Directional constructions in Matukar Panau

A Bayesian approach to assessing variation

Kira Davey and Danielle Barth
Australian National University

The Oceanic language Matukar Panau has three equivalent morphosyntactic strategies for describing the direction of the event represented by a verb, with a system of ten directional morphemes that can appear in each construction. This variation is explored using a corpus from Matukar Panau and analysed quantitatively with Bayesian regression analyses to assess what factors influence the choice of directional construction. Variables pertaining to the lexical verb and the directional morphemes are found to be the most important factors affecting the variation in the directional system. Sociolinguistic factors are shown to play a less significant role. The findings have implications for the grammaticalization of directional elements, as well as the typology of directional constructions in Oceanic languages.

Keywords: Bayesian regression, Oceanic, directionals, serial verbs, morphosyntactic variation, grammaticalization, lexicalization, corpus linguistics

1. Introduction

Expressing direction is a fundamental component of language use. Languages vary in what grammatical elements they use to describe event direction, with main verbs, non-verbal elements and verb series well-attested cross-linguistically (Ameka & Essegbey, 2013; Beavers, Levin, & Tham, 2010; Talmy, 2000). The Oceanic language Matukar Panau has ten directional elements, which appear as main verbs and in three functionally equivalent constructions (1–3) that serve to express the direction of a verb's event (Barth & Anderson, 2015).

Directional – directional suffix following main verb

https://doi.org/10.1075/aplv.23005.dav | Published online: 11 January 2024
Asia-Pacific Language Variation 9:2 (2023), pp. 156–194. ISSN 2215-1354 | E-ISSN 2215-1362

(1) *…bom main di-tarutiti-**pid**-din-e*
sago TOP 3PL.S-cut-**DWN**-3PL.P-D.SEQ
'They cut down the sago…'
Clara Kusos – 50pics14-ckd_jm: 1477050 ms–1479610 ms

SVC1 – core verb serialization, V2 expresses direction

(2) *Gaun nub ilonlo **y-ep** **pid-ago**.*
Dog water inside **3SG.S-look descend-I.R.IPFV**
'The dog is looking down into the water.'
Kadagoi Rawad Forepiso – Kadagoi_Rawad_Forepiso_Frog_Story: 49449 ms–54082 ms

SVC2 – core verb serialization with erstwhile dependency[1] marking, V2 expresses direction

(3) *Ong **so-p** **pid-tap***
2SG.S **come-D.IRR descend-D.IF.COND**
'If you come down…'
Kadagoi Lovinea Rapalau Ambrose – song01_burar_awante_klra: 31124 ms-39082 ms

This directional system is unusually elaborated morphosyntactically in comparison with other Oceanic languages, where languages tend to have either non-verbal directional elements or directional verb series (Pawley, 2003). Quantitatively assessing which of the three constructions is used for what purpose is complex. Variationist linguistics has shown that we may expect influences from individual behaviour, social factors, lexical factors and grammatical factors (Meyerhoff, 2015; Tagliamonte & Smith, 2006; Tagliamonte & Baayen, 2012 *inter alia*). We use Bayesian multinomial modelling (Levshina, 2016, 2019; Grafmiller, Szmrecsanyi, & Hinrichs, 2018; Dilley, Gamache, Wang, Houston, & Bergeson, 2019; MacKenzie, 2020; Nalborczyk, Batailler, Lœvenbruck, Vilain, & Bürkner, 2019; Nicenboim & Vasishth, 2016) to assess not only which factors play a role, but also which construction types are more similar in their use. We assess whether the variation in directional construction choice is based on categorically determined rules, that is, 'rule-based' (cf. Guy, 1988), random free variation, or probabilistically determined. We find that there are clear probabilistic patterns in construction choice, with low variation across speakers, indicating that speakers

1. The erstwhile dependent marker is termed such as it used to indicate a dependent and chained relationship between the verbs but is no longer meaningful. The two clauses should be analysed as a SVC rather than a clause chain with medial marking (cf. Barth & Anderson, 2015).

are tracking co-occurrence patterns, particularly of which verb and directional element combinations are most likely to occur with each construction type. A complete description of a language's grammar should not stop at the options for expressing information, but should also account for why we see the variation we do. To that end, this paper describes the directional constructions after presenting background on the Matukar Panau language. In the methodology section, we cover the factors we assess and the justification for Bayesian modelling. The results and discussion follow.

1.1 Matukar Panau language background

Matukar Panau is spoken in the villages of Matukar and Surumurang on the North-East coast of Papua New Guinea. These villages are located around 45km north of Madang and can be seen in the map in Figure 1, represented by the red location marker.

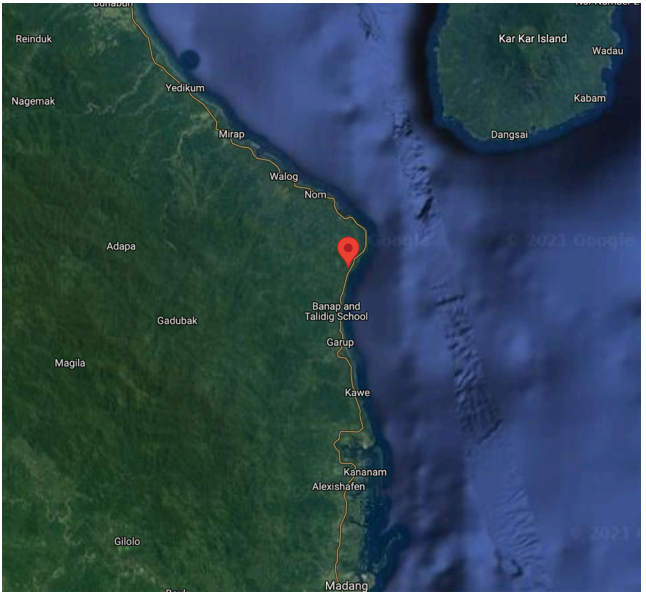


Figure 1. Map of Matukar, Papua New Guinea (Sourced from maps.google.com)

The language is highly endangered with around 300 speakers. The majority of people are under 40 years old and their first and dominant language is Tok Pisin. They largely do not speak much Matukar Panau. The next largest group of speakers are aged between 40 and 60 and their first language is Matukar Panau, but their dominant language is Tok Pisin. Lastly, the smallest group of speakers are

aged over 60 and their first and dominant language is Matukar Panau, but they also frequently speak Tok Pisin.

Matukar Panau speakers are highly multilingual, often speaking multiple Papuan and Austronesian languages, in addition to Tok Pisin. Many people speak the Papuan languages Bargam and Waskia, which are spoken in nearby areas, as well as the closely related Austronesian language Takia. In total, speakers in the Matukar Panau corpus (Barth, 2010) speak a total of around 20 different languages, many of which are under described.

Genealogically, Matukar Panau is an Oceanic language of the Bel family and belongs to the Western Bel subgroup (Barth & Ross, in press). Bel languages display morphosyntactic features that are unusual for Oceanic languages, such as verb-final word order and clause chaining. These innovations are a result of Papuan bilingualism, where Bel grammatical features have been restructured based on a Papuan model (Ross, 2008). Clause chains consist of multiple clauses that are chained together to form one sentence, typically containing one or more medial verbs with an independent verb in the final clause (Foley, 2010; Pawley & Hammarström, 2017; Longacre, 2007). Semantically, Matukar clause chains consist of multiple simultaneous, sequential or overlapping events (Barth & Anderson, 2015; Barth & Ross, in press). For example, in (4) we see a series of sequential events in a chain.

(4) *Alo nagi-yau-we ngau nga-mado-nge, ngau tamat ngale-nge.*
later give.birth-1SG.P-SEQ:D 1SG 1SG.S-stay-SEQ:D 1SG man marry-R:I:PFV.
'Later she gave birth to me, I stayed and I got married.'
Margaret Lem Kaluk – Life_Story_of_Margaret_Lem_Kaluk: 43509 ms–47628 ms

Clause chaining is a typical Papuan construction (Foley, 2010; Ross, 2008; Pawley & Hammarström, 2017) and an extremely productive method of linking clauses in Matukar Panau (Barth & Anderson, 2015). These innovations are at least partially responsible for Matukar Panau's extensive directional system.

The Matukar Panau language documentation project is ongoing since 2010.[2] There is currently more than 60 hours of video data recorded, much of which has been transcribed and translated into English and/or Tok Pisin (190,000+ words) from close to 100 speakers.

2. There is a website with more details for those interested: matukar.wordpress.com as well as archives at http://hdl.handle.net/2196/00-0000-0000-0012-388F-3, https://catalog.paradisec.org.au/collections/DGB1, and parts of https://catalog.paradisec.org.au/collections/SocCog

1.2 Matukar directional constructions

Matukar Panau is a nominative-accusative language that has a verbal complex with obligatory prefixal subjects and a three-part system of TAM suffixes. TAM information is always indexed, except in the case of non-final verbs in core serial verb constructions (see SVC1 constructions). There are three paradigmatic sets of TAM suffixes which cannot co-occur. These are for (i) independent verbs, (ii) dependent verbs, and (iii) subordinate verbs. Independent verbs appear in simple clauses or as the final verb in the last chain of multiple clauses containing dependent verbs. Verbs marked with dependent medial morphology appear only in these clause chains (cf. Example 5, Example 14). Verbs marked with subordinating morphology appear in conditional or adversative clauses (cf. Example 3, Example 9). The morphotactics of the Matukar verbal complex are shown in Table 1, along with the affixal morphology. Where affixal constituents are in parentheses, they are optionally affixed to the verb. Argument NPs are optionally expressed. Matukar has variable constituent order with SOV preferred, although usually transitive clauses will only have one or no expressed NP argument. Noun phrase and pronoun arguments are optionally expressed (cf. Example 10 vs Example 13).

Table 1. The verbal complex and affixes

Subject-	Verb	(-Directionals)	-Object*	-TAM
				INDEPENDENT
1SG nga-		-(y)a AND	-au 1SG	Ø I.R, I.IRR.IMP
2SG wa-/ø-		-(a)la AND	-o 2SG	-e~-nge~-we I.R.PFV
3SG ya-/ø-		-(a)pid DWN	-i/-ø 3SG	-go I.R.IPFV
1PL.EXCL ngam-		-(a)si DWN	-amam 1PL.EXCL	-gokai I.R.IPFV.HAB
1PL.INCL ta-		-(a)das UPW	-ad 1PL.INCL	-ba I.IRR
2PL a-		-(a)mais UPW	-ang 2PL	-bawai I.IRR.DESID
3PL di-		-(a)sa UPW	-din 3PL	SUBORDINATE
		-(a)mul VEN	(-ai) 3PL.PART	-dope D.WHEN.COND
		-(a)palum VEN		-tape D.IF.COND
		-(a)so VEN		-kai D.ADVS
				MEDIAL
				-do D.R
				-e~-nge~-we D.SEQ
				-ma D.HAB
				-p~-dop D.IRR

* 'Object' in this instance refers to both the direct object (patient) of a transitive verb and the indirect object (recipient) of a ditransitive verb. Both the patient and recipient can be indexed on ditransitive verbs, but often object suffixes are omitted in discourse.

As previewed above, Matukar Panau has three different construction types to express the direction of a verbally expressed concept.

Directional constructions

In directional constructions, directional morphemes are suffixed to the verb and fill the slot immediately after the verb, appearing before any object or TAM suffixes. Directional suffix forms are identical to directional verb forms, so the presence/absence of other morphemes help determine whether they are suffixes or verbs. As will be seen later, there is sometimes ambiguity as to their affixal/verbal status.

Directional suffixes can be suffixed to both transitive (5, 6) and intransitive verbs (7). In (5) the geocentric directional *-das* characterizes the movement of the caused motion verb *nage* as upwards. In (6) the venitive deictic suffix *-so*, denotes the direction of the change of possession verb *ngale* as towards the speaker. Object suffixes are often omitted when a directional is present, as in (6), but the object suffix may also be omitted here due to it being third person singular. Example (7) shows the andative suffix *-la* on the intransitive autonomous motion verb *si* describing the direction of the motion as away from a previously established deictic centre.

