Discourse motivations for pronominal and zero objects across registers in Vera'a

STEFAN SCHNELL

The University of Melbourne and Centre of Excellence for the Dynamics of Language

DANIELLE BARTH

Australian National University and Centre of Excellence for the Dynamics of Language

ABSTRACT

The choice between pronominal and zero form for objects in the Oceanic language Vera'a is investigated quantitatively in texts from two registers with discourse topics of three different ontological class memberships. Discourse topicality is found to predict best the choice between pronoun and zero, outranking the factors of ontological class membership, antecedent form, and antecedent function. Contrary to current models of referent tracking, antecedent distance does not show any effect at all. It is concluded that (a) discourse structure and activation are not universally the most significant factors in referential choice and (b) ontological class and discourse topicality can be teased apart through appropriate text sampling, and it is the latter that is most significant. This bears important implications for the grammaticalization of object agreement and the typology of differential object marking.

Objects in Vera'a can take forms of the following three types: a full noun phrase (1a), a pronoun (1b), or zero (1c). Vera'a syntax is strictly configurational, and objects are marked solely by their position immediately following the verbal predicate. Like objects, subjects can also be left zero, as is the case in all examples here[1]:

(1) a. *[...] ø ne   'ēn e   **tētē**   ne   **vō-ruō** [...]*
            AOR:3SG see PERS.ART infant NUM.ART NUM-two

    '[She went up to throw out garbage] when (she) spotted **two babies** [crawling around].'   ANV.017

  b. *ø van mom   **dir** lē   =n   gamal*
         go put   3PL LOC   =ART mens'.house

    '[He cheered about **their** coming, saying: "Come here, come here! We will be staying together here for a while."] (He) took **them** to his house.'   JJQ.274

  c. *ø mom kal   sag **ø** lē   =n   wōwō   nev*
         put move.up up   LOC   =ART on.top   fire

    '[Then they went up, smashed the devil's head] and put (**it**) on top of a fire.'
                                                                    HHAK.138

Following some of the literature on referential choice (e.g., Du Bois, 1987), we classify noun phrases (NPs) as *lexical* and pronouns and zeroes as *nonlexical*. Despite some remaining uncertainties and open questions, we assume here that the choice between lexical and nonlexical expression of objects, like that of any other grammatical relation, is largely accountable for in terms of information packaging, namely referent activation (or "*accessibility*" in Ariel's work) (Ariel, 1990, 2008; Chafe, 1976; Givón, 1983a, 1998; Prince, 1992, 1981) and information structure (Krifka, 2008; Lambrecht, 1994). Roughly, lexical expressions are used where the speaker does not assume that the addressee can clearly identify an intended participant in a particular role, and nonlexical expressions are used for the reverse situation where the addressee is assumed to unambiguously identify the participant in question.

The issue we are concerned with in the present paper is the choice between the two types of nonlexical expression for objects. Some scholars acknowledge the essential functional closeness of the two forms, for instance Givón (1998:57), but nonetheless claim that the choice is primarily determined by considerations of information packaging and activation. Thus, Givón (1998:57) postulated a slightly lower degree of "*thematic continuity*" in triggering the use of a(n unstressed) pronoun, and for Ariel (1990:58–61; 2000:202–210) the choice follows from the same principles of "*accessibility*" that determine the choice between nonlexical and lexical form. The functional closeness of zero and pronoun is, on the other hand, reflected in the cross-linguistic classification of languages with and without radical pro-drop (see Neeleman and Szendröi [2007] for a relatively recent account; and Huang [2000:50–90] for critical discussion): a language such as English generally requires overt expression of objects, confining zero objects to a restricted set of contexts (see Ariel [1990:59–60] for examples). In Japanese, on the other hand, zero anaphor is the default choice for nonlexical objects, and pronominal objects are confined to contexts where specific aspects of the social relations between interlocutors come into play. Thus, Clancy (1980) found no pronominal objects at all in her Japanese narrative data, a fact she explained by the absence of relevant social

considerations that would trigger the use of a pronominal form in this text type. In languages such as English, zero reference is seen as highly constrained by syntactic constellations of clause combining, whereas in languages such as Japanese and Mandarin, the choice has been attributed to complex pragmatic mechanisms of interpretation (Huang [1994] on Mandarin).

