Language Variation and Change, **32** (2020), 267–291.
© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited. 0954-3945/20 $16.00
doi:10.1017/S0954394520000125

Expression of anaphoric subjects in Vera'a: Functional and structural factors in the choice between pronoun and zero

STEFAN SCHNELL
Otto-Friedrich-Universität Bamberg & Centre of Excellence for the Dynamics of Language

DANIELLE BARTH
Australian National University & Centre of Excellence for the Dynamics of Language

ABSTRACT

The choice between a pronoun and zero anaphor for the expression of third-person subjects is examined in a corpus of Vera'a (Oceanic). While predominantly expressed by a pronoun, subjects are found to permit zero form with referents that have low anaphoric distance. Within this context, zero is found to be preferred with a subset of verbal predicates that take a specific tense-aspect-mood-polarity (TAMP) marker that historically retains subject agreement. The strong preference for pronouns is related to the clitic behavior of adjacent TAMP morphology and the rudimentarity of agreement. Animacy and number also bear on subject variation. Effects of clause-combining and the use of connectives do not align with findings from studies of the same choice in other languages. Our findings underscore the prominent role of purely structural over functional motivations for the choice of pronouns over zero.

The instantiation of subjects as either an overt form of expression or zero anaphor is one of the most widely studied cases of structural variation across and within languages, both from a theoretical (Roberts & Holmberg, 2010) and a typological (Dryer, 2013) perspective. While some previous work had established a distinction between null-subject and nonnull-subject languages (Rizzi, 1982), work in the variationist tradition has cast doubts on this crosslinguistic distinction (e.g., Torres Cacoullos & Travis, 2019): although zero subjects in discourse from, for instance, Spanish are much more frequent than in English discourse, both languages show considerable variation and are subject to similar discourse, semantic, as well as structural factors. Despite some research on non-European languages (e.g., Li & Bayley [2018] on Mandarin), including some lesser-studied languages (e.g., Meyerhoff, 2000, 2009), most of the discussion of variable subject realization to date has focused on Indo-European, in particular Germanic and Romance, languages.

In this paper, we broaden the typological scope of the variationist research on variable subject realization in reporting on a case study of Vera'a, an Oceanic

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

language spoken by approximately 500 speakers in North Vanuatu. While first- and second-person subjects are nearly categorically pronouns (Schnell, 2018), third-person subjects show substantial variation in the use of pronouns and zero anaphor, as illustrated in (1): In (1)a, the subject position is filled by the personal NP *e Qo'*, and it precedes the predicate which is introduced by an *iamitive* ("already" perfect) marker; in (1)b the same position is held by the pronoun *dir*. In the second clause in each example, the subject position is left unfilled, and this zero is understood as coreferent with the preceding subject (in the examples, zero is indicated by ∅).

(1) a. *nei* [*e Qo'*]^{SBJ} [***man*** *sal gala dirē*[1]]^{PRED} ∅^{SBJ} [***man*** *van*]^{PRED}
now PERS *pers.name* IAM float past 3PL IAM go
'Then Qo' had already floated past them, (he) had already gone.' JJQ.228

b. [*dir*]^{SBJ} [***man*** *'ēqēl ma*]^{PRED} ∅^{SBJ} [***man*** *kalu*]^{PRED} [*den ēn wio*]^{OBL}
3PL IAM descend hither IAM exit ABL ART bamboo
'They had already come down, had already left the bamboos.' JJQ.346

One goal of our study is to determine the global range of factors driving the choice between pronoun and zero, leaving aside here the case of full lexical NP subject expression.[2] A specific structural aspect relevant in Vera'a is the variable presence of subject-predicate agreement. Although Gilligan (1987) showed that agreement cannot generally be seen as a necessary condition on the possibility to leave subjects zero across languages, works in the theoretical-generative tradition suggest that the variable presence of agreement in the predicate does relate to the possibility and greater likelihood of zero subject expression; see, for instance, Rosenkvist (2018) and Fuß (2005) on the greater likelihood of zero subjects with fully transparent and nonsyncretic subject-agreement morphology in Germanic and Romance; also Meyerhoff's (2009:308–9) finding that Tamambo (a Vanuatu language with full-fledged subject agreement) shows a significantly greater rate of zero subjects than Bislama. In Vera'a, only clauses in so-called *prospective aspect*[3] show subject-predicate agreement, with the predicate being introduced by a portmanteau morpheme expressing aspect as well as person/number of the subject. Hence, in (2)a, the subject is third-person singular, and the respective prospective allomorph *ne* is used, whereas, in (2)b, the subject is third-person plural and the prospective aspect form is *k*.

(2) a. [=*n reñe anē*]^{SBJ} [***ne*** *wotoqtoqo*]^{PRED}
=ART woman DEM.ADDR PROSP:3SG pregnant
∅^{SBJ} [***ne*** *visis*] [*ēn ni'i reñe*]^{OBJ}
PROSP:3SG give.birth ART small woman
'Then that woman got pregnant, and then (she) gave birth to a little girl.' ANV.003

b. [=*n 'erē 'ama' anē*]^{SBJ} [***=k*** *van lik ma*]^{PRED}
=ART PL spirit DEM.ADDR =PROSP:NSG go more hither
∅^{SBJ} [***=k*** *siñ qēqē lik*]^{PRED} ∅^{OBJ}
=PROSP:NSG set.fire RED:finish more
'Then all the devils arrived and lit the rest of them (the coconut leaves).' JSV.E.096

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

Examples in (2) showing subject-predicate agreement thus contrast with those in (1): unlike prospective aspect, iamitive *man* is used with subjects regardless of number and person. Given its theoretical prominence and some empirical support, we consider the presence of agreement—in the form of prospective aspect marking—as a separate factor that may bear significantly on the expression of subjects as zero anaphor. Hence, in addition to our first goal of determining the range of independent factors contributing to the variation, we also aim at clarifying the specific role of agreement.

We find a strong preference for pronouns in anaphoric subject expression, and our third goal is thus to account for this preference vis-à-vis findings from other comparative studies of subject expression (e.g., Barbosa, Duarte, & Kato, 2005; Meyerhoff, 2009; Torres Cacoullos & Travis, 2019). The paper is structured as follows: we first provide a short descriptive account of Vera'a clause-level grammar and subjects as relevant to our concerns. We then turn to factors pertaining to the alternation at hand, as discussed in the literature, and how we treated them in our corpus study. The corpus we are investigating consists of narrative texts that belong to the oral literature of the region (they are not stimulus-based). We then describe the statistical methodology applied and report the results before discussing our findings.

VERA'A SUBJECTS AND THE REMAINS OF SUBJECT-PREDICATE AGREEMENT

Vera'a (glottocode: vera1241; Oceanic, North Vanuatu) is an analytic, configurational language which encodes syntactic relations solely by means of rigid SVO word order. In (3) and (4), the subject NP precedes the predicate in both a transitive and an intransitive clause, whereas the object in (3) follows it.

(3) [*e ote*]^{SBJ} [***ne*** *sursur*]^{PRED} [*ēn nes*]^{OBJ}
PERS mum PROSP:3SG RED:sing ART song
'Mum is singing a song.' HHAK.077

(4) [*lie*]^{SBJ} [***ne*** *wak*]^{PRED}
cave PROSP:3SG open
'Then the cave opened.' 2.PALA.025

The verbal predicate is realized by a complex phrase typically consisting of at least a marker of tense, aspect, mood, and also polarity (TAMP) in initial position and a verb in second position. We term this structure *verb complex* (VC) here, following a well-established Oceanist tradition. The subject slot is separated from the VC by a further slot that hosts certain types of adverbs and conjunctions, as shown in (5) where the conjunction *wo* intervenes between the subject pronoun and the VC.

(5) [*di*]^{SBJ} ***wo***^{ADV} [***ne*** *'ēn*]^{PRED} [*ēn 'añsara*]^{OBJ}
3SG CONJ PROSP:3SG see ART person
'(When it comes out,) if it spots a human, (it will go back in (its cave).)
Fish112.E.003

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

This slot is, however, rarely filled (thirty instances among 2,489 clauses), so that in actual language use subject constituents are nearly always placed adjacently to the VC, with conjunctions more commonly occurring in clause-initial position.