(5) *Nub wananan* ***nga-nage-das-e.***
water hot **1SG.S put-UPW-D.SEQ**
'I put up/brough up hot water.'
Rebecca Willie – Rebecca_Willie_Saksak: 249193 ms–251436 ms

(6) ***Ngale-so-p*** *tai-te-p…*
take-VEN-IRR:D 1PL.INCL.S-see-D.IRR
'bring it and we'll look at it…'
Kadagoi Rawad Forepiso – SocCog-mjk01-krf_spw_3: 353155 ms–354424 ms

(7) *ab-ate* ***nga-si-la-nge***
house-LOC **1SG.S-descend-AND-I.R.PFV**
'I went home.'
Maria Saiyak Kaluk – DGB1-daily_life11-st_msk: 26525 ms–31146 ms

SVC1 constructions

Matukar Panau has SVCs that consist of two or more component verbs that comprise one complex event (Barth & Anderson, 2015). The verbs are combinatory and compositional. That is, they are semantically compatible and the meaning of the SVC is transparent from the constituent verbs. Matukar SVCs are contained within a single clause, verbs share TAM information, there is no

marker of coordination, subordination or dependency of any kind, they describe a single event and they share core arguments (Aikhenvald, 2018; Bisang, 1995; Crowley, 2002; Durie, 1988; Haspelmath, 2016). In this way Matukar represents a typical Oceanic language, as SVCs are widespread across Oceanic languages (Crowley, 1987; Early, 1993; Hamel, 1993; Næss, 2011; Sperlich, 1993; Schokkin, 2013 *inter alia*).

Matukar SVCs are obligatorily inflected for subject and both constituent verbs can carry object suffixes (Barth & Anderson, 2015). Since person marking can intervene between the two verb roots, Matukar SVCs are non-contiguous (Aikhenvald, 2018). Only the final verb in the series carries TAM information and as such all verbs share the same aspect and mood values, as well as polarity. Both transitive and intransitive verbs are free to enter Matukar SVCs. Matukar has core layer serial verb constructions synchronically (cf. Crowley, 2002; Foley & Olson, 1985). Each verb forms its own nucleus at the core layer of the clause and so both can be indexed for subject and object.[3] Serialization at the core layer is very common across Oceanic languages (Crowley, 1987; Hamel, 1993; Ross, 2013; Sperlich, 1993 *inter alia*).

Semantically, Matukar verb serialization represents simultaneous or closely overlapping parts of the same macro event. The most frequent type of asymmetric (cf. Aikhenvald, 2018) serialization in Matukar is directional, where the directional V2 in an SVC serves to specify the direction of the lexical V1 (Barth & Anderson, 2015). Directional verbs are the most common semantic class to enter into SVCs cross-linguistically (Aikhenvald, 2018).

Oceanic SVCs can be classified as either geographic or deictic (Ross, 2004a). Geographic directional SVCs have a geocentric directional morpheme as V2. In Matukar, they specify the direction upwards (*das, mais, sa*), downwards (*pid, si*), inland (*sa*) or seawards (*si*). In (8) *mais* expresses the direction of *en* as upwards with the meaning 'living up' and in (9) *pid* specifies the direction of *la* as downwards resulting in the meaning 'go down'. Note that *en* is a stative non-motion verb, so *mais* is specifying where *en* is happening, rather the direction in which it happens.

(8) *…ai luwadi-te* ***diy-en*** ***di-mais-agokai.***
tree middle-LOC **3PL.S-live 3PL.S-ascend-I.R.IPFV.HAB**
'They live up in the middle of the tree.'
Berry Kuyau – PNGpics12-jb_bk: 1941887 ms–1945446 ms

3. This is in contrast to nuclear SVCs where both verbs together form the nucleus, so no material can intervene between them and the series is only marked once for subject and object.

(9) *…Ilonlo* ***ngam-la*** ***ngam-pid-e…***
inside **1PL..EXCL.S-go** **1PL..EXCL.S-descend-D.SEQ**
'…we went down inside…'
Berry Kuyau – Berry_Kuyau_Funeral_Custom: 107154 ms–109940 ms

Vertical and horizontal geocentric directional verbs can co-occur in SVCs to create a tri-verb serial verb construction. These specify direction along two axes. In (10), *sa* describes inland horizontal motion and *pid* specifies downward vertical motion. The semantics of Matukar SVCs are always semantically transparent from the component verbs and this extends to three verb SVCs.

(10) ***Nga-ngale*** ***nga-sa*** ***nga-pid-ope…***
1SG.S-take **1SG.S-go.inland** **1SG.S-descend-D.WHEN.COND**
'When I take it down home…'
Margaret Lem Kaluk – Life_Story_of_Margaret_Lem_Kaluk: 170573 ms–175506 ms

Deictic directional SVCs characterize the action of V1 in terms of the deictic centre. The second verb is one of the deictic morphemes and they specify whether the direction of the first verb is towards (*so, palum, mul*) or away from the deictic centre (*a, la*). Deictic directional SVCs can similarly occur with transitive or intransitive first verbs. In (11), an intransitive V1 is followed by an andative V2 giving the meaning 'paddle away'.

(11) ***…nga-feni*** ***ng-a-ba*** *wasing maya-lo…*
1SG.S-paddle **1SG.S-go-I.IRR.DESID** fish far-in
'…I paddle away to the far away fish…'
Bruce Kainor Kaluk – Bruce_Kainor_Kaluk_Life_Story: 414576 ms–421092 ms

Similarly, a deictic V2 can specify the direction of a geographic V1. In (12), the deictic verb *a* specifies the direction of the geographic verb *das* as away from the deictic centre.

(12) *…main abeng dadubman-te* ***di-das*** ***diy-a-we…***
PROX place cold-LOC **3PL.S-ascend** **3PL.S-go-D.SEQ**
'…they went all the way up to the cold place…'
Veronica Kubod – DGB1-PNGpics09-gk_vk: 800333 ms–805357 ms

Verbs are commonly marked by the zero-morpheme prefix *ø-* in the 2nd and 3rd person singular. When a directional V2 in a Type 1 SVC is marked by the zero morpheme *ø-* there is difficulty in distinguishing between a directional suffix and a Type 1 SVC, as in (13).

(13) ***…ngale palum-ap…***
get **come-D.IRR**
'…he will bring it…'
Rosa Kibis Dikoi – 50pics07-rkd_csw: 788111 ms–792810 ms

Object marking and certain phonotactic constraints can help to disambiguate these cases. Object marking typically appears after directionals and in these cases, the construction can be classified as a directional suffix. If the object suffix appears before the directional component, the construction is classified as an SVC. Certain consonant sequences are not permitted. If the addition of a directional suffix to a V1 would result in one of these prohibited sequences, an epenthetic *a* is inserted between the V1 and the directional suffix, as in (14) where with the andative directional-*la* becomes *-ala*. In these cases, an audible *a* would mean classifying the construction as a directional suffix and no audible *a* would mean classifying the construction as an SVC.

(14) *…ngau* ***ng-ep-ala*** *i* *mado-nggo mon-e.*
1SG.S **1SG.S-look-AND** 3SG.S sit-I.R.IPFV DIST-DEFID
'I looked away and saw her sitting there.'
John Bogg – PNGpics12-jb_bk: 1141578 ms–1143600 ms

Where there are no morphophonological changes for clues, there is ambiguity. For example, where V1 ends with a vowel and the directional or V2 begins with a consonant.

SVC2 constructions

Clause chaining in Matukar is a productive means of linking together simultaneous, sequential or overlapping events contained in separate clauses into one sentence (Barth & Anderson, 2015). Clause chains are a typically Papuan feature (Foley, 2010; Ross, 2008; Pawley & Hammarström, 2017) and so the presence and prevalence of these structures in Matukar is likely a result of contact with neighbouring Papuan languages (Barth & Anderson, 2015; Barth & Ross, in press; Ross, 2008). Clause chains are defined by having one independent verb and a number of non-finite verbs with some morphological marker of dependency (Foley, 2010; Pawley & Hammarström, 2017; Longacre, 2007). In Matukar, the verb in the final clause has independent morphology which specifies the aspect of the verb and the mood of the entire chain. Clause chaining is thus a rich source for grammaticalization. A similar process has taken place, resulting in the reanalysis of a clause chain construction that contains a lexical V1 and a directional verb as V2 (see Barth & Anderson, 2015; Barth & Ross, in press). These constructions are the Type 2 directional SVCs.

A construction containing a V1 with surface clause chain morphology and a directional V2 has been reanalysed as having the same functional semantics as a directional SVC (Barth & Anderson, 2015; Barth & Ross, in press). The directional verb in this construction specifies the direction of first verb, just as it does in SVC1s. The first verb in this construction is marked as dependent with either the *-e~-nge~-we* suffixes for realis or the *-p* suffix for irrealis (Barth & Ross, Forthcoming). No other type of medial marking can occur on the V1 in Type 2 SVCs. Both verbs have identical subject marking. The second verb specifies the mood and aspect of the SVC, just as with the SVC1s. Example (15) shows the verb *ngale* with dependent marking, followed by *a* specifying andative direction. In (16), the verb *ro* has medial marking and is followed by *palum* characterizing the direction as towards the deictic centre.

(15) *Matan di-nage-ndop* ***di-ngale-p*** ***diy-a-p*** *ha-di*
money 3PL.S-put-D.IRR **3PL.S-take-D.IRR** **3PL.S-go-D.IRR** POSS-3PL
aba-te-ge…
house-LOC-FOC
'They will put the money and take it away to their house…'
Wendy Pulu – 50pics06-wp_rw: 1802541 ms–1804103 ms

(16) *…mam hona* ***di-ro-ye*** ***di-palum-e…***
plane on **3PL.S-fly-D.SEQ** **3PL.S-go-I.R.PFV**
'…they flew away on the plane…'
Rebecca Willie – 50pics06-wp_rw: 2118526 ms–2120485 ms

An alternative analysis might be that these constructions are a special type of clause chain that describe a single event (Barth & Ross, in press). However, since they align semantically and functionally with Matukar SVCs and are distinguished from clause chains in their monoclausality, the analysis of serialization is preferred. Barth and Anderson (2015, p. 221) use the circumclitic *duba…mig* 'quickly' to evidence monoclausality. This circumfix encloses single clauses only and can be seen in (17). Further evidence to substantiate the serialization analysis comes from negation. Example (18) shows that the negative prefix *ti-* takes scope over the whole SVC. While Matukar word order is somewhat flexible, the direct object of a verb will always be in the same clause as the transitive verb. Example (19) shows the object *anan* appearing after *palume*, an intransitive verb, so this must be monoclausal to be grammatical.

(17) *Duba nub* ***a-ngale-p*** ***a-so=mig!***
HAST water **2PL.S-get-I.IRR.IMP** **2PL.S-come-HAST**
'Hurry up and bring the water!'
Kadagoi Rawad Forepiso – DGB1-2011_005-waiwaik_3_making_drink_video_narration: 73000 ms-76000 ms

(18) *Man hona* ***ti-ro-ap*** ***palum-ap.***
Bird with **NEG-fly-D.IRR come-D.IRR**
'You will not fly here on a plane.'
Rebecca Willie – 50pics06-wp_rw: 2118526 ms–2120485 ms

(19) ***Ngale-nge*** ***palum-e*** *anan tai*
take-D.SEQ come-I.R.PFV yam DUB
'She took them a yam.'
Monika Malik Gim – mjk04-mmg_1: 54200 ms–56505 ms

Reanalysis of grammatical categories often occurs when there are multiple possible analyses or there is ambiguity (Traugott & Trousdale, 2010). The third Matukar construction has been reanalysed as an SVC as a result of such ambiguity. Grammaticalization is a process that occurs on a continuum from lexical to grammatical forms to more grammatical forms (Heine & Kuteva, 2002) and is generally considered to be a gradual process of which synchronic gradience is a symptom (Traugott & Trousdale, 2010). The variation in the Matukar Panau directional system constitutes a gradient system and is an indication that there are ongoing grammatical changes taking place.