For other languages, the proportion of pronominal and zero objects has been found to be more balanced, essentially leading to greater freedom in the realization of nonlexical objects and the corresponding question about what factors determine the choice. The choice between pronoun and zero in *objects* has received some treatment in variationist studies. For instance, Schwenter (2006, 2014) reported considerable variability between pronominal and zero objects in Brazilian Portuguese and numerous South American varieties of Spanish. He attributed the variation to the semantic saliency features of animacy and specificity. Meyerhoff (2002) reported on constraints on null objects in Bislama: her findings confirm discourse-structural considerations such as activation, as well as priming effects by form and function of an object's antecedent, but she also stated that these effects are not as pronounced in Bislama as they have been found for other languages (Meyerhoff, 2002:338).

Language typology has treated the variable encoding of objects under the label of differential object marking (DOM; Aissen [2003] among many others) or conditioned agreement (Siewierska, 2004:148–162). Some corpus studies of lesser-studied languages also deal with the variable expression of objects as pronoun or zero according to semantic parameters such as animacy or specificity typically associated with DOM: for instance, Genetti and Crain (2003) found that in narrative discourse from Nepali, pronominal objects are restricted to human referents. Schwenter (2014) linked his finding of animacy and specificity to mostly predict the variation between pronominal and zero objects to patterns of case-marking DOM in Spanish: animate specific objects tend to be expressed as pronouns and receive special case marking; animate nonspecific as well as inanimate ones tend to be zero, and receive no special case marking. This suggests a deeper relationship between alternations in grammatical relation systems such as DOM, and the variable treatment of grammatical relations in discourse.

In the generative-oriented literature, the discussion focusses on the correct classification of zero objects as one type of "anaphora"—'Ā-bound variables' (e.g., Huang [1984] for Chinese)—versus *pro* (e.g., Rizzi [1986] for Italian; Chung [1984] for Chamorro); for the purposes of the current paper, we exclude these theoretical considerations from our discussion (see Huang [2000:79–88] for a succinct critical overview).

Unlike English or Japanese, and more similar to Portuguese, Spanish, Bislama, or Nepali, Vera'a does not show a general preference for either pronominal or zero nonlexical objects. Vera'a discourse is characterized by a roughly even proportion of these two forms, and examples like (1b) and (1c) are equally common in our text data. The aim of the present paper is thus to single out the relevant factors that drive the choice between pronominal and zero form for objects. To this end, we investigate a corpus of richly annotated texts (25,646 words) that have been

collected by one of the authors over 10 years of documentary work on the Vera'a language, and that is composed not only of oral narratives, but also of oral descriptions of fish and plant species. This minimal variation in register within our text sample has some bearing on our findings, since it helps tease apart the factors of animacy—in particular humanness—and what we call *discourse topicality* here (see the section on alternations in object realization for explanation), two features that converge in narrative texts. This leads us to the conclusion that the best predictor of pronominal forms in objects in Vera'a is in fact that of discourse topicality rather than the ontological salience of animacy as such. With our study of Vera'a, we thus hope to also contribute generally to some better understanding of referential choice in objects in Oceanic languages and beyond.

The paper unfolds as follows: in the next section, we provide a short overview of the structure of simple clauses, grammatical relations, and forms of reference in Vera'a. We then summarize those factors that have commonly been associated with the alternation under investigation, namely those factors pertaining to information packaging and those pertaining to semantic saliency features, most notably animacy. This is followed by an outline of our corpus data, methodological approach, and coding decisions for specific variables and their proportions in the data. We also discuss issues of potential collinearity and interactions. Next, we present findings from our analysis and discuss them. In the last section, we summarize a number of conclusions from our findings that are relevant not only for cross-linguistic findings on the topic, but that also have some bearing on hypotheses of grammaticalization of object agreement.