TAMP morphology

VC-initial TAMP markers are free-standing grammatical words (rather than affixes) that inflect a predicate. Some of the markers have a morphological shape that fulfils the minimal requirement of a phonological word of CV syllable structure, for instance, the future particle *me* in (6). Other markers are clitics, consisting of a single consonant, with an alternant VC allomorph, the allomorphy being phonologically conditioned, as can be seen from the two forms of the perfect marker *=m* ~ *ēm* in (7) and the nonsingular prospective aspect forms *=k* ~ *ēk* in (8).

(6) [*n wēweñ anē*]^{SBJ} [***me*** *dam 'i*]^{PRED} *sa lē =n mē'ēm̃ē anē*
ART k.o.basket DEM.ADDR **FUT** hang DEL EMPH LOC =ART door DEM.ADDR
'Alright, so the *wēweñ* basket will just be hanging right in the doorway.'
ASMS.077

(7) [*'ama-ru*]^{SBJ} [***=m*** *van*]^{PRED} *ø*^{SBJ} [***=ēm*** *wana*]^{PRED}
father-3DL =PRF go =PRF squeeze.kava

ø^{SBJ} [***=m*** *wana*]^{PRED}
=PRF squeeze.kava
'Their (two) father had gone and been squeezing and squeezing (it; kava) out, [had been sitting in the men's house (where kava is drunk) on and on (until he came home)].'
MVBW.119

(8) a. [*n 'erē 'ama' anē*]^{SBJ} [***=k*** *van lik ma*]^{PRED}
ART PL spirit DEM.ADDR **=PROSP:NSG** go more hither
ø^{SBJ} [***=k*** *siñ qēqē lik*]^{PRED} *ø*^{OBJ}
=PROSP:NSG light RED:finish more
'Then all the devils arrived and lit the rest of them (the coconut leaves).'
JSV.E.096

b. [*dir*]^{SBJ} [***=ēk*** *vusvus kel dir*]^{PRED} *sir*
3PL **=PROSP:NSG** RED:kill back 3PL for
'They would kill each other for (it).'
ISWM.073

The phonological conditioning of allomorphy in *m* and *k* is dependent on the *proceeding* sound, so that the vowel-initial allomorph occurs after a consonant-final word and the consonant-only allomorph after a vowel-final word. It is in this sense that we classify clitic TAMP markers as inherently *enclitic* rather than *proclitic*: they form a phonological word together with any *preceding* word, be that a pronoun, as in (8)b, or some constituent of a subject NP, for example, the head noun *'amaru* in (7) or the demonstrative *anē* in (8)a. Given that TAMP clitics occupy a phrase-initial position, cliticization thus crosses a phrase-structure border. Such clitics are termed *detached* (Bickel &

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

Nichols, 2007) or *ditopic* (Himmelmann, 2014): they attach phonologically to an element outside the phrase that they have functional scope over.[4] As can be seen from (7) and (8a), respectively, clitic TAMP markers also occur with zero subjects, and, in these cases, their host word can be the final word of the preceding clauses, namely the verbs *van* 'go' or *wana* 'squeeze' in (7) resulting in the phonological words *vanēm* and *wanam*, respectively, or the directional particle *ma* 'hither' in (8)a, resulting in *mak*. Hence, the realization of subjects as pronoun or zero is clearly connected to the status of TAMP markers as particles or clitics, since pronouns are a potential host for TAMP clitics. There can, however, also be a conjunction occurring between subject and VC, as in (9). Only eighteen such cases of cliticization to an intervening conjunction occur in the corpus of 2,489 clauses.

(9) *ba* [*duru*]^{SBJ} [***wo***]^{CONJ} [***=m*** *vanvan*]^{PRED}
but 3DL CONJ =PRF RED:go
'Once they move around, (they might get all around the island).' Fish92_JH.012

Finally, note that the range of fifteen TAMP categories is rarely exploited in language use. Only two, the prospective and the perfect aspect, account for 90% of all verbal clauses in our corpus; the future (marked by a particle *me*) also occurs with some frequency.

Retained subject agreement

Like many other Oceanic languages (e.g., Hyslop [2001] on Lolovoli; Jauncey [2011] on Tamambo; Thieberger [2006] on Nafsan), Vera'a retains a form of subject-predicate agreement from Proto-Oceanic (POc) (Lynch, Ross, & Crowley, 2011:67), namely in the form of prospective aspect marking: if a predicate is marked for prospective aspect, the form depends on the person/number of the subject, as per the paradigm in Table 1. The paradigm has person distinctions in the singular, and all nonplural forms show a somewhat quirky syncretism with the first person singular in the form of *=k*.

In (10) the prospective marker is third-person singular *ne*, and in (11) it is the nonsingular clitic form *=k*.

(10) PROSPECTIVE ASPECT – 3RD PERSON SINGULAR SUBJECT (PARTICLE)
[*n* *'añsara*]^{SUBJ} [***ne*** *van* *wi-wiēg*]^{VC}
ART person PROSP:**3SG** go RED-work
'Then the person (who is on that day) would go work.' GARP.174

(11) PROSPECTIVE ASPECT – 3RD PERSON NON-SINGULAR SUBJECT (CLITIC)
[*'erē* *'añsara*]^{SUBJ} [***=k*** *'evē* *su-sur* *va'i*]^{VC} [*=n 'erē wōqa'aga*]^{OBJ}
PL person =PROSP:**NSG** throw RED-down still =ART PL post
'The people are still throwing down all the posts for it.' house29JAN0705.006

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

TABLE 1. *Paradigm of Vera'a prospective plus subject agreement marker*

	Singular	Nonsingular
1		
2	*ē*	*k*
3	*ne*	

Prospective aspect alone accounts for 50% of all verbal clauses, which means that, in half of the verbal clauses, we find subject-predicate agreement in TAMP marking. Hence, we find a contrast between clauses with subject-predicate agreement as in (8), (10), and (11) and those without subject-predicate agreement, as in (6), (7), and (9).

Interim summary

In sum, Vera'a grammar shows a number of structural properties immediately relevant to the realization of subjects as pronoun or zero anaphor: subjects are usually adjacent to the VC and subject pronouns are an ideal candidate as a TAMP clitic host, with the potential for these two forms to be treated as a formal unit. Since TAMP morphology variably co-expresses person/number features of subjects, we can hypothesize that this is ideally done only once per clause under the assumption that redundancy be avoided. Table 2 summarizes possible constellations of the two forms of realization under investigation, pronoun and zero, and three categories of TAMP morphology, involving contrasts between particle and clitic form, and those between agreement in number or not. Hence, with any given constellation, we find a predictable host for an enclitic in the form of a pronoun or not, we find number overtly expressed (and hence some explicit reference to the subject participant) or not, and we find either a preferred constellation of expressing such information only once, or doubled. The checked cells represent ideal constellations for any of these three properties, and the exclamation marks those that theoretically should be disfavored.

FACTORS DETERMINING THE CHOICE BETWEEN PRONOUN AND ZERO

Languages have been traditionally classified in terms of subject expression as nonnull-subject languages like English or German that basically require an overt subject in most contexts and, thus, show a high rate of overt expressions in discourse, and null-subject languages like Spanish or Italian that lack such a rule and favor zero (Rizzi, 1982; Roberts & Holmberg, 2010). More detailed distinctions are occasionally found in the typological literature, for instance Bickel's (2003) classification of languages according to their degree of overt expressions (of all arguments, including subjects), a finer-grained version of

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

TABLE 2. *Hypothesized preferences of subject realization in relation to cliticization and agreement*

		TAMP CATEGORY	HOST CLITIC	EXPRESS NUMBER	AVOID NUM DOUBLING
di	*ne*	prospective	-	√	!
ø	*ne*	prospective	-	√	√
dir	*=(ē)k*	prospective	√	√	!
ø	*=(ē)k*	prospective	!	√	√
di / dir	*me*	future	-	√	√
ø	*me*	future	-	!	√
di / dir	*=(ē)m*	perfect	√	√	√
ø	*=(ē)m*	perfect	!	!	√

Ross' (1982) typology of 'hot,' 'medium,' and 'cool' languages, or Dryer's (2013) classification that takes possible combinations of clause-level subject expressions and co-present agreement into account. Common to all these works, however, is their essentially holistic approach, seeking to capture a language's profile in terms of reference production rather than its specific conditions.