The development of SVCs over time can be partially distinguished based on their symmetry (Aikhenvald, 2018). Symmetrical core SVCs are compositional and thus are unlikely to grammaticalize, tending instead to lexicalize. Asymmetrical SVCs behave differently and tend towards grammaticalization like Matukar directional SVCs. The relative instability of SVCs features prominently in the Oceanic literature (Barth & Anderson, 2015; Hamel, 1993; Næss, 2011; Ross, 2004a, 2004b; Schokkin, 2013). A number of Oceanic languages show synchronic evidence that their directional SVCs are undergoing grammaticalization to prepositions, prefixes or suffixes (directionals) (Hamel, 1993; Næss, 2011; Schokkin, 2013). In contrast, Matukar directional core (Type 1) SVCs do not show evidence of grammaticalization and are used frequently and productively, like the directional suffixes. Similarly, the third Matukar directional construction shows no synchronic evidence of instability.

2. Methodology

2.1 Corpus

The data used in this study comes from a corpus of 86 spoken texts from Matukar Panau (Barth, 2010) collected between 2010 and 2020. The data is transcribed in

ELAN and translated into Tok Pisin and English. There are 48 speakers (female $n=27$, male $n=21$) across the texts, born between approximately 1930 and 1993.

Roughly 25% of the corpus consists of descriptive picture tasks, where the speakers describe a series of pictures. One task involved describing pictures of people engaging in everyday tasks and activities in Germany (Terrell, Tschirner, & Nikolai, 2005). The other task involved pictures of people and places from the area surrounding the villages of Matukar and Surumurang in Papua New Guinea. Another 10% of the corpus texts are social cognition tasks, where the speakers engaged in narrative problem-solving tasks and descriptive picture tasks (Barth & Evans, 2017; San Roque et al., 2012). Traditional songs make up approximately 5% of the corpus. Procedural texts where the speaker is describing a custom or the correct way to conduct a daily task comprise another 5% of the corpus. The rest of the corpus (55%) consists of descriptive life stories, descriptions of daily life and interviews where the speaker describes themselves, their family and their life, as well as traditional stories and historical stories.

The data was annotated manually in ELAN, adding tiers to existing corpus files. Annotation noted the construction type (directional suffix vs. Type 1 SVC vs Type 2 SVC), subject person of the verbs/SVCs, the presence of an object noun phrase, the verb(s) involved, and TAM marking. A total of 788 tokens were tagged. Several types of tokens were excluded: cases where the construction type was ambiguous due to ∅- 2/3 person subject marking ($n=147$), SVCs where V1 in the SVC occurred with a directional suffix, making the token an instance of more than one construction type ($n=9$), and cases with the directional morpheme *mais*, as it rarely appeared in the corpus ($n=3$). This left 629 tokens for analysis for quantitative analysis.

2.2 Factors influencing directional construction choice

Statistical analysis allows for determining whether variation is random or not. By using quantitative methods linguists can model variation, revealing significant factors associated with people's use of one variant over others. The use of quantitative methodologies in modelling linguistic variation necessitates accepting the inherent variability of language (Weinreich, Labov & Herzog, 1968) and understanding that variation is inherently probabilistic (Guy, 1988; Wolk, Bresnan, Rosenbach, & Szmrecsanyi, 2013). This study uses Bayesian regression modelling to determine the influence of independent variables on the 3-way dependent variable of directional construction type, but see Davey (2021) for other quantitative approaches to Matukar directional variation.

The dependent variable for this study is the directional construction type: directional suffix, SVC type 1, or SVC type 2. Table 2 summarises these construc-

tions for reference. Binary logistic regression models are quite common in quantitative variationist linguistics (Coupé, 2018; Speelman, 2014; Speelman, Heylen, & Geeraerts, 2018 *inter alia*), but having a three-way categorical DV requires a more complicated modelling type: multinomial regression. In multinomial logistic regression, we can assess the probability of more than two outcomes (cf. Section 2.3).

Table 2. Summary of the Matukar Panau directional constructions

	Directional suffix	**Type 1 SVC**	**Type 2 SVC**
Morphosyntax	S-VERB-DIR-TAM	S-VERB S-DIR.VERB-TAM	S-VERB-D S-DIR.VERB-TAM

Ten independent variables are included in the statistical analysis, six internal (linguistic) variables and four external (sociolinguistic) ones.

I. *Internal variables*

1. *Semantic type of directional morpheme*

There are no occurrences of geocentric directional suffixes on deictic verbs in the corpus, suggesting that this combination is not possible. There then seems to be conditioning based on the semantic type of the directional morpheme, based on how they combine with verbs.

Tokens were therefore coded for the type of directional in the construction. That is, whether the directional is one of the deictic verbs (*a/la/mul/palum/so*) ($n = 336$), one of the geocentric verbs that operates on a vertical axis (*das/mais/pid*) ($n = 239$) or one of the geocentric verbs that operates on both a vertical and horizontal axis (*sa/si*) ($n = 52$). The token counts for each directional construction by the type of directional can be seen in Table 3.

Table 3. Semantic type of directional morpheme by directional construction

Semantic type of directional morpheme	**Directional suffixes**	**Type 1 SVCs**	**Type 2 SVCs**
Deictic	73	154	109
Geocentric (vertical)	25	13	15
Geocentric (horizontal and vertical)	41	191	8

2. *Frequency of the directional morpheme as a main verb in the corpus*

Some Oceanic languages contain fully grammaticalized directional morphemes that are no longer members of the source category from which they grammaticalized (Pawley, 2003). It may be the case that some of the Matukar directionals are

in the process of losing their status as main verbs, particularly since the system contains multiple synonyms. We would then expect these morphemes to appear less frequently as main verbs than as directionals.

A corpus search for each of the directional verbs showed the frequency of their occurrence. These figures do not include cases where the verb is either the first or second verb in a SVC. The frequencies are as follows: *a*[4] is the most frequently occurring verb ($n = \sim 1206$), *palum* comes next ($n = 333$), *so* is the next most frequent ($n = 200$), *pid* follows this ($n = 122$), *si* is the next most frequent ($n = 83$), *mul* follows this ($n = 80$), *das* comes next ($n = 79$), following this is *sa* ($n = 68$) and lastly the least frequently used verb is *la* ($n = 23$). The variable was log transformed, a common technique in corpus linguistics when dealing with word frequencies that often have a highly skewed distribution. Linearized corpus frequencies can be more easily compared with other data (Gries, 2010, p. 272).

3. *Lexical semantics of the main verb*

It is well attested that the lexical semantics of a verb constrains the syntactic constituents of a phrase. For example, one constraint for the dative alternation in English is that verbs that license spatial goals do not alternate (Bresnan & Nikitina, 2010). Thus, verbs with differing lexical semantics entail different clausal constituents, which might affect construction choice. We might also expect the semantics of a verb to affect the directional morphemes it can combine with and the directional construction it is most associated with. We classified main verb/V1 semantics based on the clausal components that each semantic category might entail, as well as the type of movement that they involve. Autonomous motion verbs ($n = 391$) (e.g., *ro* 'to fly') are verbs where the subject is the mover. Change of possession verbs ($n = 113$) (e.g., *ngale* 'to get') represent the transfer of an object and mainly consist of verbs that result in meanings of bringing and taking when included in a directional construction. In these cases, the subject and the object are moving and there is no constraint on direction. In caused motion verbs ($n = 81$) (e.g., *bal* 'to throw'), the object of the verb is the mover and the subject is not necessarily moving. Finally, non-motion verbs ($n = 44$) (e.g., *ep* 'to look') have no inherent movement associated with them. We might expect non-motion verbs to behave differently to verbs of motion and we might expect verbs that entail restrictions on direction to pattern differently to verbs where the movement is unconstrained. Table 4 and Figure 2 show the distribution of main verb

4. It is too difficult to search the whole corpus for the verb *a* as this would yield all words with the letter *a*, as 65% of the texts are not parsed and glossed. However, of the 35% that are parsed and glossed there are 422 instances of *a* as a main verb, which is higher than the frequency recorded for any other directional verb in the whole corpus. If we assume that this proportion is representative of the corpus, there would be roughly 1,206 instances of *a* in the entire corpus.

(and V1) semantics and directional construction type, where it does seem that there is some semantic influence on construction choice.

Table 4. Lexical semantics of the main verb by directional construction

Lexical semantic category of V1	Directional suffixes	Type 1 SVCs	Type 2 SVCs
Autonomous motion	33	290	68
Caused motion	58	9	14
Change-of-possession	12	56	45
Non-motion	36	3	5

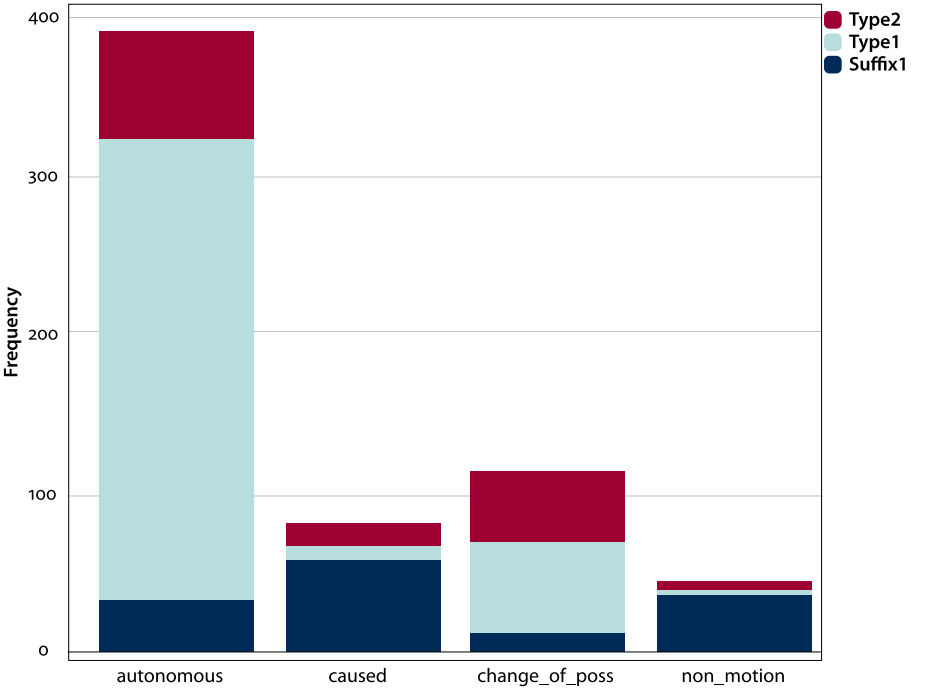


Figure 2. Bar plot of directional construction frequency by the lexical semantics of the main verb

4. *Transitivity of the main verb*

Verb transitivity has also been shown to affect morphosyntactic variation (Lazard, 2002). For example, Spanish exhibits morphosyntactic variation in its system of case clitics and this is influenced in part by the transitivity of the verb (Guajardo, 2021). Therefore tokens are coded for whether the main verb is intransitive

($n=371$) or transitive ($n=257$). The token counts for each directional construction by the transitivity of the main verb can be seen in Table 5.

Table 5. Transitivity of the main verb by directional construction

Transitivity of the main verb	Directional suffixes	Type 1 SVCs	Type 2 SVCs
Intransitive	41	289	41
Transitive	98	68	91

5. *Subject as speech act participant (SAP)*

Whether or not the subject of a construction is a SAP may interact with the type of directional morphemes used. As deixis is linked to the speech act context (Levinson, 2003), the choice of directional might differ depending on whether the subject is involved in the speech act. Tokens were coded for whether the subject of the directional construction is a SAP (1st or 2nd person) ($n=174$) versus 3rd person ($n=455$), token counts are in Table 6.