VERA'A BASIC CLAUSE STRUCTURE AND VARIABLE REALIZATION OF OBJECTS

Vera'a is an Oceanic language from North Vanuatu, spoken by a growing population of approximately 500 on Vanua Lava (Banks Islands) (Schnell, 2011), in close proximity to the language Vurës (Malau, 2016). The language is essentially configurational, and grammatical relations of subject and object are encoded by means of ordering relative to the verb complex (VC; rendered in boldface in this section), with subjects preceding the VC and objects occupying a postverbal position:

(2) a. [*di*]^{SBJ} ***me kur   ma-ma'*** [*ēn gusuwō*]^{OBJ}
          3SG   FUT devour RED-dead ART rat

          '[When today a cat finds a rat], it will kill and eat the rat.'   GABG.121

     b. [*vēvē-ruō   ga   garaqa*]^{SBJ} ***ne   da*** [=*n*^{2} *go-ru   =n gengen*]^{OBJ}
          mother-3DU STAT new   AOR:3SG do =ART POSS.EAT-3DU =ART RED:eat

          'Then their new mother prepared their meal.'   MVBW.034

(3) a. [*di*]^{SBJ} ***ne   vrigē***
          3SG   AOR:3SG rush

          'And then he ran.'   2.PALA.086

b. *'enei* [=*n   goto*]^{SBJ}   ***ne***   ***m̄es***
   now   =ART hermit.crab   AOR:3SG fall

   'And now hermit crab fell down.'   ISGG.051

The VC functions as the predicate of a verbal clause. With the exception of imperative and some nonfinite clause-chaining constructions,[3] it consists of minimally a marker for tense-aspect-mood-polarity (TAMP) and one verb. As is common for Oceanic languages, multiple verbs can serialize within the same VC, as in (2a), where *ma-ma'* 'be dead' is a serialized verb. The argument structure of verb serializations is composed of that of the individual verbs and bears a combined transitivity value as well as object relation. In addition, different types of adverb can follow onto a single or series of verbs. The right periphery of the VC is optionally occupied by two types of directional, namely directional adverbs, like *sar* in (4a), and directional particles, like *ma* in (4b):

(4)   a. *dir'ōl   **m**   **rēv   kal**   **sar**   [*ēn   nak   su-suō*]^{OBJ}   *anē'ē*
         3TL   TAMP1 drag move.up bushwards ART canoe RED-paddle DEM1.A

         'When they had pulled the canoe ashore […]'   2.PALA.006

      b. *duru =**m**   **'ō'   kal**   **ma**   [= *n   sul*]^{OBJ}
         3DU   =TAM1 carry move.up hither   =ART folk

         'They took the(ir) people here.'   GMV.027.A

Where both types of directional co-occur, the directional particle follows onto the adverb; directional particles constitute the final element of a VC. As for the realization of objects, the examples in (4) show that full lexical NPs follow directional adverbs and particles and are thus outside the VC. Object pronouns, on the other hand, occur *within* the VC where they precede any directional adverb, like *sar* in (5a), or directional particle, like *ma* in (5b):

(5)   a. *dir =**ēk   qērē   ba'a** [**di**]^{OBJ} **sar** *lē   =n   m̄o-gi   =n   nim̄ē*
         3PL =AOR push into 3SG   in   LOC =ART POSS.HOUSE-3SG =ART house

         'They would then push her into her house.'   ISWM.171

      b. *taem   duru =**m**   **van   'ō'   kal**   [**dirē**]^{OBJ} **ma**   *anē*
         when 3DU =TAM1 go   carry upwards 3PL   hither DEM1.A

         'When they had brought them here […]'   GMV.028

The paradigm of free personal pronouns is given in Table 1. They distinguish categories of person, number, and clusivity. Some pronominal forms show a general tendency to drop their final vowels; this is a general phonological process found with different lexical elements, not just pronouns. All forms listed in Table 1 are found in different syntactic contexts. Thus, object pronouns are not systematically reduced in terms of their segmental phonological shape, and Vera'a does not have a specialized set of pronominal forms for objects. However, in their VC-internal position, they are by default not accessible to