Variationist work has focused on specific factors applicable mostly to individual languages (e.g., Cameron & Flores-Ferrain, 2004; Meyerhoff, 2000, 2009; Travis & Lindstrom, 2016). More recently, Torres Cacoullos and Travis (2019) have proposed a way to overcome the gap between variationist studies of language-internal variation and cross-language comparison. Comparing subject expression patterns in discourse from two languages, English and Spanish, they found that, besides certain idiosyncratic conventions, for example, the restrictions of zero subjects in English to clausal chains and initial position in intonation units in declarative main clauses, which are absent in Spanish (Torres Cacoullos & Travis, 2019:668), the choice between pronoun and zero is sensitive to the same factors across the two languages, which, however, exhibit differences in the contexts of variation as well as in the magnitudes of impact (Torres Cacoullos & Travis, 2019:671–3). In determining these details of variation, the authors arrive at an empirically more satisfying account of the observed differences in subject pronoun rates across the two languages than traditional approaches. In what follows, we outline the factors found to be relevant for subject expression, in keeping with the goal of our study to situate our findings in a comparative perspective in accounting for pronoun preference in Vera'a subjects.

One set of discourse-related factors pertains to referent accessibility, a concept established building on Chafe's (1976) seminal work. On this view, the choice between pronoun and zero is a matter of recipient design, the contrast being essentially of the same nature as that of the choice between a lexical and any kind of nonlexical expressions (see, e.g., Ariel, 1990:74–81). This view is reflected in Givón's (1983) work on referent tracking and persistence as well as

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

Ariel's (2014[1990]) Accessibility Theory (AT). According to Givón (1983) then, more explicit expressions are used for new mentions, followed by less material for subsequent mentions. This would result in chains of full NPs followed by pronouns in turn followed by zero anaphor (cf., Cameron & Flores-Ferrán, 2004:50), thus reflecting a rise in accessibility. Tracking would then continue with zero until accessibility of the referent drops and a more explicit form is to be used again, and so forth. Relevant factors of accessibility are the distance between mentions of a referent, continuity/change in syntactic function, presence of competing referents, and—to some extent—their form of expression (see Ariel, 1990: Chapter 2). As pointed out by many scholars in this area, the most central environment is one where these four factors converge in same-subject chains in consecutive clauses (Givón's [2015] *chain-medial* clauses with same-subject referent continuity). Hence, zero is overall preferred in contexts of one clause anaphoric distance, continuity of subject function and where the previous mention has been zero (cf., Li & Bayley, 2018:151; Torres Cacoullos & Travis, 2019:672–4, among many others).

Related to antecedent distance and function, in particular, the chaining of co-referent subjects is the spatiotemporal coherence of sentences in discourse. On this view, co-reference relations interact with temporal sequencing, so that the most predictable referent is one which is a continuous topic (as per Givón, 1983) in a sequence of foregrounded, narrative clauses involving the same referent(s), whereas referents in clauses involving background information and the like are less accessible. Previous studies, for instance Myhill (1997), found that zero subjects are particularly likely in temporal-sequencing contexts, in particular with chains of co-referent subjects in this type of context (see also Torres Cacoullos & Travis [2019:672–3] on semantic refinements in defining the same-subject chain context). Connected to semantic aspects of sequencing are structural aspects of clause chaining: thus, in some languages, zero subjects are more likely in clauses that are overtly coordinated, as is the case in English according to Torres Cacoullos and Travis (2019:661).

Other factors that have been found to bear on the choice between a pronoun and zero are those pertaining to semantic properties of referents, for instance, their animacy and number. Here, it has been found that, in many languages, reference to human (or animate) beings is preferably made by means of a pronoun, whereas nonhuman (inanimate) reference is preferably made by zero (e.g., Genetti & Crain [2003] on Nepali). This tendency is more typically associated with objects than with subjects (cf., Schnell & Barth, 2018), and it is also reflected in hierarchical splits in object agreement systems, that is, indexing-based DOM (Haig, 2018; Siewierska, 2004:145–62), where nonhumans, or inanimates, etc., are less likely to show agreement. Similarly, paradigmatic zeroes in verbal paradigms are typically restricted to the singular (e.g., DuBois [1987] on Sakapultek), indicating that zeroes on clause level may be likewise more likely to occur in the singular (cf., Carvalho, Orozco, & Shin [2015] on Spanish). Yet, Li, and Bayley (2018) found the converse preference in Mandarin Chinese where nonsingular subjects (second- and third-person ones) were more

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

likely to be zero. More corpus studies from a wider range of languages are required to establish universal tendencies and typological differences in this regard.

A further line of thinking focusing on structural environment emphasizes the role of frequency of use: a specific form tends to occur more and more frequently together with another adjacent form, so that their co-occurrence becomes predictable and regularized. Once such adjacent co-occurrence has reached a critical minimum, language users tend to process these as a single structural unit (Barth & Kapatsinski, 2017; Bybee, 2006; Bybee & Scheibman, 1999; Krug, 1998; *inter alia*), and, hence, this phenomenon often involves processes of formal reduction on one or both co-occurring elements. Such processing-related factors do not pertain exclusively to referential choice but resemble general principles of language use. A key aspect of these considerations is that they do not assume any functional motivation; instead, preferences for specific forms are fairly idiosyncratic. In Vera'a, it is the position of subjects and TAMP morphology, and the formal status of the latter that is of particular relevance.

A similar factor, but not entirely excluding functional considerations, is the co-presence of subject-predicate agreement. This is relevant in the sense of what is known as Taraldsen's Generalization (Taraldsen, 1978; see Seo [2001:Chapter 2] or Simonenko & Crabbé [2019] for discussion) or Nichols' (2018) *complementarity hypothesis*, whereby pronouns are more likely where agreement is absent or not sufficiently transparent, thus ensuring marking of relevant features at least once but never more than once (Meyerhoff [2009:309] on Tamambo versus Bislama) (cf., Table 2). This latter aspect of morphological structure is discussed under the heading of *Morphological Uniformity* after Jaegli and Safir (1989) and has been subject to corpus-linguistic work on Germanic, Slavic, and Romance languages. For instance, Fuß (2005) and Rosenkvist (2018) found that pronouns are significantly more frequent in subjects where verb forms are less transparent (and hence less informative) due to syncretism in a number of Romance and Germanic varieties, respectively (cf., also Simonenko & Crabbé [2019] for its relevance in the history of subject expression in French). Similarly, Seo (2001:165) found a higher frequency of subject pronouns with preterit tense than with present tense predicates in Russian, where only the latter show subject agreement in person and number; her data also show significant differences across parallel corpora from Russian and four other Slavic languages in this regard, so that the rate of zero subjects is higher in the latter languages that also show verbal subject agreement in person and number in all tenses. Nichols (2018), on the other hand, did not find complementarity confirmed in her corpus from the Northeast Caucasian language Ingush.

There are pragmatic and psycholinguistic motivations for the complementarity hypothesis. Going all the way back to Grice and his Maxim of Manner, listeners should prefer for speakers to give them clear and concise information, providing neither too much nor too little information (Grice, 1975). Psycholinguistic experiments have shown reaction times to favor additional cues in ambiguous environments but disfavor redundancy in already clear environments (Bates &

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

MacWhinney, 1989; Caballero & Kapatsinski, 2015; Kail, 1989). We consider the complementary hypothesis as a possible motivation for the co-presence of subject-predicate agreement in ambiguous environments.

What is relevant for our concerns in regards to co-present subject-predicate agreement here—and generally in the spirit of Torres Cacoullos and Travis (2019)—is the shift away from broad correlations yielding holistic typologies of languages (Gilligan, 1987; Roberts & Holmberg, 2010) and toward considerations of language-internal structural variability in co-determining the expression of subjects as pronoun or zero. In Vera'a, it is prospective aspect marking that also constitutes subject agreement in approximately half of all finite verbal clauses. The lack of agreement in all other TAMP categories then gives rise to the possibility of complementarity: zero subjects would be more frequent in prospective aspect clauses than with other TAMP values. This factor overlaps with that of temporal sequencing, since prospective aspect is typically used to refer to the new event that ensues at a given point in discourse, being thus the primeval aspect for temporal sequencing clauses.

CORPUS DATA AND CODING

Corpus

We investigate a corpus of ten narrative texts from eight speakers of Vera'a, comprising 20,851 words. The texts were recorded by the first author during his first one-year fieldtrip in 2007. All narrators volunteered proactively to have their stories recorded. As far as such distinctions can be meaningfully made, we can state that most stories belong to the oral literary tradition of Vanua Lava (where Vera'a is spoken) rather than being adaptations of Western stories, although considerable influences from the latter tradition can be assumed. Mythical stories have human heroes who get in conflict with evil supernatural villains; these stories often explain social or environmental facts and may contain a moral. Fables have human-like animals as heroes, and are often of a more entertaining character, while also explaining some facts about nature—for instance why cats chase mice.