Table 6. Subject as a SAP by directional construction

Speech act participant	Directional suffixes	Type 1 SVCs	Type 2 SVCs
Yes	58	89	27
No	81	269	105

II. *External variables*

6. *Text genre*

Text context and speaker motivation have been shown to be important in influencing grammar (Fleischman & Waugh, 2016). Different text genres naturally affect lexical and grammatical choices as the context necessitates. For example, in Matukar Panau, Barth (2019) found that the choice of kinship possession term was influenced by text type, with less formal situations associated with less direct possession patterns. We might expect speaker motivation to affect directional choice and the discourse context to interact with the types of verbs that are used.

Tokens were therefore coded for the genre of the text. There are four levels to this variable, consisting of descriptive texts ($n=484$), narratives ($n=92$), procedurals ($n=31$) and songs ($n=22$), token counts in Table 7.

Table 7. Genre of text frequency by directional construction

Genre of text	Directional suffixes	Type 1 SVCs	Type 2 SVCs
Descriptive	99	292	93
Narrative	15	48	29
Procedural	17	6	8
Song	8	12	2

7. *Gender of the speaker*

In sociolinguistic studies gender is often found to condition variation, though this seems to be less significant in the case of morphosyntactic variation (Meyerhoff, 2013). However, in Barth's (2019) study of the variation in the Matukar Panau kinship possession system, gender was found to be significant in driving the changes to the system. It is possible that the choice of a certain construction signals some social meaning. As directional constructions are completely unstudied in sociolinguistics, it is worthwhile to investigate whether gender may play a role. All of the speakers in the corpus were either female ($n = 461$) or male ($n = 168$), token counts by directional construction are in Table 8.

Table 8. Gender of the speaker by directional construction frequency

Gender of speaker	Directional suffixes	Type 1 SVCs	Type 2 SVCs
Female	108	258	95
Male	31	100	37

8. *Speaker age*

Speaker age was also found to be significant in Barth's (2019) study of the variation in kinship possession strategies, with younger women and older men found to be driving the shift to indirect possession strategies. We therefore investigate whether the same pattern holds for the variation in directional constructions. Further, younger speakers are more likely to be Tok Pisin dominant, which may result in them using SVC strategies (found in Tok Pisin) in Matukar Panau, more than directional suffixes, which do not have a Tok Pisin equivalent.

Tokens were coded for the age of the speaker based on whether they were older than 60 ($n = 196$) or younger than 60 ($n = 433$). Token counts by directional construction are included in Table 9.

Table 9. Speaker age by directional construction frequency

Speaker age	Directional suffixes	Type 1 SVCs	Type 2 SVCs
Older	46	101	49
Younger	93	257	83

9. *Speaker*

We might expect individual speaker behaviour to affect the choice of directional construction (cf. Tagliamonte & Baayen, 2012 *inter alia*). The tokens were coded for the speaker, with a total of 48 speakers contributing tokens to the corpus.

2.3 Bayesian multinomial modelling

Linguistic variationist studies typically employ frequentist methods that test the likelihood of the data based on the null hypothesis (Vasishth & Nicenboim, 2016). In the context of this paper, the potential null hypothesis could be thought of as the case where the predictor has no effect on the variation in the directional constructions. Predictors would be shown to be significant when the likelihood of the data under the null hypothesis is shown to be significantly small enough. This is usually determined to be when the significance value p is less than 0.05 (Vasishth & Nicenboim, 2016). One criticism of frequentist methods is the use of p-values as a measure of significance, as they do not provide any indication of the strength of the hypothesis that is being tested and are easily misinterpreted (Halsey, 2019).

More recently Bayesian approaches that can directly test the hypothesis (Nicenboim & Vasishth, 2016) have been adopted into linguistic studies (Flego & Forrest, 2021; Levshina, 2016; MacKenzie, 2020; Nalborczyk et al., 2019; Nicenboim & Vasishth, 2016; Vasishth, Chen, Li, & Guo, 2013 *inter alia*). One major advantage of taking a Bayesian approach is the ability to incorporate a researcher's prior knowledge into the model through the setting of priors. Priors are probability distributions that are based on expert knowledge or the results of previous studies before examining the data (Bürkner, 2017a). The size of the dataset generally determines the effect of priors on the model, with larger datasets less affected by priors (Levshina, 2016, 2019; Nicenboim & Vasishth, 2016; Vasishth et al., 2013).

Bayesian approaches provide a credible interval (CI), which shows the "range over which we can be reasonably sure that the true parameter value lies" (Nicenboim & Vasishth, 2016, p. 593). This way of quantifying the uncertainty of the parameter values based on the posterior distributions is another important way that a Bayesian approach differs from a frequentist one. Bayesian models use

Markov chain Monte-Carlo (MCMC) sampling to "approximate aspects of posterior distributions that cannot be directly calculated…[by] drawing a sequence of samples from the posterior, and examining their mean, range, and so on" (van Ravenzwaaij, Cassey, & Brown, 2018, p. 144). In other words, MCMCs are used to estimate the posterior distribution, which cannot usually be "derived analytically" (Nicenboim & Vasishth, 2016, p. 594).

In this study, the multinomial mixed-effects Bayesian regression model was implemented through the brm function in the {brms} package (Bürkner, 2017a, b) in R (R Core Team, 2021). The brms package provides an interface that allows the use of MCMC samplers to approximate the posterior distribution through the programming language Stan (Carpenter et al., 2017). The Monte-Carlo part of the name refers to the process of estimating properties of the posterior distribution from many random samples. The Markov Chain part of the name refers to the practice of generating random samples from prior random samples, forming a sequential chain (van Ravenzwaaij et al., 2018, p. 143). A sufficient number of iterations is required to generate enough random samples. Since the first sample in the chain is not based on a prior sample, MCMCs typically have a burn-in or warm-up period to control for any random samples that are "unlikely to come from the target distribution" (van Ravenzwaaij et al., 2018, p. 146). That is, a number of early iterations are omitted.[5]

Bayesian modelling requires the setting of priors based on how the variable is expected to behave. These expectations are formed based on prior knowledge or previous studies. By default, the model sets flat priors that identify all variables as equally likely to be significant. This results in similar results to a frequentist approach (Lemoine, 2019). In this study, weak priors were chosen, as the hypotheses outlined in the variables Section (2.2) are based on other studies of morphosyntactic variation. However, since this study is the first of its kind, there is not enough information available to provide the model with fully informative priors. Weak as opposed to non-informative priors are preferred as the dataset is relatively small and non-informative priors can result in high Type 1 and Type M errors (Lemoine, 2019). Furthermore, weak priors can constrain the coefficient values to avoid unrealistic values, which is particularly beneficial in the case of small and sparse datasets (Levshina, 2019). Weakly informative Cauchy priors were specified for the fixed effects and the default Student's *t* priors were used for the intercept, standard deviation and random effects. The model was tested for sensitivity to different weak prior distributions (as recommended in Nicenboim &

5. The MCMC sampler used in the model in this article is the no-U turn sampler (NUTS), which is an extension of the Hamiltonian Monte Carlo (HMC) sampler. Bürkner (2017) notes that these samplers improve efficiency, as well as the quality of the random samples.

Vasishth, 2016), with similar results for four commonly used weakly-informative prior distributions showing that the model is relatively robust to different weak prior distributions (see Appendix 2 for brief discussion). Variables were tested *a priori* for collinearity. The lexical semantics of the main verb was found to be collinear with transitivity ($\chi^2 = 169.27$). Including the frequency of the directional as a main verb caused convergence issues and so was omitted from the model. The leave-one-out (LOO) method was used to find the best model, wherein models are compared based on the parameters included (following Vehtari, Gelman, & Gabry, 2017). The LOOIC (leave-one-out information criterion) for the final model is 859.5. The model then includes three internal variables: the semantic type of the directional, lexical semantics of the main verb and the subject as SAP; and two external variables: the gender and age of the speaker as an interaction term. All other variables were tested but not included in the final model.

To calculate the predictive power of the model, the predicted probability of directional suffix, Type 1 SVCs and Type 2 SVCs for each datapoint was compared to the actual construction that was used (following Levshina, 2016). The directional construction with the highest probability was selected for comparison. The accuracy of the model was then calculated as the proportion of correct predictions, which yielded a result of 73%. The probability of selecting the correct directional construction by chance is one in three or 33%, so the model makes a clear improvement compared to the baseline.

The 95% CI is provided for each coefficient, showing the range over which the middle 95% of the posterior values lie and thus how confident we can be in the direction of the effect.

3. Results

The model runs four chains, each with 2000 iterations with a warmup period of 1000 iterations, so the total number of post-warmup draws is 4000.[6] An inspection of the trace plots shows that they each resemble a "fat, hairy caterpillar", proof that the Markov chains converged (see the Appendix for brief discussion) (Lunn et al., 2013 in Levshina, 2016, p. 253). The Rhat statistic is the potential scale reduction factor and is another way to check that the chains have converged (cf. Gelman & Rubin, 1992). The within-chain and between-chain estimates are compared for each coefficient and if the chains have converged the Rhat statistic should be 1. The output of the brm function implemented here shows the Rhat statistics with

6. That is, after the warmup period there are 4000 posterior samples, 1000 samples per chain ((2000–1000)*4 = 4000).

all values equal to 1, showing further proof of convergence. The model was also tested for multicollinearity and predictive power.

The results of the random effects can be seen in Table 10. There is very little inter-speaker variance, with slightly increased variability for Type 2 SVCs compared to Type 1 SVCs.

Table 10. Random effects results from multinomial Bayesian regression with 3-way dependent variable

Random effects	Estimate	Estimated error	Lower boundary of 95% CI	Upper boundary of 95% CI
Intercept: Type 1 SVC	**0.45**	**0.26**	**0.03**	**1.03**
Intercept: Type 2 SVC	**0.78**	**0.32**	**0.15**	**1.44**

The model summary returns 95% credible intervals for each coefficient with parameters deemed to have an effect if the interval does not include zero. The default variables are arbitrary and are by default the first level of each variable alphabetically. Posterior probabilities are also provided, showing the probability that the coefficient is less than or greater than zero providing us with an indication of the likely effect of the variable even when the CI includes zero. The posterior probability is the probability of the event after taking prior information into account and the posterior mean provided is the mean of the posterior distribution. Negative posterior means are associated with higher rates of directional suffix use. When the CI for a parameter does not include zero we can be confident about the direction of the effect and these lines are bolded in the results tables below.

The intercepts represent the initial probabilities of the Type 1 and Type 2 SVCs compared to the default, the directional suffixes, before considering the effects of the predictor variables and the results are shown in Table 11. The results show that the difference between the directional suffixes and Type 1 SVCs is greater than the difference between the directional suffixes and Type 2 SVCs (2.08 vs 1.57), perhaps indicating longer term separation and diversification of the older constructions.

Table 11. Results of intercepts from multinomial Bayesian regression with 3-way dependent variable

	Posterior mean	Lower boundary of 95% CI	Upper boundary of 95% CI	P($\beta < 0$)	P($\beta > 0$)
Intercept: Type 1 SVC	**2.08**	**1.28**	**2.87**	**0%**	**100%**
Intercept: Type 2 SVC	1.57	0.56	2.50	0.22%	99.78%

The results in Table 12 show the parameters relating to the lexical verb. The posterior means for the caused motion parameters are negative, showing that these verbs are less likely to appear in either type of SVC compared to directional suffixes. The effect is stronger for Type 1 SVCs. Non-motion verbs are less likely to be observed in Type 1 and Type 2 SVCs than with directional suffixes and similarly the direction of the effect is larger for Type 1 SVCs. We can also conclude that autonomous motion and change of possession verbs are more likely to be used with either type of SVC than caused motion and non-motion verbs, since the posterior means are greater. The CI for change of possession verbs includes zero for both types of SVCs, meaning we cannot be certain of the effect that this has on the variation, but it might also suggest that they exhibit the greatest variation. So, we see Type 1 and Type 2 SVCs are similar in the direction of the effect for each parameter relating to the lexical semantics of the verb in comparison to the directional suffixes. However, the size of the effect varies. Type 2 SVCs are more associated with caused motion verbs, change of possession verbs and non-motion verbs than Type 1 SVCs, since all of the posterior means are comparatively higher. Directional suffixes are more strongly associated with caused motion and non-motion verbs than either type of SVCs, since the posterior means are negative.