TABLE 1. *Paradigm of Vera'a personal pronouns*

	Singular	Plural	Dual	Trial
1 IN	—	*gidē*	*gidu(ō)*	*gid('ōl*
1 EX	*no*	*kamam*	*kamadu(ō)*	*kamam'ōl*
2	*nik(ē)*	*kimi*	*kumru(ō)*	*kimi'ōl*
3	*di(ē)*	*dir(ē)*	*duru(ō)*	*dir'ōl*

prosodic stress, which falls on the last constituent word of a sentence ('house' and the demonstrative, respectively, in (5)). The only instances where object pronouns are stressed is where they are not followed by anything else and thus happen to occupy a sentence-final position, as in (6):

(6)   *alē*   **nik** **ē**   **vus** [***dirē***]^{OBJ}
    alright 2SG AOR:2SG hit 3PL

    'Alright, then you kill them [these fish].'   fish98.022

More research into the prosodic structure of Vera'a is required to ascertain these stress-related observations. Yet, we take examples such as (5) and (6) as essentially suggesting that we are dealing here with by-default unstressed pronouns, much in the sense of Givón (1983a:17) or Ariel (1990:73). Finally, note that object pronouns are mutually exclusive with a coreferent object NP, that is they never "double" an NP and thus do not constitute verb-object agreement (Corbett, 2003; Haspelmath, 2013).

    As shown in (1c), objects may be left implicit. Example (7) shows two instances of zero anaphor objects:

(7)   *dir* *=ēk* **mul**   **'ō**   **kal**   **Ø** *lē*   *n*   *vono-re*   *dir*   *=ēk* **big** **Ø**
    3PL =AOR return carry upwards   LOC =ART home-3PL   3PL =AOR eat

    'Then they would go back up the shore with their catch, fish, mussels, crabs, whatever they found], they would take (it) up to their village and eat (it).' ISAM.006

As zero objects, we only count those instances of absence of an overt object expression where an intended referent is recoverable from discourse context. We do not consider semantic entailment a sufficient criterion to assume an unexpressed object argument (Dowty, 1982).[4] Generally, zero anaphors in Vera'a are pragmatically rather than syntactically conditioned (Huang, 2000). For zero objects, this can be seen from instances of split antecedence, where a zero anaphor relates back to two antecedent expressions, shown in example (8):

(8)   *trust*   *mu-gi*   *mē* *no* *trust*   *mu-k*   *mē* *diē*
    trust   POSS.GEN-3SG DAT 1SG trust   POSS.GEN-1SG DAT 3SG
    kadu   *=m*   **mom**   **serge**   **Ø**
    1DU.EX =TAM1 put   together

'His trust in me, my trust in him, we put (them/both) together [and that lead to all these ideas].'   MVRP.077

Vera'a also allows left-dislocations correlating with objects, and in this case, we either find a resumptive pronoun in its regular position, as in (9a), or a zero anaphor, as in (9b):

(9) a. [*e*   *ruwa*   *re-reñe*   *anē*]ᵢ *duru*ₖ *=m*   *da* '*ō*' *duruō*ᵢ
         PERS.ART   HUM:DU   NSG-female   DEM1.A   3DU   =TAM1   do   carry   3DL

         'Hence, these two girls, the two [parents] looked after them [the girls].'
                                                                                        1.PALA.009

     b. *rekso*   [*= n* *dudu-gi*]ᵢ *kamam*ₖ *me* *as*   *Øᵢ*
         like   =ART leaf-3SG   1PL.EX   FUT   rub

         'Like its leaves, we crush (them) in our hands.'   GAMO.012

Whether a left-dislocated phrase correlates with the object or some other (or no) grammatical relation within the clause is entirely a matter of semantic interpretation and pragmatic inference. Thus, in example (10), an interpretation like 'But birds, they (human beings or some animals) really like (them, that is, the birds)'[5] would be entirely consistent. Only contextual information triggers the correct interpretation of the left-dislocated NP as coreferential with the subject pronoun:

(10) *ba*   [*= n* *men* *sa*]ᵢ *dir*ᵢ *ga*   *mōrōs* *Ø*ₖ
      but   =ART bird   STAT   3PL   EMPH STAT   like