Variables coded for in our analysis

The corpus contains 3,037 clauses that have a subject, and, of these, 1,404 were selected for analysis in a mixed-effects regression model. The tokens in the analysis were restricted to third-person subjects that were expressed as a pronoun or zero, had discourse-given referents (and hence an antecedent in the preceding discourse) and were not in relative clauses. We excluded certain types of nonfinite clauses for the obvious reason that these do not contain a subject relation. The dependent variable was the form of the subject, with pronoun and zero as the two possible values.

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

The following independent variables were considered in our statistical model:[5]

- *Animacy*: The effect of animacy may be not very dramatic in subject expression for this corpus since the majority have human reference. We only distinguish here between human and nonhuman referents, neglecting further distinctions within the latter category. Where generally nonhuman referents are anthropomorphized to the extent that they are treated entirely like human beings, for instance, spirits in myths or animals in fables that can speak and are capable of planning and thought, these are coded as a separate category but lumped with humans in the current analysis. The prediction is that human and human-like referents favor expression by pronouns.
- *Number:* We only consider the contrast between singular and nonsingular numbers, the latter encompassing plural, dual, and trial. Initial testing showed no difference between nonsingular numbers, and some numbers are too infrequent (our dataset contains only forty-four tokens of trial subjects) for model convergence. The prediction is that nonsingular numbers favor expression by pronouns.
- *Anaphoric distance:* We measure anaphoric distance in clause units. A distance of one clause means that the antecedent is in the immediately preceding clause, and so forth. Distances of two or higher are treated as a single category, as initial analyses showed no difference in subject expression for anaphora beyond two clauses. Since clause combining is generally paratactic in Vera'a, we do not distinguish between main and subordinate clauses. A distance of zero would comprise cases where the antecedent of a subject is a left-dislocated phrase in preclausal position. There are no other possibilities for a subject to have an antecedent within the same clause, given its initial position. However, there were no cases in our dataset where there was a left-dislocated phrase followed by a zero subject, so cases of zero-clause anaphora distance were not included in our analysis, as the conditioning was not variable. Left-dislocated phrases followed by pronominal subjects were excluded.
- *Antecedent function and temporal sequencing*: We consider antecedent function to be a potential factor in the sense of accessibility theory, but, for the purposes of the current study, we assume that it loses its impact when the antecedent is more than three clauses away, so all functions considered here apply only to those within three clauses distance. Nonsubject functions considered here are object and a range of functions that we collapsed into a single category "other," namely, nonverbal clause subjects, oblique, possessor, left-dislocation (in preclausal position of the same clause or any other preceding clause [Chinese-style topicalization]) and antecedents that are only partially co-referential. Nonsubject antecedence is predicted to favor pronominal form of subjects, because of the difference in function between the subject and antecedent.

We also consider the effect of temporal sequencing on the expression of subjects. Clauses that are part of a foregrounded sequence of events, one after another, are coded as sequenced. Clauses that are backgrounded, predicted, generally true, a description, a clarification, or merely express duration are coded as nonsequenced. In accordance with previous studies of subject expression, we explore the interaction of temporal sequencing with the type of antecedent.

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

Namely, we want to determine if temporal sequencing predicts more zero-subject expression in cases of clause-chaining of co-referent subjects. Table 3 below shows the distribution of subject form in these cases. It should be immediately apparent that pronominal reference is proportionally most frequent when the antecedent of a sequenced clause is not a subject (92%). Zero subjects are proportionally most frequent in the sequence of co-referent subjects (36%); however, zero-subject expression is not considerably different across different temporal and functional antecedent configurations. This is why it is important to include this factor alongside other possible predictive IVs.

TABLE 3. *Subject form by temporal sequencing and function of antecedent* (n = *1,404)*

Temporal sequence	Antecedent function	Subject: Pronoun	Subject: Zero	Total *n*
Sequenced	Subject	437 (64%)	241 (36%)	678
	Other	134 (92%)	12 (8%)	146
Non-sequenced	Subject	334 (78%)	96 (22%)	430
	Other	128 (85%)	22 (15%)	150

Note: Percentages of form expression for each factor level appear in parentheses.

- *Presence of connectives:* We code the presence or absence of a connective, as previous research has shown zero expression to be favored in cases of overt coordination (Torres Cacoullos & Travis, 2019:661). There are various Vera'a conjunctions and connective particles such as *wo* 'and, when' (cf., examples 5, 9, 25) and *ē* 'and then' (cf., 15). Connectives are relatively rare, but, when present, they occur more often with a pronominal subject.
- *TAMP expression:* We code for whether the TAMP marker following the subject has a clitic form (like *=m* or *=k*) versus a particle (like *me* or *ne*) and whether the TAMP marker includes information about person/number of the subject. While the form of TAMP marker does not determine the form of the subject, it may be possible that clitic TAMP markers are beginning to form a morphological unit with their preceding element, when that element is a subject pronoun. Hence, it may be less favorable to produce a clitic TAMP marker without a pronoun, and, thus, we would expect clauses with clitic TAMP markers to also show a propensity for the use of subject pronouns rather than zero (cf., Table 2 and related examples).

  The presence of subject agreement in the form of a prospective aspect marker, as opposed to other TAMP categories without agreement (cf., examples [10] and [11] versus [6] and [7]) may also have an effect. Since we are restricting our investigation to third persons, this distinction comes down to a singular (*ne*) versus nonsingular (*=k*) contrast in subject-predicate agreement. If speakers of Vera'a adhere to complementarity, we would expect the favoring of pronouns where number is not coded by prospective aspect and avoiding a pronoun in the co-presence of prospective aspect (cf., Table 2). Thus, the preference would be for person/number features to be realized precisely once in a clause. The form

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

(clitic or particle) of the TAMP and whether or not it expresses person interact, and, therefore, are bundled as one independent variable. The prospective TAMP marker *ne,* particle and singular, should favor zero subjects. The nonagreement TAMP clitic *=m* should favor a pronoun so that person is expressed and the clitic has a subject host. However, the remaining two combinations, of the prospective nonsingular clitic *=k* and the nonagreement particle TAMP markers like future *me* and iamitive *man*, could theoretically go either way, depending on what factor has a stronger impact on the data.

Table 4 shows the distribution of the data by both factors.

TABLE 4. *Subject form by TAMP form and person inflection (*n = *1,404)*

TAMP expression	TAMP agreement	TAMP form	Subject: Pronoun	Subject: Zero	Total *n*
me, man, etc	No agreement	Particle	159 (90%)	17 (10%)	176
=m		Clitic	500 (87%)	73 (13%)	573
ne	Agreement (prospective)	Particle - 3s	192 (44%)	240 (56%)	432
=k		Clitic - 3p	275 (87%)	41 (13%)	316

RESULTS AND DISCUSSION

A mixed-effects generalized linear (GLMM) regression model was produced with a stepwise procedure, using the packages *lme4* (Bates, Mächler, Bolker, & Walker, 2015) and *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017) in *R* (R Core Team, 2018). A random intercept was included for speaker. Our results (Table 5) show that Animacy, Number, Anaphoric Distance, Temporal Sequencing, Connectives, and TAMP expression all have a significant role to play in predicting subject expression.

Overall, we find a strong preference in Vera'a narratives to express subjects by pronouns rather than zero anaphors. Of the 1,404 third-person subjects investigated here, only 357 (26%) are zero. Differences across individual speakers are marginal, suggesting homogenous behavior of narrators with regards to subject realization. We discuss individual factors in turn.