Table 13 shows the results of the parameters pertaining to the directional morphemes. Geocentric directional morphemes that operate over a horizontal and vertical axis are more likely to appear as directional suffixes than in Type 1 and Type 2 SVCs, though the CI contains zero so we cannot be certain of the direction of this effect. The geocentric morphemes that operate on a vertical axis are more likely to appear in Type 1 SVCs than as directional suffixes or in Type 2 SVCs. They are less likely to appear in Type 2 SVCs than as directional suffixes.

Taken together, we see that Type 1 and Type 2 SVCs pattern more similarly when it comes to the lexical semantics of the main verb but diverge when it comes to the semantics of the directional morphemes.

Table 14 shows the results for whether the subject is a speech act participant. Both Type 1 and Type 2 SVCs are less likely than directional suffixes when the

Table 12. Results of lexical semantics parameter from multinomial Bayesian regression with 3-way dependent variable

Parameter	Posterior mean	Lower boundary of 95% CI	Upper boundary of 95% CI	P($\beta < 0$)	P($\beta > 0$)
Lexical semantics (default – autonomous motion)					
Caused motion: Type 1 SVC	**−4.42**	**−5.36**	**−3.57**	**100%**	**0%**
Caused motion: Type 2 SVC	**−2.15**	**−2.98**	**−1.32**	**100%**	**0%**
Change-of-possession: Type 1 SVC	−0.06	−0.74	0.62	57.12%	42.88%
Change-of-possession: Type 2 SVC	0.35	−0.40	1.08	17.33%	82.67%
Non-motion: Type 1 SVC	**−4.60**	**−5.92**	**−3.17**	**100%**	**0%**
Non-motion: Type 2 SVC	**−3.11**	**−4.33**	**−1.93**	**100%**	**0%**

Table 13. Results of directional semantics parameter from multinomial Bayesian regression with 3-way dependent variable

Parameter	Posterior mean	Lower boundary of 95% CI	Upper boundary of 95% CI	P($\beta < 0$)	P($\beta > 0$)
Semantics of directional morpheme (default – deictic)					
Geocentric (horizontal/vertical): Type 1 SVC	−0.29	−1.15	0.57	75.3%	24.7%
Geocentric (horizontal/vertical): Type 2 SVC	−0.26	−0.98	0.55	73.55%	26.45%
Geocentric (vertical): Type 1 SVC	**1.23**	**0.51**	**1.87**	**0%**	**100%**
Geocentric (vertical): Type 2 SVC	**−1.73**	**−2.62**	**−0.76**	**100%**	**0%**

subject is a SAP, with the strongest negative effect for Type 2 SVCs. This could be due to the proximity of the participants to the event. That is, when participants are more involved in the directed event, directional suffixes are used and when they are less directly involved SVCs are preferred. As already mentioned, there seems to be an association between SAP subjects and the morpheme *la*, which is mostly used as a directional suffix. It might be the case that *la* is the preferred deictic morpheme when the event is proximal to the SAPs and *a* is preferred when the subject is further removed from the speech act.

Table 14. Results of SAP parameter from multinomial Bayesian regression with 3-way dependent variable

Parameter	Posterior mean	Lower boundary of 95% CI	Upper boundary of 95% CI	P($\beta<0$)	P($\beta>0$)
Subject as speech act participant (default – not a SAP)					
SAP: Type 1 SVC	**−1.09**	**−1.69**	**−0.50**	**100%**	**0%**
SAP: Type 2 SVC	**−1.42**	**−2.11**	**−0.72**	**100%**	**0%**

Table 15 shows the results for the external variables. Speaker gender and age are not valuable predictors as individual parameters, but as an interaction term we see that younger males are associated with higher rates of SVC usage. Although the CI contains zero, the results for the individual gender and age parameters show the inverse of this, implying that older women are also associated with higher rates of SVC usage. Interestingly, this mirrors the results seen in Barth's (2019) study of the variation in the Matukar kinship system where younger males and older women were found to be the demographics driving the shift to indirect possession strategies.

In sum, the language internal factors are largely responsible for the strongest predictors for the choice of Matukar Panau directional constructions, as expected based on the morphosyntactic variationist literature (Meyerhoff, 2013). Type 1 and Type 2 SVCs are generally associated with similar parameters relating to the lexical semantics of the verb when compared to the directional suffixes. However, the size of the effect differs across these parameters.

Table 15. Results of external parameters from multinomial Bayesian regression with 3-way dependent variable

Parameter	Posterior mean	Lower boundary of 95% CI	Upper boundary of 95% CI	P($\beta<0$)	P($\beta>0$)
Speaker gender (default – female)					
Male: Type 1 SVC	−0.08	−1.05	0.94	56.38%	43.62%
Male: Type 2 SVC	−0.40	−1.65	0.85	73.62%	26.38%
Speaker age (default – older)					
Younger: Type 1 SVC	−0.34	−1.06	0.37	82.80%	17.20%
Younger: Type 2 SVC	−0.41	−1.39	0.55	80.17%	19.83%
Speaker age and gender					
Younger male: Type 1 SVC	**1.37**	**0.02**	**2.70**	**1.60%**	**98.40%**
Younger male: Type 2 SVC	1.67	−0.08	3.36	2.65%	97.45%

4. Discussion and conclusion

The choice of directional construction is largely influenced by internal variables related to the directional morpheme, the lexical semantics of the verb, whether the subject is a SAP and sociolinguistic variables, though to a lesser extent. The fact that internal variables relating to the linguistic system were shown to be more significant than external variables in conditioning the variation is typical of grammatical variation, where there are less likely to be significant social variables than significant linguistic variables influencing variation (Meyerhoff, 2013).

One of the most stark differences between the three construction types is the propensity for particular directional expressions to be associated with either the directional suffix construction or the SVCs. Notably, the most frequently attested main directional verbs in simple clauses are the least likely to appear as directional suffixes and, correspondingly, the least frequently attested main verbs are strongly associated with directional suffix use. This implies that some directional morphemes are losing their status as main verbs. The directional morpheme *la* is used more frequently as a morpheme in the three directional constructions ($n=80$) than as a main verb in a simple clause ($n=23$) in our corpus. Similarly, *sa* is used more frequently as a directional element ($n=71$) than as a main verb

($n=68$). In many Oceanic languages this process is already complete. For example, Woleai has three postverbal particles *tage, tiwe/tiw* and *waiu/weiu* that do not occur synchronically as main verbs but are reflexes of Proto-Oceanic directional verbs **sake*, **sipo* and **[w]atu* (Pawley, 2003, p. 164). In Vaeakau-Taumako, the directional morpheme *ange* 'go out, go along' is rarely used as a main verb and is very close in meaning to the frequently used verb *hano* 'to go' (Næss, 2011). Næss (2011) suggests that the morphemes may be nearing complementary distribution. As *la* and the frequently used verb *a* are synonymous, with both representing direction away for the deictic centre, we might expect a similar case of complementary distribution, with *la* a fully grammaticalized directional suffix and *a* a main verb that is also used in SVCs. If we also consider the results of the SAP variable, there may be a subtle semantic distinction between *la* and *a* as well, with *la* used for events proximal to the speech act and *a* used for events distant from the speech act. As part of their grammaticalization, these morphemes may be narrowing in meaning and what constructions they can occur in. If some directional morphemes are losing verbal status and becoming fully grammaticalized as directional suffixes, we then might not expect the less frequently used directional verbs to appear more frequently as main verbs in the future. Moreover, if there is a constraint in Matukar Panau for SVCs to contain directional verbs that can be used as main verbs in simple clauses, we may also see them lose their SVC usage in the future. This would be an area worthy of further study, ideally involving tasks designed around the proximity of the event to the speech act.

In addition to grammaticalization, we also see clear lexicalization, with certain directional elements becoming strongly associated with particular main verbs and semantic classes of main verbs. This, in turn, influences which verb + directional combinations appear in which construction type. Directional suffix constructions are strongly associated with non-motion verbs and caused motion verbs. Non-motion verbs generally do not entail physical movement of the subject from a location. Caused motion verbs require that the subject causes an object to move in some way, but this distance is limited by the force that a person can exert on an object and generally entails limited distance. In comparison, autonomous motion verbs that are more frequently used in both type of SVCs do entail the physical movement of the subject and the distance that the subject can travel is for the most part unrestricted. Change of possession verbs show the greatest amount of variation and we could posit that they may also show the greatest variation in distance travelled, with the distance dependent on the object or means of travel. From a cognitive semantic perspective, this could be interpreted as iconicity of distance (Croft, 2008) or it may be that an SVC also helps iconically reflect multipart events: initiating motion and then moving along a trajectory. Chelliah and Utt (2017) observe a similar pattern in the Himalayan language Lamkang, which

has a pair of directional prefixes reserved for short range movements and synonymous but formally distinct directional verbs used in SVCs for long range movements.

We also see that directional suffixes are more likely when the subject is a SAP and SVCs more likely when the subject is third person. This could be interpreted as a preference for directed events that are more closely related to the speech act to use directional suffixes and events that are less closely related to the speech act associated with SVCs. The relationship between the distance entailed by the directional construction and the choice of directional construction would seem to warrant further investigation.

Croft (2008) argues that more semantically compatible expressions will be expressed more cohesively. In the Mayan language Mam, England (2011, pp. 170–172) identifies a relationship between the semantics of transitive verbs based on their ability to combine with directionals. These groups of transitive verbs are: general verbs that are unconstrained with respect to direction and can combine with all directionals, verbs that entail some direction so are naturally restricted from combining with certain directional morphemes and verbs that entail no or very little movement and combine with a limited number of directionals. In Matukar Panau, some verbs are only compatible with one or two of the directional morphemes. For example, caused motion verb *tariututi* 'to cut' only appears with the morpheme *pid* with the resultant meaning 'to cut down.' Based on England's (2011) findings, non-motion and caused motion verbs might be less compatible with a large number of morphemes, due to the fact that they do not entail physical movement and the verbs might have more specific semantics. Autonomous motion verbs and change of possession verbs are relatively less constrained, since they entail movement and can for the most part be characterized as moving in any direction. For example, verbs like *ro* 'to fly', *tor* 'to walk' or *sol* 'to run' or change of possession verbs like *ngale* 'to take' that do not entail any inherent directional component but are associated with motion. Caused motion and non-motion verbs that are closely associated with only one or two morphemes might then be more likely to appear with a directional suffix compared to autonomous motion and change of possession verbs that are less strongly associated with a particular morpheme. The results also showed a relationship between the semantic group of the directional and the semantics of the verb adding further support to this proposal. Due to the size of the dataset, *hapax legomena* are an issue. That is, some verbs only appear once or twice, and it is difficult to always discern whether the verb is semantically compatible with other morphemes.

In sum, the directional morphemes seem to be grammaticalizing and lexicalizing at different rates, with the morpheme *la* the most grammaticalized and possibly losing its status as a verb. Frequently used verbs seem to be associated

with SVC usage, possibly to retain verbal features. The relationship between the lexical semantics of the main verb and the choice of directional construction is less clear and warrants further investigation. A well-constructed experiment in the field would be an appropriate next step, aimed at discerning how different verbs and morphemes interact. This would necessitate speaker involvement and should aim to have speakers identify which directional morphemes are compatible with each verb, as well as including tasks designed around proximity to the speech act and the distance represented by the verb's event.

Funding

Our sincere thanks to the funders of the Matukar Panau language documentation project: Living Tongues Institute for Endangered Languages, National Geographic Enduring Voices, the Firebird Foundation for Anthropological Research, the ARC Centre of Excellence for the Dynamics of Language and the Endangered Languages Documentation Program (MDP 0382). Open Access publication of this article was funded through a Transformative Agreement with Australian National University.