      '[People don't eat (them),] but (it is) birds (who) really like (them) [that is, the fruits of a tree].'   MVMGV.020

Note that the same is essentially true for pronominal in situ objects that share the same person and number values with dislocated phrases, as in (9a) and (11): here, the left-dislocated NP can be understood as coreferential with either subject or object, and only context and encyclopedic clues guide the intended interpretation:

(11) *Qo'*   *di* *ga* *galgal*   *di*   *so*
      pers.name   3SG STAT RED:lie   3SG   QUOT

      'Qo', he betrayed him, saying: […]'   JJQ.409

We conclude that left-dislocated NPs in Vera'a are outside the clausal core and do not themselves enter grammatical relations of subject or object. Where coreferent with subjects, they are always resumed by a subject pronoun, but where they are coreferent with the object, this may be pronominal or zero, as in any other sentence without left-dislocation. The fact that this alternation is preserved under left-dislocation entails that these constructions cannot be regarded a sufficient determining factor for the use of an object pronoun, as may be suggested by Givón's (1976) seminal paper on the development of subject and object

agreement. Hence, a broader approach considering a range of possible factors is required to tackle the alternation at hand.

ALTERNATIONS IN OBJECT REALIZATION: RELEVANT FACTORS

In the literature on pronominal versus zero anaphors, there are essentially two approaches: the generativist approach in terms of parametric constraints on (radical) pro-drop (Neeleman & Szendröi, 2007) and the usage-based approach taken for instance by Bickel (2003), who essentially replaces the parametric approach with a corpus-based one. Both approaches are basically holistic, classifying languages according to specific features.[6] They are not concerned with the factors that drive the formal choice in discourse from individual languages (cf. Neeleman & Szendröi, 2007:673–674).

The factors we considered in our study can be grouped into two categories: (a) Factors pertaining to discourse structure, that is mainly features of the anaphoric relation between object anaphor and its antecedent, but also the global thematic status of object referents, that is to say whether a given text is primarily *about* this referent. (b) Factors pertaining to inherent semantic properties of objects, that is features that are independent of the specific relationship to discourse context, like person and number, and the ontological class of the discourse entity in question.

The first set of factors comprises antecedent distance, function, and form, as well as discourse topicality. Of these, antecedent distance has probably been the most prominent, see for instance studies relating to Givón's concept of look-back (Givón [1983b] and contributions therein) and is also the one central to Ariel's (1988, 1990, 2000) Accessibility Theory (AT). AT and similar theoretical approaches predict that the use of a more explicit form correlates with greater antecedent distance. Although AT focuses on the choice between lexical and nonlexical expressions (see Ariel, 1990:58, 2008:46–47), the choice between "high-accessibility markers" of different types of pronoun and zero has also been attributed to accessibility (see Ariel [1990:58–68, 76–79] for detailed discussion). In our case, pronominal objects would, according to AT, be predicted to have antecedents at greater distance than zero objects. Note that since both pronoun and zero anaphor are high accessibility markers, both forms should show relatively low anaphoric distances, and differences in distance between the two types of expression are expected to be quite minimal. Moreover, since Vera'a pronouns do not mark gender, they are not particularly amenable to distinguishing competing referents. A switch of syntactic function —and related participant role—also diminishes the accessibility of an antecedent, and more explicit forms tend to be used in switch-function contexts; in our case, where an object's antecedent is not itself an object, it should tend to be realized as a pronoun rather than zero and vice versa. Finally, the form of the antecedent is relevant in connection with so-called priming effects—that is the tendency for speakers to reuse the same kinds of structures in series. Research on subject expression shows that the form of an earlier subject may prime a

latter subject to have the same form (Cameron & Flores-Ferrán, 2004; Torres Cacoullos & Travis, 2013; Travis, 2007). Other studies have shown that a speaker's use of other kinds of variants can prime the further use of that variant over another, such as agreement marking or lack thereof (Poplack, 1980; Scherre & Naro, 1991) or contraction of verbs and auxiliaries (Barth & Kapatsinski, 2017). In our case, this means that the form of the antecedent of an object should trigger the same form in the anaphoric object; thus, zero objects should tend to have zero object antecedents, and pronominal objects pronominal object antecedents.