Animacy: When subjects are humans, they are more likely to be expressed with a pronoun than when they are nonhuman animates ($\beta = 0.58, p < 0.05$). This contrast is reflected in the first two clauses of (12) compared to the final two clauses in (12):

(12) ***di***^{i} ***ne*** *mi'ir suwō* ***di***^{i} ***ne*** *qōrqōr*
3SG PROSP:3SG sleep down 3SG PROSP:3SG dream

[*n me'*]^{k} *ne* *van ma* *ϕ*^{k} *ne* *tēk mē diē so*
ART reef PROSP:3SG go hither (3sg) PROSP:3SG speak DAT 3SG QUOT
'He went to sleep and had a dream: the reef came and spoke to him.' ISAM.051

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

TABLE 5. *Results of mixed-effects generalized linear regression: factors predicting subject pronoun expression* (n = *1,404*)
Log likelihood: −557.9 AIC: 1141.8 BIC: 1210
Speaker variance: 0.36 ± 0.60

	n (%)	Estimate	*SE*	*Z*	*p*
Intercept		−0.68	0.54	−1.26	0.21
Animacy = deity (reference level)	230 (53%)				
Animacy = human	**1111 (80%)**	**0.58**	**0.23**	**2.57**	**0.01**
Animacy = inanimate	**63 (32%)**	**−1.30**	**0.38**	**−3.45**	**0.00**
Number = singular (reference level)	807 (61%)				
Number = non-singular	**597 (90%)**	**1.36**	**0.29**	**4.66**	**0.00**
Anaphoric distance = 1 clause (reference level)	1127 (68%)				
Anaphoric distance = 2+ clauses	**277 (95%)**	**1.73**	**0.31**	**5.58**	**0.00**
Temporal sequencing = nonsubject antecedent, nonsequential (reference level)	150 (85%)				
Temporal sequencing = subject antecedent, sequential	678 (64%)	0.26	0.33	0.79	0.43
Temporal sequencing = subject antecedent, nonsequential	430 (78%)	−0.06	0.33	−0.19	0.85
Temporal sequencing = nonsubject antecedent, sequential	**146 (92%)**	**2.13**	**0.49**	**4.33**	**0.00**
Connective = absent (reference level)	1092 (70%)				
Connective = present	**312 (88%)**	**1.02**	**0.22**	**4.61**	**0.00**
TAMP expression = number agreement, clitic (reference level)	307 (87%)				
TAMP expression = ambiguous number, particle	**153 (90%)**	**1.62**	**0.43**	**3.77**	**0.00**
TAMP expression = ambiguous number, clitic	**516 (87%)**	**0.99**	**0.32**	**3.11**	**0.00**
TAMP expression = number agreement, particle	**428 (44%)**	**−0.77**	**0.34**	**−2.25**	**0.02**

Note: Token number is for observations per level. Percentage is for pronoun expression within each level. Positive coefficients are associated with higher pronoun expression, significant effects in bold.

When subjects are inanimates, they are more likely to be zero (β = -1.30, $p < 0.01$); an illustrative example is (13) where the referent of the subject in the final clause is the 'wood' mentioned in the preceding clause.

(13) *wo di ne diñ [ēn woqe'enge]*^{k} *ϕ*^{k} *ne bur*
and 3SG PROSP:3SG flick ART wood (3sg) PROSP:3SG shut
'And then he flicked the wood shut.' JJQ.386

This is one of the few conditions to significantly favor zero subject expression. We take this to reflect a restriction of pronouns to human beings, as suggested by Genetti and Crain (2003) for Nepali (see also, Haig, Schnell, & Wegener, 2011): the rationale here is that pronouns as a form class, which include first- and second-person forms referring to speaker and addressee, are so strongly associated with humanness that they are avoided for nonhumans. This does not preclude zero form for human reference but favors zero for nonhuman ones.

Number: Nonsingular subjects are significantly more likely to be expressed with a pronoun (β = 1.36, $p < 0.01$) than singular ones. Examples (14) and (15) illustrate

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

this: in (14), the singular prospective marker *ne* is used, and zero subjects occur in all noninitial clauses in this sequence. By contrast, in (15) the nonsingular form of the prospective marker *=k* attaches to a preceding plural pronoun *dir* in both clauses.

(14) ***buskat***^{i} ***di***^{i} *=m* *dar* *kēlkēl* *dar* *kēlkēl* *vanvan*
Cat 3SG =PRF wait all.about wait all.aboutout RED:go

ø ^{i} *ne* *rem*
(3SG) PROSP:3SG climb

ø^{i} *ne* *rem* *kal* *sag* *diñ* *ma* *=n* *wova'al* […]
(3SG) PROSP:3SG climb upwards up reach hither =ART pawpaw

ø^{i} *ne* *'ēn* *suw* *lē* *=n* *lōlō-gi*
(3SG) PROSP:3SG see down LOC =ART inner-3SG

'Buskat, he waited and waited until (he) then started to climb, climbed up to the top of the pawpaw (that was ripe) and looked down into it.' GABG.109-110

(15) ***dir =ēk*** *mul* *kal* ***dir =ēk*** *bigbig* *ēn* *go-rē*
3PL =PROSP:NSG go back 3PL =PROSP:NSG RED:eat ART POSS.EAT-3PL

=n *vovoñodo*
=ART RED-catch

'They went back up (the shore) and they ate their catch (with staples).'
ISAM.005

Although our model suggests an independent effect of number, example (14) illustrates a typical case of a singular zero subject co-occurring with the TAMP marker, which is both a particle and an agreement marker. The independent effect of number can be regarded as a markedness effect (cf., Seo, 2001), whereby the default expected case is, in particular for humans, to make an appearance as single individuals, so that reference to multitudes thereof triggers overt expression; see Schnell (2019) on nominal number in Vera'a.

Anaphoric distance: Subjects whose antecedent is two or more clauses away are significantly more likely to be expressed with a pronoun than subjects whose antecedent is only one clause prior ($\beta = 1.73$, $p < 0.01$). Hence, similar to English, zero subjects in Vera'a are unlikely where the distance is larger than one clause unit. A rare example of a zero subject with high anaphoric distance is given in (16), where the spirit (indexed i) is taken up again by zero after two intervening clauses. Example (17), on the other hand, represents the more typical constellation where the same referent is the subject in a sequence of clauses.

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

(16) *ϕ*^{i} *ne* *'añ* *qal* *sa* *=n* *'uwañēre*^{k}
(3SG) PROSP:3SG handle hit EMPH =ART first.born

ϕ^{k} *ga* *mi'ir* *qēlēñ* *wal* *di*^{k} *e* *rōñrōñ* *wal* *rōs*
(3SG) HAB sleep lost INT 3SG NEG1 RED:feel INT NEG2

alē *ϕ*^{i} *ne* *gis* *sa* *=n* *wōbin* *mu-gi*
INTERJ **(3SG)** PROSP:3SG grab EMPH =ART penis POSS.GEN-3SG

ϕ^{i} *ne* *rēv*
(3SG) PROSP:3SG pull

'(He (the spirit) came back in,) touched the older brother (who) was fast asleep and he didn't notice a thing. Alright, (he, the spirit) grabbed his penis and pulled on (it).' 1.AS.049-051

(17) ***gusuwō*** *ne* *kal* *raka*
rat PROSP:3SG rise out

ϕ^{i} *ne* *rōw* *lē* *=n* *nanara* *lumasag* *[...]*
(3sg) PROSP:3SG jump LOC =ART tree.sp on.top

ϕ^{i} *ne* *rōw* *kal* *lumasag* *kēkē* *ϕ*^{i} *ne* *sag* *'i*
(3sg) PROSP:3SG jump up on.top high **(3sg)** PROSP:3SG sit DEL

'Rat got up, jumped on top of the *nanara* (trunk), jumped up on top and sat down there.' GAQG.028

Crucially, the effect picked up by our model is for anaphoric distance higher than one clause to prefer expression by pronoun. In the context of one-clause anaphoric distance, pronouns are actually still the more frequent form of subject expression, as exemplified by (18).

(18) ***di*** *ne* *sagsag* *'i* ***di*** *ne* *rērē* *qēl* *suwē*
3SG PROSP:3SG RED:sit DEL **3SG** PROSP:3SG crane.neck descend down

di *ne* *'ēn* *suwē* *=n* *gusōwō* *ϕ*^{i} *ne* *kalu* *[...]*
3SG PROSP:3SG see down =ART rat (3sg) PROSP:3SG exit

'He sat there for a bit, then looked down and saw that down there the rat was coming out (of the hole in the nanara trunk.)' GAQG.008

This yields something like a *privative opposition*—in classic structuralist terms—triggering the use of pronouns at distances greater than one and leaving the choice more open at a distance of one, while still favoring pronouns. While this constellation is obviously a reflection of accessibility in a wide sense, it does contradict two major assumptions of the more specific framework of

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

Accessibility Theory (AT), as proposed by Ariel (2014 [1990]). First of all, distance is not a scalar factor in Vera'a; the choice of pronoun versus zero does not increase in parallel to an increase in distance and an assumed concomitant decrease in accessibility. Secondly, only the use of zero bears any functional load during comprehension, pointing the addressee to retrieve its antecedent in the immediately preceding clause; the use of a pronoun, however, does not provide any relevant clues. This contradicts the predictions of AT where any contrast in referential form corresponds to a contrast in referential function in a one-to-one mapping relation.