Acknowledgements

We would like to acknowledge and sincerely thank community members of Matukar and Surumarang especially Kadagoi Rawad Forepiso and Rudolf Raward. We further thank Stefan Schnell and anonymous reviewers of Davey (2021) and this article.

Gloss abbreviations

AND	andative	IRR	irrealis
ADVS	adversative	LOC	locative
COND	conditional	NEG	negative
D	dependent	P	patient-like argument/object
DEFID	definite	PFV	perfective
DESID	desiderative	PL	plural
DIST	distal	PROX	proximate
EXCL	exclusive	R	realis
FOC	focus	S	subject (i.e., S [single] and A [agent-like] argument reference)
HAB	habitual		
HAST	hasten		
I	independent	SEQ	sequential
IMP	imperative	SG	singular
IPFV	imperfective	TOP	topic
INCL	inclusive	UPW	upwards

References

Aikhenvald, Alexandra Y. (2018). *Serial verbs*. Oxford University Press.

Ameka, Felix K. & Essegbey, James (2013). Serialising languages: Satellite-framed, verb-framed or neither. *Ghana Journal of Linguistics*, *2* (1), 19–38.

Barth, Danielle (collector) (2010). *Matukar Panau Language Documentation*. Collection DGB1 at catalog.paradisec.org.au [Open Access].

Barth, Danielle (2019). Variation in Matukar Panau kinship terminology. *Asia-Pacific Language Variation*, *5* (2), 138–170.

Barth, Danielle & Anderson, Gregory D. S. (2015). Directional constructions in Matukar Panau. *Oceanic Linguistics*, *51* (1), 206–239.

Barth, Danielle & Evans, Nicholas (2017). SCOPIC design and overview. In Danielle Barth & Nicholas Evans (Eds.), *The Social Cognition Parallax Interview Corpus (SCOPIC): A Cross-linguistic Resource* (pp. 1–21). University of Hawai'i Press.

Barth, Danielle & Ross, Malcolm (In Press). Clause chaining in Matukar Panau (Oceanic). In Hannah Sarvasy & Alexandra Aikhenvald (Eds.), *Clause chaining in the languages of the world*. Oxford University Press.

Beavers, John, Levin, Beth & Tham, Shiao Wei (2010). The typology of motion expressions revisited. *Journal of Linguistics*, *46* (2), 331–377.

Bisang, Walter (1995). Verb serialization and converbs – differences and similarities. In Martin Haspelmath & Ekkehard König (Eds.), *Converbs in Cross-Linguistic Perspective* (pp. 137–188). Mouton de Gruyter.

Bresnan, Joan & Nikitina, Tatiana (2010). The gradience of the dative alternation. In Lian-Hee Wee & Linda Ann Uyechi (Eds.), *Reality exploration and discovery: Pattern interaction in language and life* (pp.161–185). CSLI Publications.

Bürkner, Paul-Christian (2017a). brms: An R Package for Bayesian Multilevel Models using Stan. *Journal of Statistical Software*, *80* (1), 1–28.

Bürkner, Paul-Christian (2017b). Advanced Bayesian multilevel modeling with the R package brms. arXiv preprint arXiv:1705.11123.

Carpenter, Bob, Gelman, Andrew, Hoffman, Matthew D., Lee, Daniel, Goodrich, Ben, Betancourt, Michael, Brubaker, Marcus, Guo, Jiqiang, Li, Peter, & Riddell, Allen (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, *76* (1), 1–32.

Chelliah, Shobhana L., & Utt, Tyler P. (2017). The Syntax and Semantics of Spatial Reference in Lamkang Verb. *Himalayan Linguistics*, *16* (1), 28–40.

Coupé, C. (2018). Modeling linguistic variables with regression models: Addressing non-Gaussian distributions, non-independent observations, and non-linear predictors with random effects and generalized additive models for location, scale, and shape. *Frontiers in psychology*, *9*, Article *513*, 1–21.

Croft, William (2008). On iconicity of distance. *Cognitive Linguistics*, *19* (1), 49–57.

Crowley, Terry (1987). Serial verbs in Paamese. *Studies in Language*, *11* (1), 35–84.

Crowley, Terry (2002). *Serial verbs in Oceanic: A descriptive typology*. Oxford University Press.

Davey, Kira (2021). A quantitative study of directional constructions in Matukar Panau [Honours thesis, Australian National University]. Australian National University Open Research. https://openresearch-repository.anu.edu.au/handle/1885/274367

Dilley, Laura, Gamache, Jessica, Wang, Yuanyuan, Houston, Derek M., and Bergeson, Tonya R. (2019). Statistical distributions of consonant variants in infant-directed speech: evidence that /t/ may be exceptional. *Journal of Phonetics*, *75*, 73–87.

Durie, Mark (1988). Verb serialization and "verbal-prepositions" in Oceanic Languages. *Oceanic Linguistics*. https://www.jstor.org/stable/3623147.

Early, Robert (1993). Nuclear layer serialization in Lewo. *Oceanic Linguistics*, *32* (1), 65–93. http://www.jstor.org/stable/3623097.

England, Nora C. (2011). *A grammar of Mam, a Mayan language*. University of Texas Press.

Flego, Stefon & Forrest, Jon (2021). Leveraging the temporal dynamics of anticipatory vowel-to-vowel coarticulation in linguistic prediction: A statistical modeling approach. *Journal of Phonetics*, *88* (1): 101093.

Fleischman, Suzanne & Waugh, Linda R. (2016). *Discourse pragmatics and the verb: The evidence from Romance*. Routledge.

Foley, William A. & Olson, Mike (1985). Clausehood and verb serialization. In Johanna Nichols & Anthony C. Woodbury (Eds.), *Grammar inside and outside the clause* (pp. 17–60). Cambridge University Press.

Foley, William A. (2010). Clause linkage and nexus in Papuan languages. In Isabelle Bril (Ed.) *Clause linking and clause hierarchy: Syntax and pragmatics* (pp. 27–101). John Benjamins Publishing Company.

Gelman, Andrew & Rubin, Donald B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7* (4), 457–472.

Grafmiller, Jason, Szmrecsanyi, Benedikt, and Hinrichs, Lars (2018). Restricting the restrictive relativizer: constraints on subject and non-subject English relative clauses. *Corpus Linguistics and Linguistic Theory*, *14*(2), 309–355.

Gries, Stefan Th. (2010). Useful corpus-linguistics statistics. In Aquilinio Sánchez & Moisés Almela (Eds.), *A mosaic of corpus linguistics: Selected approaches* (pp. 269–291). Peter Lang.

Guajardo, Gustavo (2021). Co-occurrence Strength and Transitivity Effects on Spanish Clitic Case Variation With Reverse-Psychological Predicates. *Frontiers in Psychology*, *12*:712959.

Guy, Gregory (1988). Advanced VARBRUL analysis. In Kathleen Ferrara, Becky Brown, Keith Walters, & John Baugh *Linguistic change and contact: NWAV-XVI (Texas Linguistic Forum 30)* (pp. 124–136). University of Texas, Department of Linguistics.

Halsey, Lewis G. (2019). The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum?. *Biology letters*, *15* (5), 20190174.

Hamel, Patricia J. (1993). Serial Verbs in Loniu and an Evolving Preposition. *Oceanic Linguistics*, *31* (1), 111–132. https://www.jstor.org/stable/3623099.

Haspelmath, Martin (2016). The serial verb construction: Comparative concept and cross-linguistic generalizations. *Language and Linguistics*, *17* (3), 291–319.

Heine, Bernd & Kuteva, Tania (2002). *World lexicon of grammaticalization*. Cambridge University Press.

Lazard, Gilbert (2002). Transitivity revisited as an example of a more strict approach in typological research. *Folia Linguistica*, *36* (3–4), 141–190.

Lemoine, Nathan P. (2019). Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128* (7), 912–928.

Levinson, Stephen C. (2003). *Space in language and cognition: Explorations in cognitive diversity*. Cambridge University Press.

Levshina, Natalia (2016). When variables align: A Bayesian multinomial mixed-effects model of English permissive constructions. *Cognitive Linguistics*, *27* (2), 235–268.

Levshina, Natalia (2019). *Comparing Bayesian and frequentist models of language variation: The case of help + (to) infinitive*. Unpublished manuscript.

Longacre, Robert E. (2007). Sentences as combinations of clauses. In Timothy Shopen (Ed.), *Language Typology and Syntactic Description* (pp. 372–420). Cambridge University Press.

MacKenzie, Laurel (2020). Comparing Constraints on Contraction Using Bayesian Regression Modeling. *Frontiers in artificial intelligence*, *3*, *58*.

Meyerhoff, Miriam (2013). Syntactic variation and change: The variationist framework and language contact. In Isabelle Léglise & Claudine Chamoreau (Eds.), *The interplay of variation and change in contact settings* (pp. 23–52). John Benjamins Publishing Company.

Meyerhoff, Miriam (2015). Turning variation on its head: Analysing subject prefixes in Nkep (Vanuatu) for language documentation. *Asia-Pacific Language Variation*, *1*(1), 78 – 108.

Næss, Åshild (2011). Directional verbs in Vaeakau-Taumako. *Oceanic Linguistics*, *50*(1), 120–139.

Nalborczyk, Ladislas, Batailler, Cédric, Lœvenbruck, Hélène, Vilain, Anne, & Bürkner, Paul-Christian (2019). An introduction to Bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard Indonesian. *Journal of Speech, Language, and Hearing Research*, *62*(5), 1225–1242.

Nicenboim, Bruno, and Vasishth, Shravan (2016). Statistical methods for linguistic research: Foundational Ideas – Part II. *Language and Linguistics Compass*, *10*(11), 591–613.

Pawley, Andrew (2003). Grammatical categories and grammaticisation in the Oceanic verb complex. In Anastasia Riehl and Thess Savella (Eds.), *Cornell Working Papers in Linguistics*, *19*, (pp. 149–172).

Pawley, Andrew & Hammarström, Harald (2017). The Trans New Guinea family. In Bill Palmer (Ed.) *The languages and linguistics of the New Guinea Area : A comprehensive guide* (pp. 21–156). De Gruyter.

R Core Team (2021). *R: A Language and Environment for Statistical Computing [Computer software]*. R Foundation for Statistical Computing. https://www.R-project.org

Ross, Malcolm (2004a). The grammaticalization of directional verbs in Oceanic languages. In Isabelle Bril & Françoise Ozanne-Rivierre (Eds.), *Complex predicates in Oceanic languages: Studies in the dynamics of binding and boundness* (pp. 297–329). Mouton de Gruyter.

Ross, Malcolm (2004b). Demonstratives, local nouns and directionals in Oceanic languages: A diachronic perspective. In Gunter Senft (Ed.), *Deixis and demonstratives in Oceanic languages* (pp. 175–204). Pacific Linguistics.

Ross, Malcolm (2008). A history of metatypy in the Bel languages. *Journal of Language Contact*, *2*(1), 149–164.

Ross, Malcolm (2013). Takia. In John Lynch, Malcolm Ross & Terry Crowley (Eds.), *The Oceanic languages* (pp. 216–248). Routledge.

San Roque, Lila, Gawne, Lauren, Hoenigman, Darja, Miller, Julia C., Spronck, Stef, Rumsey, Alan, Carroll, Alice & Evans, Nicholas (2012). Getting the story straight: Language fieldwork using a narrative problem-solving task. *Language Documentation and Conservation*, *6*, 135–174. https://hdl.handle.net/2066/139704

Schokkin, Dineke (2013). Directionals in Paluai: Semantics, use, and grammaticalization paths. *Oceanic Linguistics*, *52*(1), 169–191.

Speelman, Dineke (2014). Logistic regression: A confirmatory technique for comparisons in corpus linguistics. In Dylan Glynn & Justyna A. Robinson, *Corpus methods for semantics: Quantitative studies in polysemy and synonymy* (pp. 487–533). John Benjamins Publishing Company.