Another group of factors considered here composes the more discourse-stable semantic feature of ontological class membership, most notably animacy, already found to be relevant by Schwenter (2006, 2014) for the same alternation in Spanish and Portuguese, and Genetti and Crain (2003) for Nepali. We distinguish human, animate, and inanimate referents. Drawing on this previous research, we expect pronominal objects to be more likely with human referents than nonhuman ones and more likely with animate than with inanimate ones.

A factor featuring prominently in the discussion of DOM and object agreement is that of (sentence) topicality (Dalrymple & Nikolaeva, 2011; Krifka, 2008). This is related to Givón's (1976) seminal hypothesis that argument agreement arises from so-called *topic shift* constructions involving the left-dislocation of the object NP (see Givón, 1976:156–160), which in turn would trigger the use of object pronouns that resume the referent in situ. Object pronouns should be particularly frequent with topicalized, left-dislocated NPs, since these objects must be topics.

Finally, we include what we call here *discourse topicality*. *Discourse topic* is understood here in the sense of Chafe (1994:120–121): it is that discourse entity that an entire text is about and that makes the text interesting. As such it is different from both *sentence topics* (Krifka, 2008) and from topicality in the Givónian tradition, which is essentially a measurement of referential continuity of individual participants across sentences in a given (stretch of) discourse (Givón, 1983a). What we are interested in here is the communicative purpose of a text: with regard to what entities is a text relevant? As such, discourse topic is part of what Biber and Conrad (2009:39–47) termed the "*situational characteristics*" of a text with specific register and genre properties.[7] The Vera'a corpus for this study consists of three groups of texts (narratives, fish descriptions, plant descriptions) that differ with regard to their discourse topics: for narratives, we assume that all human or human-like referents (including anthropomorphized animals in fables) are the discourse topics, since these are the entities that will be of most concern to the audience, and thus constitute part of a narrative's tellability. And for the two types of descriptive texts, it is the fish or plant species, respectively, excluding other fish and plants, since the communicative purpose of producing the text is to provide information about these species. We will see in the results section that discourse topicality is the best predictor for the choice between pronominal and zero objects, with discourse topics favoring expression by pronouns. Although discourse topicality and ontological class/animacy are independent of each other, they show a

tendency of convergence in human referents. Thus, in order to tease apart these two factors, we require a sample of texts with discourse topics of different ontological classes and an evaluative methodology that can directly compare animacy and discourse topicality.

METHODOLOGY

Corpus data

Data for the present study comes from a total of 36 texts in the Vera'a corpus, collected by the first author from 2007 to 2013 on Vanua Lava, North Vanuatu. The larger part of the data comprises 10 traditional narratives. These are supplemented by descriptive texts, namely 9 descriptions of fish species and 17 descriptions of plant species, which together make up a much smaller part of the data, an overview of which is provided in Table 2. The three types of text differ according to register properties (in the sense of Biber and Conrad [2009:39–47]), including different discourse topics, as we have outlined. Narratives represent a register in itself that has the broad communicative function of relating adventures of human or anthropomorphized protagonists to the addressee who may feel more or less sympathetic with their characters; there may occasionally be a moral to a story. Descriptions are of a different register, bearing the primary communicative function of presenting factual information about the fish or plant species under discussion. Thus, the three types of text have three different discourse topics in terms of ontological class membership: narratives have human beings as their topics, whereas the descriptions have fish and plant species as their topics, respectively.

Variables considered for statistical analysis

Based on the literature and knowledge of the Vera'a data, we have identified seven variables as having a potential impact on pronominal versus zero object expression: (a) animacy, (b) discourse topicality, (c) form of the antecedent, (d) function of the antecedent, (e) distance of the antecedent, (f) discourse interruption, and (g) speaker.

Animacy.   Referents are coded either as animate or inanimate. Humans, spirits, and anthropomorphized animals capable of speech, thought, and planning, and ordinary animal referents are coded as animate, and plants, plant parts, body parts, and other inanimate objects as inanimate. Animacy may interact with the following predictor, discourse topicality. We are interested in the relative strength of these predictors and in determining when each, if any, matters.