Temporal sequencing: We find that tokens with subjects for antecedents show no significant effect on subject expression. This holds whether or not the clauses are sequential ($\beta = 0.26$, $p = 0.43$) or not ($\beta = -0.06$, $p = 0.85$). Previous studies, such as Myhill (1997) found a greater likelihood of same-subject anaphors in temporally sequenced clauses to be zero. In temporally sequenced same-subject clause chains in our corpus, the favored form of subject expression is still a pronoun, as shown in (18), and there is no significant difference to same-subjects in nonsequential clause chains as in (19). However, it is important to note here that subjects with subject antecedents do not significantly favor zero, and they do not significantly favor pronouns either and are one of the few conditions in our model that do not lean heavily toward pronoun-subject expression.

(19) ***duru*** *=m 'ōg van a Lēmērig*
3DU =PRF stay go LOC.SP place.name

duru *ga mōrōs nō-ru =n raw*
3DU HAB want POSS.DOM-3DU =ART intrasex.pig
'They (two) lived and lived in Lēmērig, and they wanted a bisexual pig.'
1.AS.004-005

We do see an effect for nonsubject antecedents in a sequential clause chain to favor pronoun expression ($\beta = 2.13$, $p < 0.01$). In this context, the subject is almost categorically a pronoun. We see this in (20), where the function of the antecedent of the third-person dual subject *duru* is a possessor, as indicated by the suffix *-ru* on the possessive classifier *go-* (indexed *k*).

(20) [*reñe anē*]^{i} *ne mōn go-**ru**^{k} =n dom*
woman DEM.ADDR PROSP:3SG roast POSS.EAT-**3DU** =ART yam

*ē **duru**^{k} =k van 'ō*
DISC **3DU** =PROSP:NSG go with

'So, the woman roasted some yams for the two to eat, and then they went, (taking them) with (them).' ANV.050-051

In sum, this means that zero subjects are largely restricted to contexts of temporal-sequencing (cf., Table 3). But even within this context, the pronominal form is still

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

favored. This is parallel to our finding regarding antecedent distance where the zero form is largely restricted to one clause distance, but most subjects with antecedent subjects in the preceding clause are also pronouns rather than zero. The most obvious explanation is that pronouns are just so predominant as a referential choice that such functional distinctions are not picked up through this choice. At most we can identify same-subject chains as a context that allows zero expression more often than other constellations.

Clause connectives: Clauses with a connective are significantly more likely to have subjects expressed as pronouns ($\beta = 1.02$, $p < 0.01$). This is in sharp contrast to findings from other studies, for example, see Torres Cacoullos and Travis (2019:659–65) for discussion and references. While in cases like (20), the connective is used with a shifted subject, connectives are also found in same-subject chains, as in (21).

(21) ***dir*** *=ēk* *van* *diñ* *ma* *=n* *'uvu* *baka* *anē*
3PL=PROSP:NSG go reach hither =ART base nambangga DEM.ADDR

ø *=k* *'ēn* *sag* *ēn ..* *ēn* *tōo*
(3PL) =PROSP:NSG see upwards ART ART fowl

dir *=ēk* *'ēn'ēn* *suwō* *ø* *'ēn'ēn* *qal* *rōs* *duruō*
3PL =PROSP:NSG RED:see down (3pl) RED:see hit NEG.GEN2 3DU

ē ***dir*** *=ēk* *sikē*
DISC **3PL** =PROSP:NSG search

'[They were coming into the forest, (they) heard the two chickens sing out,] (they) came to a nambangga tree, looked up and spotted the chickens, looked down, did not find them (the two girls), and thus went on searching.' ANV.083-085

One possible explanation for the frequent use of pronouns after connectives is that connectives in Vera'a always bear discourse-structuring functions rather than coordinating and sequencing meanings: for instance, *ē* has the effect of concluding a quintessence of what was said before and turn to the following events. It is in this sense not immediately comparable to many of the clause connectives in English, in particular not *and*. Specifically, temporally sequenced clause chains in Vera'a are always asyndetic, including final segments, which is where connective *and* in English is extremely common.

TAMP expression: TAMP morphology, including rudimentary subject agreement, comes out as significant in our model, with both of our theoretical predictions borne out. Overall clitics are favored to have a pronominal host and number expression is favored to occur once per clause. Pronouns are more likely when the TAMP marker is a clitic, and there is no subject agreement ($\beta = 0.99$, $p < 0.01$). Hence, examples like (22) where the TAMP marker has a pronominal host and subject number is expressed only once are favored over those in (7), repeated here.

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

(22) *da bē* ***di*** ***=m*** *kalu ma* ***di*** ***=m*** *'ō =n bē nanara*
make water **3SG** **=PRF** exit hither **3SG** **=PRF** carry =ART trunk tree.sp
'Because of that the water came out and (it) took the *nanara* trunk.' GAQG.024

(7') ***'ama-ru*** ***=m*** *van ø* ***=ēm*** *wana ø* ***=m*** *wana*
father-3DU **=PRF** go (**3sg**) **=PRF** squeeze.out (3sg) **=PRF** squeeze.out

ø ***=m*** *'ōg lē =n gamal van vavan*
(**3sg**) **=PRF** stay LOC =ART men's.house go RED-go

'Their (two) fathers had gone and been squeezing and squeezing (it; kava) out, had been sitting in the men's house (where kava is drunk) on and on (until he came home).' MVBW.119

When the TAMP marker is a particle and agrees in number (prospective third-singular *ne*), subjects are likely to be zero ($\beta = -0.77$, $p < 0.05$), meaning subject number is expressed only once, as in (23).

(23) ***di***^{i} ***ne*** *van qēl ma*
3SG **PROSP**:**3SG** go descend hither

di^{i} ***ne*** *le =n mēdēr ne vōwal*
3SG **PROSP**:**3SG** take =ART rock NUM.ART one

ø^{i} ***ne*** *bur gōr* ***ø***^{i} ***ne*** *van qēl suwō* […]
(**3SG**) **PROSP**:**3SG** shut secure (**3SG**) **PROSP**:**3SG** go descend down

'She went down, she took a rock, blocked off (the opening of the cave) and went down …' HHAK.045-046

Additionally, where the TAMP marker is a particle but does not co-express agreement, pronominal subjects are strongly favored ($\beta = 1.62$, $p < 0.01$). In these cases, number is expressed once in the clause, through the subject pronoun, as seen in example (24).

(24) *ba* ***di*** ***me*** *kalu* ***di*** ***me*** *van 'ag ēn dōdōmia*
but **3SG** **FUT** exit **3SG** **FUT** go follow ART thought
'But he will come out and (he will) go according to the idea that …' ISWM.302

This means that person and number values for subjects tend to be expressed only once per clause, either by a pronoun, as in (22), and also (24), or as a TAMP marker, as in (23), and also (14) and (17). Conflicting tendencies arise where a subject shows number agreement and is a clitic (nonsingular prospective *=k*). A huge majority of these subjects are realized as a pronoun, as in (25) where the clitic attaches to the preceding dual pronoun *duru*. The clitic form preference to have a pronominal host overrides the avoidance of doubling the expression of person/number features.

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

(25) ***duru*** ***=k*** *kalraka* ***duru*** ***=k*** *tēk* *mē* *di* *so*
3DU **=PROSP:NSG** get.up **3DU** **=PROSP:NSG** speak DAT 3SG QUOT
'Then they (two) got up and said to him:' HHAK.117

These findings do lend some support to the hypothesis that subject pronouns and clitic TAMP markers form a structural unit to some extent. But it leaves open the question as to why pronouns should be even more common where the TAMP marker is not a clitic, as in (24). One possibility is that examples like (7) are restricted to a small range of lexical host words other than pronouns, as was suggested by Schnell (2018) for first- and second-person zero subjects. Finally, it is to be noted that, even where agreement is co-expressed by a prospective particle *ne*, a pronoun is still used in 44% of these cases (see Table 4 and example [18]). Hence, Vera'a does show a considerable degree of redundancy of person/number expression in subjects but seems to restrict zero subjects largely to contexts of co-present agreement.

