back; redeem; go	/o/+zero+other	4	–	5	–	–	1	10
fun	work; touch	/o/+/a/	–	1	6	–	–	–	7
bun	hide	/o/+/a/+zero	2	8	1	–	–	–	11
mii	set; sit (inanimate)	/o/+other	–	–	1	–	–	1	2
faafa	look for	/a/+zero	11	9	–	–	–	–	20
ilai	look	/a/+zero	16	1	–	–	–	–	17
sooi	ask	/a/+zero	6	1	–	–	–	–	7

I. *(continued)*

			Frequency per marking strategy						
Verb	**Gloss**	**Alternation profile**	**zero**	**/a/-series**	**/o/-series**	**/e/-series**	**other: ‘DAT’**	**other: ‘SBEN’**	**Total**
taang	descend; descend from	/a/+zero	3	3	–	–	–	–	6
buh	lift; cradle; pick up	/a/+zero	1	2	–	–	–	–	3
koo	stay	zero+other	1	–	–	–	–	2	3
baa	say to	other+other	–	–	–	–	2	11	13

II. Non-alternating verbs (including verbs attested once only)

Verb	Gloss	Total	Verb	Gloss	Total
/a/-series			zero		
tak	see	23	*see*	arrive	33
wai	return	17	*maa*	walk	19
oo	be born; give birth to	13	*lak*	know	11
tau	meet	11	*mu'tan*	fall from above	10
mai	not want; refuse	7	*mota*	throw implement	9
maaung	want	6	*me*	come	8
ah	eat; feed	5	*puisi*	crack	8
wei	wash	4	*tewe*	enter	7
serang	get up	4	*lila*	fly	7
tom	pay for	3	*ite*	put inside	6
maa	sound	2	*wehe*	go down	6
suma	compare; try	2	*tanei*	cry	6
maitang	hungry	2	*pilan*	lego-lego (dance)	6
bo'ra	die	2	*fanenee*	chase after	5
beh	order	2	*met*	take	4
yok	shake	2	*eisin*	ask	4
kut	stab	1	*taam*	cook	4
rot	cut off	1	*dok*	move	4
toka	peck	1	*bau*	crack canarium nuts	3
saakda	die of old age	1	*awila*	full	3
sui	bother	1	*mauu*	war	3
lai2	delight in	1	*nika*	marry [Malay]	3
beei	can; able	1	*keeka*	tear	3
tan	be in contact with	1	*suboita*	leap	3
pa	have fun	1	*paan*	kill	3
tara	store in; save in	1	*wete*	go up	3
tafanee	receive	1	*wee*	use	2
toota	lay crosswise	1	*silik*	strip	2
			metaang	come up	2

II. *(continued)*

Verb	Gloss	Total	Verb	Gloss	Total
/o/-series			*kangka*	whine	2
ra	carry	15	*lai*	finish	2
iti	call out	9	*tabei*	promise	2
baa	make	5	*taangme*	come up	2
o	follow	4	*ne*	drink	2
sak	dry in sun	3	*liwang*	be clear; explain oneself	2
um	teach; discuss	3	*paliin*	squash down	2
simai	??	2	*isu*	draw water	1
nok	one	2	*buka*	open [Malay]	1
tawaah	assemble	2	*taak*	cut	1
pakah	carry weight	2	*tulen*	share	1
aakai	trap; trick	2	*luh*	wipe	1
suusa	difficulty	1	*milai*	be angry at someone	1
tee	cover	1	*saawan*	count	1
kata	squeezed in	1	*suh*	collide	1
yaka	take care of children	1	*diit*	pop out	1
ileh	drag	1	*tumung*	up there anywhere	1
pok	bundle	1	*lulus*	graduate	1
wal	sing	1	*letla*	kick	1
paipa	swoop	1	*ide*	roast	1
suksiteh	desire	1	*loo*	walk	1
su	be three	1	*menyelai*	graduate [Malay]	1
			fasi	swell	1
/e/-series			*mitan*	complete	1
n	give to	18	*bui*	cut down; fell	1
n	marry	6	*taaliki-*	carry cheering	1
sulee	continue	3	*taalawee*		
biee	afraid	2	*tafe*	plant seed	1
kaang	be present	2	*taitin*	wiggle	1
gai	associate with	1	*ai*	catch	1
wet	exchange	1	*sau*	not know	1
mai	hear	1	*lamu*	swallow	1
koh	make a hole	1	*tooning*	be like	1
lil ka	be firm	1	*mat*	be sick	1
sika	finished	1	*wal*	be	1
other: ‘DAT’			other: ‘SBEN’		
sah	conclude; block	4	*wofaafa*	be surprised	4

Address for correspondence

Katherine Walker
University of Amsterdam
P.C. Hoofthuis
Spuistraat 134
1012VB Amsterdam
The Netherlands
k.walker@uva.nl
https://orcid.org/0000-0002-5423-7308

Co-author information

Pegah Faghiri
University of Amsterdam
Amsterdam, The Netherlands
p.faghiri@uva.nl
https://orcid.org/0000-0001-9121-6330

Eva van Lier
University of Amsterdam
Amsterdam, The Netherlands
e.h.vanlier@uva.nl

Publication history

Date received: 25 November 2021
Date accepted: 21 October 2022
Published online: 27 June 2023