Speelman, Dirk, Heylen, Kris, & Geeraerts, Dirk (Eds.) (2018). *Mixed-effects regression models in linguistics*. Springer.

Sperlich, Wolfgang B. (1993). Serial verb constructions in Namakir of Central Vanuatu. *Oceanic Linguistics*, *32* (1), 95–110. http://www.jstor.org/stable/3623098.

Tagliamonte, Sali A., & Baayen, R. Harald (2012). Models, forests, and trees of York English: Was/ were variation as a case study for statistical practice. *Language Variation and Change*, *24*(2), 135–178.

Tagliamonte, Sali A., & Smith, Jennifer (2006). Layering, competition and a twist of fate: Deontic modality in dialects of English. *Diachronica*, *23*(2), 341–380.

Talmy, Leonard (2000). *Toward a cognitive semantics: Vol. II: Typology and process in concept structuring*. MIT press.

Terrell, Tracy D., Tschirner, Erwin & Nikolai, Brigitte (2005). *Picture File to accompany Kontakte: A Communicative Approach, 5th edition*. The McGraw-Hill Companies, Inc.

Traugott, Elizabeth C., & Trousdale, Graeme. (2010). Gradience, gradualness and grammaticalization: How do they intersect? In Elizabeth C. Traugott & Graeme Trousdale (Eds.), *Gradience, gradualness and grammaticalization* (pp. 19, 44.). Amsterdam: John Benjamins.

van Ravenzwaaij, Don, Cassey, Pete & Brown, Scott D. (2018). A simple introduction to Markov Chain Monte–Carlo sampling. *Psychon Bull Rev*, *25*, 143–154.

Vasishth, Shravan & Nicenboim, Bruno (2016). Statistical methods for linguistic research: Foundational ideas–Part I. *Language and Linguistics Compass*, *10*(8), 349–369.

Vasishth, Shravan, Chen, Zhong, Li, Qiang & Guo, Gueilan (2013). Processing Chinese relative clauses: Evidence for the subject-relative advantage. *PloS one*, *8*(10), 1–15.

Vehtari, Aki, Gelman, Andrew & Gabry, Jonah (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput*, *27*, 1413–1432.

Weinreich, Uriel, Labov, William & Herzog, Marvin I. (1968). Empirical foundations for a theory of language change. In Winfred P. Lehmann & Yakov Malkiel (Eds.), *Directions for historical linguistics* (pp. 95–195). University of Texas Press.

Wolk, Christoph, Bresnan, Joan, Rosenbach, Anette & Szmrecsanyi, Benedikt (2013). Dative and genitive variability in Late Modern English: Exploring cross-constructional variation and change. *Diachronica*, *30*(3), 382–419.

Appendix 1

An inspection of the MCMC trace plots is an important aspect of assessing the convergence of a Bayesian model. Figure A1 below shows how MCMC trace plots look when the model converges compared to when it does not. At convergence the trace plots resemble "fat, hairy caterpillars" (Lunn et al., 2013 in Levshina, 2016, p. 253).

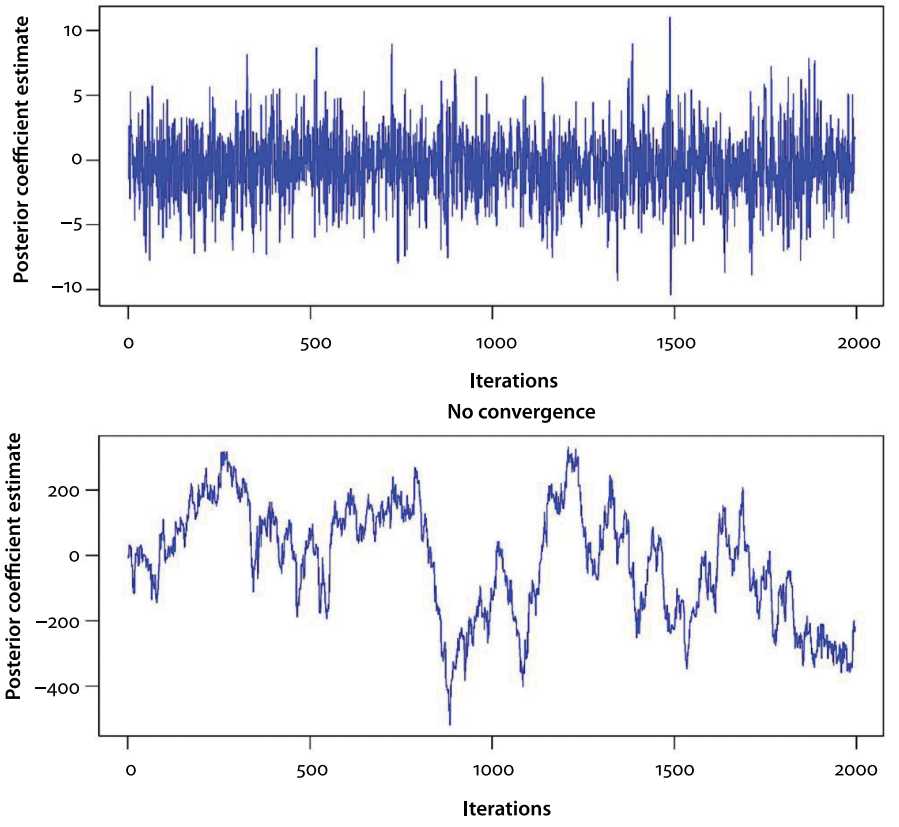
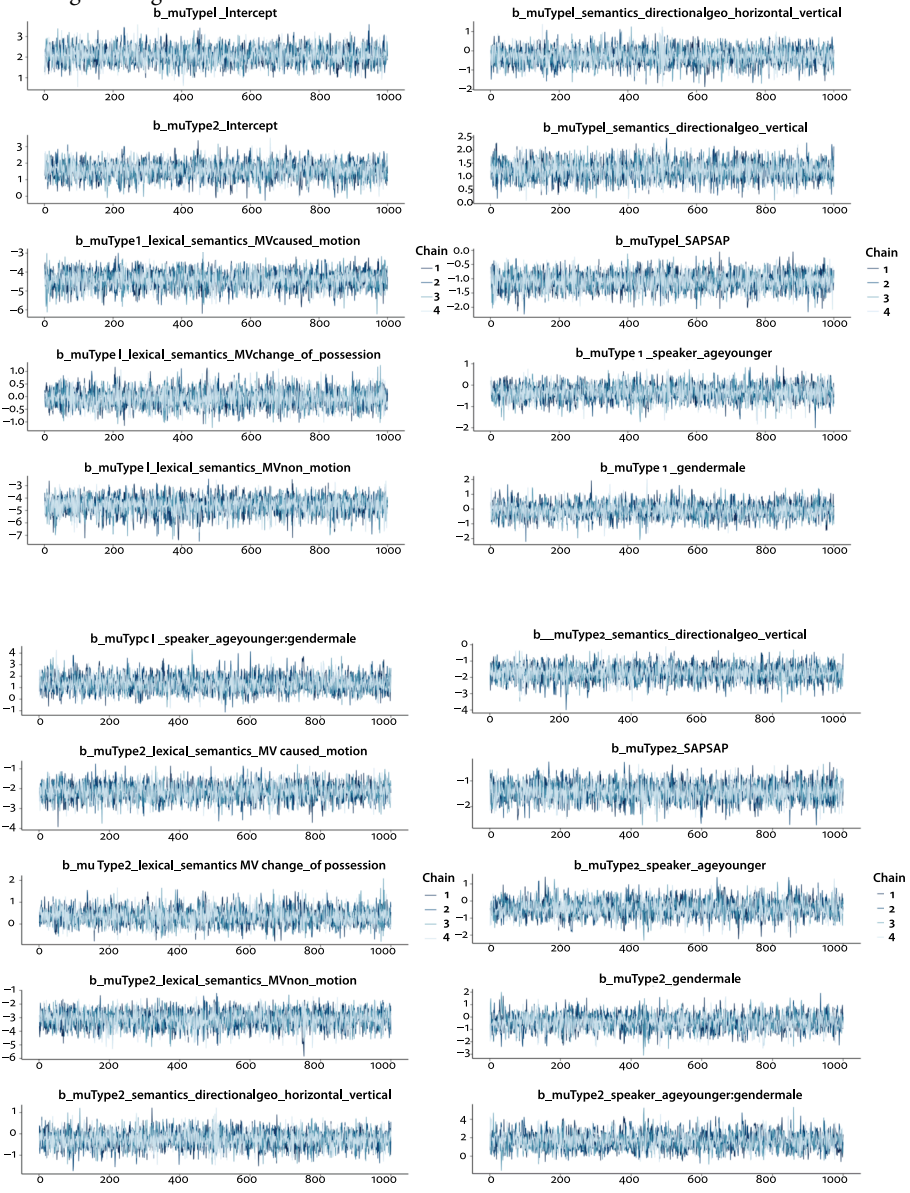


Figure A1. MCMC trace plots when a Bayesian model converges and does not converge (from Levshina, 2019, p. 8)

The trace plots below in A2 are from the Bayesian model used in this article, with all trace plots showing convergence.



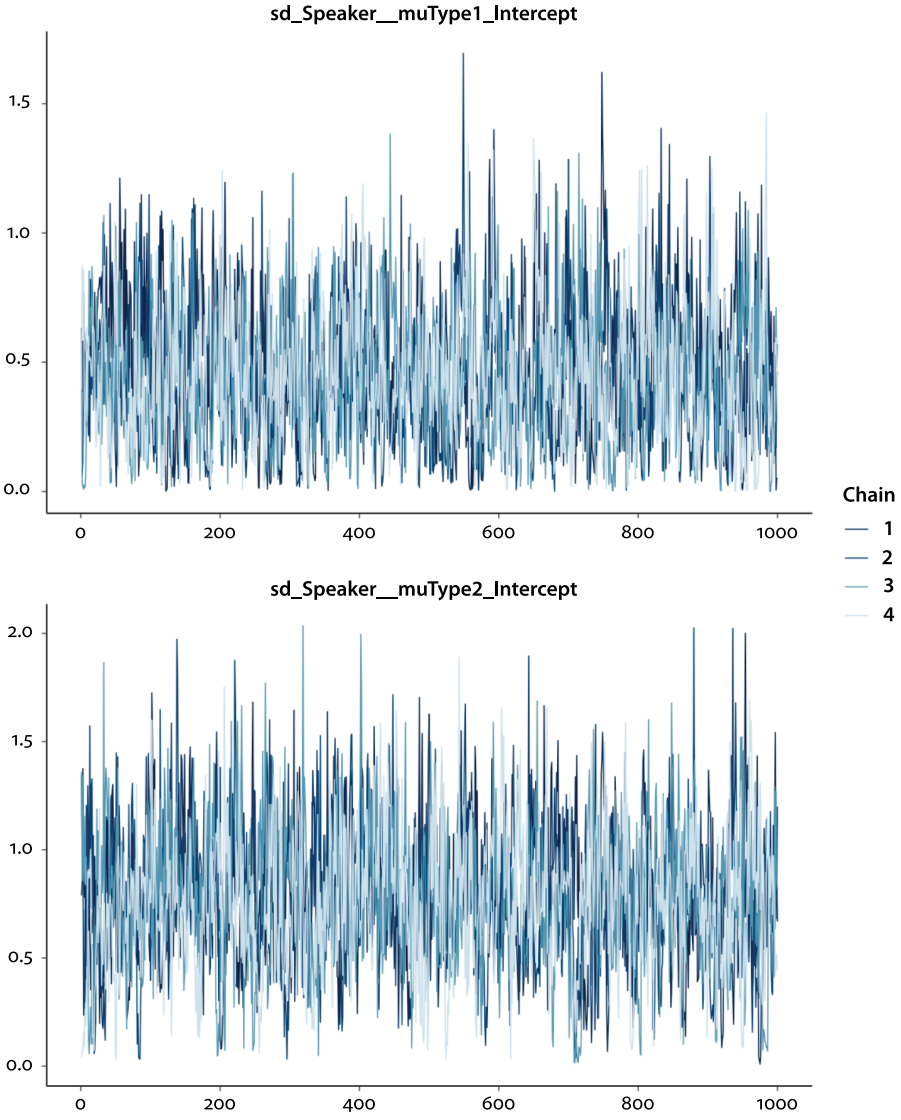


Figure A2. MCMC trace plots for the Bayesian regression model Appendix 2

Appendix 2

Normal (0.10)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(muType1_Intercept)	0.51	0.28	0.04	1.13	1.00	737	800
sd(muType2_Intercept)	0.77	0.33	9.15	1.48	1.01	959	1175