There are very few cases such as (26)[6] with a zero subject and a particle that does not express number agreement, leading to no person/number values for the subject expressed in the clause.

(26) *gav* *diñ* *wal* *rōw* *ēn* *naw* *ē*
fly reach once down ART saltwater DEM.REM

ø ***mak*** *gav* *kal* *'a-'ag* *ma* *=n* *qoro-gi* *anē*
(3sg) IMM fly up RED-follow hither =ART hole-3SG DEM.ADDR

'As soon as (they) reached the sea, (they) flew up here along the river.'
ANV.086

In sum, while all six factors that our model has picked up as significant make their own respective contribution to the variation at hand, we can identify two areas where the zero form appears to have a stronghold against the strong general trend toward realizing subjects as pronouns: same-subject chains, and singular subjects in prospective-aspect clauses, marked by a TAMP particle. By far, most of these examples are also temporal sequence clauses. The former context seems to reflect a near-categorical rule of clause combining in the language, similar to English. It is nonetheless worth noting again that, even in this type of context, the majority of subjects are realized as a pronoun. This is a convergence of various dimensions that allow for zero subjects: the TAMP marker in question is also an agreement marker, it is a particle rather than a clitic, it is the form for singular rather than nonsingular subjects, and it marks prospective aspect, a TAMP category strongly associated with temporal sequencing. All this taken together suggests an overall picture where the default expression for nonlexical subjects in Vera'a is a pronoun rather than a zero, except for those cases where it is third-person singular in a prospective clause, which also shows subject-predicate agreement. That the respective TAMP exponent also bears nonclitic

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

form is probably not a coincidence: it would, for instance, seem possible that it resists overall tendencies of morphological reduction during language change due to the commonality of co-occurring zero subjects. In sum, these multiple factors seem to conspire in allowing for zero subjects in this context.

CONCLUSIONS AND OUTLOOK

In conclusion, pronouns are the default subject expression in those contexts where no full (lexical) noun phrase is used in Vera'a narrative discourse. The functional factors our model identifies as significant for Vera'a subjects have also been found relevant in previous studies of the equivalent alternation in other languages, yet their specific bearings on the alternation in Vera'a warrants some discussion. With regards to our three goals, namely, to determine the range of factors determining the choice between pronouns and zero, the role of agreement therein and a tentative account of the overall pronoun preference, we can state the following: the variation attested in the Vera'a corpus is subject to factors similar to those found in previous studies of subject realization in the variationist tradition, with accessibility, clause chaining, and animacy and number all contributing to the variation at hand. In particular, the tendency for zero anaphor to be most prevalent in same-subject chains in temporally sequenced clauses squares with findings from English, Spanish (Torres Cacoullos & Travis, 2019:673), Mandarin (Li & Bayley, 2018), and Tamabo and Bislama (Meyerhoff, 2009:309). Hence, our study adds further support that this is universally a context more likely to show zero anaphors in subject function, regardless of their overall potentially divergent preferences (see Torres Cacoullos & Travis [2019:671] for remarks along these lines). A specific finding from the Vera'a corpus is, however, that this convergence context does not yield a preference for zero anaphors but rather defines a context where the otherwise overwhelming preference for pronouns is attenuated, so that zero anaphor has a greater share here compared to other contexts of greater anaphoric distance, discontinuing syntactic function and so forth. This can be understood as further contribution to Torres Cacoullos and Travis' (2019) general idea that functional factors relevant for referential choice may be quite similar across languages, but the specific response to these factors can differ drastically, yielding the—in this sense somewhat misleading—impression of stark contrasts in this regard, as reflected in the null-subject versus nonnull-subject typology. In regards to our second research question, we conclude that the rudimentary form of subject agreement plays a major role in determining the form of anaphoric subjects in Vera'a, thus lending further support to Taraldsen's Generalization (Seo, 2001; Simonenko & Crabbé, 2019; Taraldsen, 1978) and Nichols' (2018) *complementarity hypothesis*: regardless of some redundancy in the combination of pronouns with agreement in the predicate, subject features tend to be expressed only once in a clause, and tend to be zero only where respective features are marked transparently within verbal predicates, as found by, for

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

instance, Rosenkvist (2018) or Fuβ (2005) for some Germanic and Romance dialects of Europe. While findings from these previous studies relate primarily to syncretisms in verbal paradigms and differences across tense-specific subparadigms, we find the rudimentary form of subject agreement as restricted to only one aspectual category to be relevant. With regards to our third question, the restrictedness of subject agreement to a subset of clauses as well as the tendency for TAMP clitics to form a structural unit with pronouns, account to a large extent for the overall strong preference for pronouns. Taking these latter two conclusions into account, we can further elaborate on our conclusions regarding the first question: if we think of Vera'a as responding to the same set of functional factors as other languages do, we would consider the absence of agreement and clitic form of TAMP morphology as competing factors that lead to an overall preference for pronouns, including in the convergence context of temporally sequenced same-subject chains. Hence, zeroes remain a minority form here, but their relatively higher proportion aligns with the preference for zero form in this functional context across languages.

Overall, we find that the strong preference for pronouns in Vera'a seems to follow from independent factors on other levels of morphosyntactic representation. Major drivers are complementarity of subject feature expression as well as more general tendencies of structural reduction and conjoint processing in TAMP morphology.

NOTES

1. Grapheme to phoneme correspondences: <g> /ɤ/; <q> /kp/; <ñ> /ŋ/; <ẽ> /ɪ/; <m̃> /ŋm/; <õ> /ʊ/; other grapheme correspondences are as predictable.

2. The choice between lexical (full NPs) and nonlexical expressions seems to follow quite different considerations and has, in our view, been treated much more satisfactorily than the choice between pronoun and zero in a number of seminal contributions (e.g., Ariel, 2014; Chafe, 1976; Du Bois, 2003; Givón, 1983, 2015).

3. Prospective aspect is—roughly—a more grammaticalized equivalent of English *going to*, which can have reference times in past and present time.

4. Vera'a is, in this regard, like Kwa'kwa'ala, as discussed by Anderson (1992).

5. We initially considered a distinction between S (subject of intransitive predicates) and A (subject of transitive predicate) functions as well, following some of the typological literature on reference, but collapsed the two into a single category of subject, since initial investigations showed that the proportion of pronoun and zero is identical and that it does not bear on the choice at hand. Hence, we did not include this distinction in our model.

6. Note that the first clause in this example has been excluded from analysis, because it represents a nonfinite tail-head linkage.

ABBREVIATIONS

1	1st person
2	2nd person
3	3rd person
A	a-form of demonstrative
ABIL	ability
ABL	ablative
ADDR	addressee-oriented

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

ART	common article
AT	Accessibility Theory
CONJ	conjunction
DAT	dative
DEL	delimitative (aktionsart)
DEM	demonstrative
DISC	discourse particle
DISTR	distributive
DU	dual
DOM	domestic
EAT	eat possession
EMPH	emphatic particle
EX	exclusive
FUT	future
GEN	general (possession)
HAB	habitual
HOUSE	house possession
IAM	iamitive
IMM	immediacy (TAMP)
IN	inclusive
INTERJ	interjection
IPFV	imperfective
LOC	locative
MAN	manner demonstrative
NP	noun phrase
NSG	non-singular
NUM	numeral (prefix or article)
PERS	personal article
PL	plural
POSS	possessive classifier
PRF	perfect
PROH	prohibitive
PROSP	prospective
QUOT	quotative
RED	reduplication
REL	relativizer
SAP	speech-act participant
SG	singular
SP	specific
TAMP	tense-aspect-mood-polarity
VAL	valuable possession
VC	verb complex

REFERENCES

Anderson, Stephen. (1992). *A-morphous morphology*. Cambridge: Cambridge University Press.
Ariel, Mira. (2014 [1990]). *Accessing noun-phrase antecedents*. London and New York: Routledge.
Barbosa, Pilar, Duarte, Maria Eugenia Lamoglia, & Aizawa Kato, Mary. (2005). Null subjects in European and Brazilian Portuguese. *Journal of Portuguese Linguistics* 4, 11–52.

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

Barth, Danielle, & Kapatsinski, Vsevolod. (2017). A multimodal inference approach to categorical variant choice: construction, priming and frequency effects on the choice between full and contracted forms of *am, are* and *is*. *Corpus Linguistics and Linguistic Theory* 13(2), 1–58.
Bates, Douglas, Mächler, Martin, Bolker, Ben, & Walker, Steve. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48. <doi:10.18637/jss.v067.i01>.
Bates, Elizabeth, & MacWhinney, Brian. (1989). Functionalism and the Competition Model. In B. MacyWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing*. Cambridge: Cambridge University Press. 3–73.
Bickel, Balthasar, & Nichols, Joanna. (2007). Inflectional morphology. In T. Shopen (Ed.), *Language typology and syntactic description. Volume 3: Grammatical categories and the lexicon*. Cambridge: Cambridge University Press. 169–240.
Bickel, Balthasar. (2003). Referential density in discourse and syntactic typology. *Language* 79, 708–36.
Bybee, Joan. (2006). From usage to grammar. The mind's response to repetition. *Language* 82(4). 711–33.
_____, & Scheibmann, Joanne. (1999). The effect of usage on degrees of constituency: the reduction of don't in English. *Linguistics* 37(4), 575–96.
Caballero, Gabriela, & Kapatsinski, Vsevolod. (2015). Perceptual functionality of morphological redundancy in Choguita Rarámuri (Tarahumara). *Language, Cognition and Neuroscience* 30(9), 1134–43.
Cameron, Richard, & Flores-Ferrán, Nydia. (2004). Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context* 1.1, 41–65.
Carvalho, Ana M., Orozco, Rafael, & Shin, Naomi. (2015). Introduction. In A. M. Carvalho, R. Orozco & N. Shin (Eds.), *Subject pronoun expression in Spanish. A cross-dialectal perspective*. Georgetown: Georgetown University Press. xiii-xxxi.
Chafe, Wallace. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subjects and topics*. New York: Academic Press. 25–56.
Dryer, Matthew S. (2013). Expression of Pronominal Subjects. In: M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/101, Accessed on 2019-10-07.)
Du Bois, John. (1987). Absolutive zero. Paradigm adaptivity in Sacapultec Maya. *Lingua* 71, 203–22.
_____. (2003). Discourse and grammar. In Michael Tomasello (Ed.), *The new psychology of language. Volume 2*, 47–87. Mahwah (NJ): Lawrence Erlbaum.
Fuß, Eric. (2005). *The rise of agreement. A formal approach to the syntax and grammaticalization of verbal inflection*. Amsterdam & Philadelphia: John Benjamins.
Genetti, Carol, & Crain, Laura D. (2003). Beyond preferred argument structure: Sentences, pronouns and given referents in Nepali. In J. W. Du Bois, L. E. Kumpf & W. J. Ashby (Eds.), *Preferred argument structure: Grammar as architecture for function*. Amsterdam: Benjamins. 197–223.
Gilligan, Gary Martin. (1987). *A cross-linguistic approach to the pro-drop parameter*. Doctoral dissertation, University of Southern California.
Givón, Talmy. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam: John Benjamins. 1–41.
_____. (2015). Topic, pronoun, and grammatical agreement. In Talmy Givón (Ed.), *The diachrony of grammar. Vol I*. Amsterdam & Philadelphia: John Benjamins. 163–96.
Grice, Herbert P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech acts*. New York: Academic Press. 41–58.
Haig, Geoffrey. (2018). The grammaticalization of object pronouns: Why differential object indexing is an attractor state, *Linguistics* 56(4): 781–818.
Haig, Geoffrey, Schnell, Stefan, & Wegener, Claudia. (2011). Comparing corpora from endangered language projects: explorations in language typology based on original texts. In G. Haig, N. Nau, S. Schnell & C. Wegener (Eds.), *Documenting endangered languages. Achievements and perspectives*. Berlin: Mouton de Gruyter. 55–86.
Himmelmann, Nikolaus P. (2014). Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. *Language* 90(4), 927–60.
Hyslop, Catriona. (2001). *The Lolovoli dialect of the North-East Ambae language, Vanuatu*. Canberra: Pacific Linguistics.
Jaegli, Osvaldo, & Safir, Ken. (eds.) (1989). *The null subject parameter*. Dordrecht: Kluwer.
Jauncey, Dorothy G. (2011). *Tamambo, the language of west Malo, Vanuatu*. Canberra: Pacific Linguistics.

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press

Kail, Michele. (1989). Cue validity, cue cost, and processing types in sentences comprehension in French and Spanish. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing*. Cambridge: Cambridge University Press. 77–117.
Krug, Manfred. (1998). String frequency: a cognitive motivating factor in coalescence, language processing, and linguistic change. *Journal of English linguistics* 26. 286–320.
Kuznetsova, Alexandra, Brockhoff, Per B., & Christensen, Rune H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. R package version 3.0-1. *Journal of Statistical Software* 82.13, 1–26.
Li, Xiaoshi, & Bayley, Robert. (2018). Lexical frequency and syntactic variation: subject pronoun use in Mandarin Chinese. *Asia-Pacific Language Variation* 4.2, 135–60.
Lynch, John, Ross, Malcolm, & Crowley, Terry. (2011). *The Oceanic languages*. London and New York: Routledge.
Meyerhoff, Miriam. (2000). The emergence of creole subject-verb agreement and the licensing of null subjects. *Language Variation and Change* 12. 203–30.
_____. (2009). Replication, transfer, and calquing: Using variation as a tool in the study of language contact. *Language Variation and Change* 21(3). 297–317.
Myhill, John. (1997). Viewpoint, sequencing, and pronoun usage in Javanese short stories. In G. R. Guy, F. Crawford, D. Schiffrin & J. Baugh (Eds.), *Toward a Social Science of Language: Papers in honor of William Labov. Volume 2: Social interaction and discourse structures. Papers in honor of William Labov*. Amsterdam and Philadelphia: John Benjamins. 237–58.
Nichols, Joanna. (2018). Agreement with overt and null arguments in Ingush. *Linguistics* 56(4). 845–63.
R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Rizzi, Luigi. (1982). *Issues in Italian syntax*. Dordrecht: Foris.
Roberts, Ian, & Holmberg, Anders. 2010. Introduction: Parameters in minimalist theory. In Th. Biberauer, A. Holmberg, I. Roberts & M. Sheehan (Eds.), *Parametric variation: Null subjects in minimalist theory*. Cambridge: Cambridge University Press. 1–57.
Rosenkvist, Henrik. (2018). Null subjects and distinct agreement in Modern Germanic. In F. Cognola & J. Cassalicchio (Eds.), *Null subjects in Generative Grammar. A synchronic and diachronic perspective*. Oxford: Oxford University Press. 285–306.
Ross, John. (1982). Pronoun deleting process in German. *Annual meeting of the Linguistic Society of America.* San Diego, California.
Schnell, Stefan. (2018). When subject-verb agreement? Investigating the role of topicality, accessibility, in frequency in Vera'a texts. *Linguistics* 56.4, 735–80.
_____. (2019). Variable number marking in Vera'a. Animacy and beyond. *Asia-Pacific Language Variation* 5.2, 208–43.
_____, & Barth, Danielle. (2018). Discourse motivations for pronominal and zero objects. Across registers in Vera'a. *Language Variation and Change* 30, 51–81.
Seo, Seunghyun. (2001). *The frequency of null subject in Russian, Polish, Czech, Bulgarian, and Serbo-Croatian*. Doctoral dissertation, Ann Arbor.
Siewierska, Anna. (2004). *Person*. Oxford: Oxford University Press.
Simonenko, Alexandra, & Crabbé, Benoit. (2019). Agreement syncretisation and the loss of nulls subjects: quantificational models for Medieval French. *Language Variation and Change* 31.3, 275–301.
Taraldsen, Taralds. (1978). *On the NIC, vacuous application and the* that-*trace filter*. Indiana University Linguistics Club.
Thieberger, Nicholas. (2006). *A grammar of South Efate, an Oceanic language of Vanuatu*. Honolulu: University of Hawaii Press.
Torres Cacoullos, Rena & Catherine E. Travis. (2019). Variationist typology: shared probabilistic constraints across (non-)null subject languages. *Linguistics* 57.3, 653–92.
Travis, Catherine & Amy M. Lindstrom. (2016). Different registers, different grammars? Subject expression in English conversation and narrative. *Language Variation and Change* 28(1). 103–28.

https://doi.org/10.1017/S0954394520000125 Published online by Cambridge University Press