Population-level effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tai_ESS
muType1_Intercept	2.63	0.50	1.68	3.68	1.00	2785	2945
muType2_Intercept	2.30	0.58	1.14	3.39	1.00	2419	2327
muType1_lexical_semantics_MVcaused_motion	−4.71	0.48	−5.71	−3.80	1.00	3848	2641
muType1_lexical_semantics_MVchange_of_possession	−0.02	0.41	−0.80	0.81	1.00	4144	3583
muType1_lexical_semantics_MVnon_motion	−5.05	0.75	−6.66	−3.74	1.00	4654	2680
muType1_semantics_directionalgeo_horizontal_vertical	−0.38	0.51	−1.36	0.63	1.00	4460	2969
muType1_semantics_directionalgeo_vertical	1.29	0.36	0.59	2.02	1.00	4371	3164
muType1_SAPSAP	−1.34	0.32	−1.99	−0.71	1.00	3715	3216
muType1_speaker_ageyounger	−0.84	0.49	−1.77	0.13	1.00	2838	2503
muType1_gendermale	−0.81	0.66	−2.10	0.52	1.00	2489	2413
muType1_speaker_ageyounger:gendermale	2.66	0.89	0.98	4.43	1.00	2259	2669
muType2_lexical_semantics_MVcaused_motion	−2.50	0.45	−3.44	−1.65	1.00	4060	3225
muType2_lexical_semantics_MVchange_of_possession	0.39	0.43	−0.43	1.25	1.00	3655	3194
muType2_lexical_semantics_MVnon_motion	−3.55	0.66	−4.94	−2.33	1.00	4725	3238
muType2_semantics_directionalgeo_horizontal_vertical	−0.37	0.48	−1.31	0.57	1.00	4318	2789
muType2_semantics_directionalgeo_vertical	−1.80	0.49	−2.78	−0.88	1.00	4012	3028
muType2_SAPSAP	−1.69	0.37	−2.42	−1.01	1.00	4669	3494
muType2_speaker_ageyounger	−1.07	0.58	−2.23	0.05	1.00	2633	2581
muType2_gendermale	−1.51	0.84	−3.22	0.20	1.00	2522	3015
muType2_speaker_ageyounger:gendermale	3.41	1.06	1.38	5.59	1.00	2269	2398

Normal (0.1)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(muType1_Intercept)	0.36	0.22	0.02	0.85	1.00	1066	1790
sd(muType2_Intercept)	0.76	0.31	0.14	1.41	1.00	806	989

Population-level effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
muType1_Intercept	1.73	0.36	1.03	2.44	1.00	4119	3035

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
muType2_Intercept	1.18	0.45	0.24	2.01	1.00	3365	3070
muType1_lexical_semantics_MVcaused_motion	−3.62	0.37	−4.36	−2.92	1.00	6183	3447
muType1_lexical_semantics_MVchange_of_possession	0.17	0.34	−0.49	0.84	1.00	5774	3284
muType1_lexical_semantics_MVnon_motion	−3.26	0.45	−4.18	−2.41	1.00	7663	3187
muType1_semantics_directionalgeo_horizontal_vertical	−0.53	0.42	−1.33	0.29	1.00	6157	2740
muType1_semantics_directionalgeo_vertical	1.17	0.28	0.62	1.72	1.00	5775	3425
muType1_SAPSAP	−0.88	0.26	−1.41	−0.37	1.00	6709	3454
muType1_speaker_ageyounger	−0.28	0.34	−0.96	0.38	1.00	4302	3190
muType1_gendermale	0.01	0.45	−0.83	0.92	1.00	3506	3217
muType1_speaker_ageyounger:gendermale	1.05	0.55	−0.03	2.13	1.00	4355	3192
muType2_lexical_semantics_MVcaused_motion	−1.64	0.36	−2.33	−0.93	1.00	5643	3398
muType2_lexical_semantics_MVchange_of_possession	0.61	0.35	−0.07	1.34	1.00	4890	3341
muType2_lexical_semantics_MVnon_motion	−2.18	0.48	−3.17	−1.28	1.00	6829	2768
muType2_semantics_directionalgeo_horizontal_vertical	−0.44	0.40	−1.27	0.33	1.00	6280	3233
muType2_semantics_directiorialgeo_vertical	−1.64	0.40	−2.45	−0.89	1.00	5258	3194
muType2_SAPSAP	−1.22	0.30	−1.82	−0.64	1.00	5988	3514
muType2_speaker_ageyounger	−0.32	0.45	−1.16	0.60	1.00	3443	3038
muType2_gendermale	−0.22	0.55	−1.27	0.91	1.00	3126	3155
muType2_speaker_ageyounger:gendermale	1.22	0.66	−0.10	2.49	1.00	2921	2642

student_t(3, 0, 2.5)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(muType1_Intercept)	0.47	0.27	0.03	1.05	1.00	766	1545
sd(muType2_Intercept)	0.76	0.34	0.10	1.47	1.01	626	697

Population-level effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
muType1_Intercept	2.35	0.46	1.43	3.24	1.00	3188	3157
muType2_Intercept	1.94	0.54	0.85	2.98	1.00	2718	3058
muType1_lexical_semantics_MVcaused_motion	−4.51	0.46	−5.48	−3.62	1.00	4296	3426
muType1_lexical_semantics_MVchange_of_possession	0.01	0.40	−0.76	0.79	1.00	3957	3019
muType1_lexical_semantics_MVnon_motion	−4.65	0.69	−6.11	−3.39	1.00	4889	2680
muType1_semantics_directionalgeo_horizontal_vertical	−0.40	0.49	−1.38	0.57	1.00	4765	3399
muType1_semantics_directionalgeo_vertical	1.27	0.34	0.63	1.96	1.00	5023	3329
muType1_SAPSAP	−1.22	0.32	−1.85	−0.58	1.00	3714	2743
muType1_speaker_ageyounger	−0.61	0.45	−1.52	0.27	1.00	3336	3017
muType1_gendermale	−0.45	0.60	−1.59	0.78	1.00	2562	2042
muType1_speaker_ageyounger:gendermale	2.02	0.80	0.48	3.64	1.00	2567	3030

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
muType2_lexical_semantics_MVcaused_motion	−2.29	0.44	−3.18	−1.44	1.00	4429	3090
muType2_lexical_semantics_MVchange_of_possession	0.44	0.42	−0.35	1.26	1.00	3481	3182
muType2_lexical_semantics_MVnon_motion	−3.21	0.60	−4.45	−2.08	1.00	3964	3144
muType2_semantics_directionalgeo_horizontal_vertical	−0.39	0.47	−1.30	0.52	1.00	4941	3273
muType2_semantics_directionalgeo_vertical	−1.77	0.47	−2.72	−0.87	1.00	4481	2812
muType2_SAPSAP	−1.57	0.36	−2.28	−0.89	1.00	4564	3164
muType2_speaker_ageyounger	−0.79	0.54	−1.86	0.30	1.00	2635	3040
muType2_gendermale	−0.98	0.76	−2.44	0.54	1.00	2576	2757
muType2_speaker_ageyounger:gendermale	2.58	0.95	0.75	4.44	1.00	2510	2989

Cauchy (0.1)

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sd(muType1_Intercept)	0.45	0.27	0.03	1.05	1.00	878	1733
sd(muType2_Intercept)	0.80	0.33	0.14	1.47	1.00	635	700

Population-level effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
muType1_Intercept	2.08	0.40	1.33	2.92	1.00	3562	2797
muType2_Intercept	1.57	0.49	0.60	2.52	1.00	2660	2776
muType1_lexical_semantics_MVcaused_mot	−4.43	0.48	−5.43	−3.56	1.00	4977	2957
muType1_lexical_semantics_MVchange_of_possession	−0.07	0.34	−0.73	0.64	1.00	4934	3260
muType1_lexical_semantics_MVnon_mot	−4.61	0.73	−6.20	−3.29	1.00	5182	2633
muType1_semantics_directionalgeo_horizontal_vertical	−0.28	0.42	−1.16	0.54	1.00	6449	3463
muType1_semantics_directionalgeo_vertical	1.24	0.35	0.57	1.94	1.00	5005	3340
muType1_SAPSAP	−1.09	0.31	−1.70	−0.48	1.00	4905	3379
muType1_gendermale	−0.08	0.50	−1.07	0.89	1.00	3564	2816
muType1_speaker_ageyounger	−0.34	0.38	−1.12	0.41	1.00	3779	3064
muType1_gendermale:speaker_ageyounger	1.35	0.71	0.10	2.81	1.00	3558	2909
muType2_lexical_semantics_MVcaused_mot	−2.17	0.43	−3.05	−1.36	1.00	4597	3451
muType2_lexical_semantics_MVchange_of_possession	0.34	0.37	−0.38	1.08	1.00	4104	2918
muType2_lexical_semantics_MVnon_mot	−3.13	0.64	−4.43	−1.98	1.00	4962	3293
muType2_semantics_directionalgeo_horizontal_vertical	−0.24	0.41	−1.06	0.54	1.00	6280	3309
muType2_semantics_directionalgeo_vertical	−1.73	0.48	−2.71	−0.81	1.00	5068	3202
muType2_SAPSAP	−1.42	0.36	−2.16	−0.74	1.00	5581	3277
muType2_gendermale	−0.38	0.66	−1.78	0.89	1.00	2537	2644
muType2_speaker_ageyounger	−0.39	0.48	−1.33	0.55	1.00	2826	2952
muType2_gendermale:speaker_ageyounger	1.64	0.90	0.07	3.47	1.00	2389	2456

Cauchy priors were specified in the model. While the estimates vary slightly, the direction of the effect and the degree to which we can be confident in the results are the same for all of the weak priors tested. This means that the model is robust, but we can be less confident in the size of the effect than the direction of the effect.

Abstract (German)

Die ozeanische Sprache Matukar Panau hat drei gleichwertige morphosyntaktische Strategien zur Richtungsbeschreibung eines Ereignisses das durch ein Verb ausgedrückt wird. Das Verb kann in Verbindung mit einem von zehn möglichen direktionalen Morphemen auftretten. Diese Variationen werden hier anhand des Matukar-Panau-Korpus mit der Methode der Bayesianischen Regression analysiert. Faktoren, welche die Entscheidung für die eine oder andere Variation beeinflussen, werden bestimmt. Variablen mit Einfluss auf das lexikalische Verb und die direktionalen Morpheme haben sich hierbei als wichtigste Faktoren herausgestellt, wohingegen sozioligusitische Faktoren eine gerinngere Rolle spielen. Die Ergebnisse haben Bedeutung für die Gramatikalisierung von direktionalen Elementen wie auch für die Typologie von Direktionalkonstruktionen der ozeanischen Sprachen alsgemein.

Address for correspondence

Kira Davey
School of Culture, History and Language
Australian National University
Canberra, ACT, 2601
Australia

kira.davey@anu.edu.au
https://orcid.org/0009-0002-1542-364X

Co-author information

Danielle Barth
School of Culture, History and Language
Australian National University

danielle.barth@anu.edu.au

Publication history

Date received: 1 March 2023
Date accepted: 6 August 2023
Published online: 11 January 2024