Discourse topicality of the object referent.   Referents are coded as either topical or nontopical, in the sense of Chafe (1994), as we have outlined. In narrative texts, human beings, spirits, and anthropomorphized animals are considered to be the discourse topics and are thus coded as topical. In the

TABLE 2. *Overview of Vera'a corpus data*

Text ID	Speakers *n*	Texts *n*	Corpus size		Objects considered (pron, zero) *n*
			Clauses	Words	
Narratives	8	10	3789	20,851	300 (133, 167)
Fish descriptions	7	9	330	2439	58 (37, 21)
Plant descriptions	3	17	362	2356	92 (21, 71)
Totals		36	4481	25,646	450 (191, 259)

descriptions about fish and plants, fish and plants under consideration in the specific text are considered topical. Referents considered to be nontopical are nonhuman, animate beings in narratives and descriptions (although not the fish under discussion in a descriptive text, rather a fish species other than the one described in a text, or a fish or pig in a narrative), inanimate objects in a narrative or description (other than the plant species under discussion in a plant text, rather a bow or a basket), body parts and plant parts.

Table 3 shows the distribution of tokens by form, animacy, and topicality. We see in the Total column that animate objects are most often expressed with pronouns (76%, 173 of 227) and inanimate objects are overwhelmingly expressed with zeros (92%, 205 of 223). In the Total rows, we see that topical objects are often pronouns (78%, 186 of 240) and nontopical objects are overwhelmingly zero (98%, 205 of 210). We also see that animate objects are topical (95%, 215 of 227), while inanimate ones tend to be nontopical (89%, 198 of 223).

Form of the antecedent. The antecedent may have been expressed as either a lexical NP or a pronoun or it may have been a zero. An issue here is that for antecedents that are zeros or pronouns, there are factors that condition the form of that antecedent pronoun. Our analysis is designed to detect the strength of this predictor, given the other predictors. Table 4 shows the distribution of tokens by their form and the form of their antecedent.

Function of the antecedent. The grammatical function of the antecedent could either have been a subject of either a verbal or nonverbal clause, an object, a left-dislocated expression in preclause position, or *other*, which may have been an oblique argument, a nonverbal predicate, a possessor, an adjunct, a vocative address term, or a stretch of discourse (discourse-deictic reference). There were relatively few of each of these other kinds of functions. Where an object's antecedent is the subject in the same clause, the object is always expressed by a pronoun. This is obviously a rule outside the general mechanisms assumed by accessibility-related models. Since there are only seven instances of such subject-bound object pronouns in the corpus data investigated here, we do not expect this to bear any effect on our findings. Note that when the antecedent and target are both objects, then they have the same function. Table 5 shows the distribution of objects by their form and the function of the antecedent.

TABLE 3. *Object form by animacy and discourse topicality (n = 450)*

| | | Discourse topicality | | |
| | | | | |
Form	Animacy	Topical	Nontopical	Total
Pronoun	Animate	173	0	173
	Inanimate	13	5	18
Subtotal		186	5	
Zero	Animate	42	12	54
	Inanimate	12	193	205
Subtotal		54	205	
Total		240	210	450

TABLE 4. *Object form by form of the antecedent (n = 450)*

| | Form | | |
| | | | |
Antecedent form	Pronoun	Zero	Total
Lexical noun phrase	39 (21)	148 (79)	187
Pronoun	137 (84)	26 (16)	163
Zero	15 (15)	85 (85)	100

Note: Percentages appear in parentheses.

Distance of antecedent. Objects were coded for how many clauses previously the antecedent occurred. The value 0 was used in cases where the object antecedent occurs in the same clause, either as subject (equivalent to reflexive constructions in other languages) or possessor in some other constituent NP, but also cases where the antecedent is a left-dislocated NP in preclausal position, which in fact comprises the bulk of cases for this category. The value 1 was used for one clause previous, 2 for two clauses previous, 3 for three or more clauses previous. We also distinguished objects in relative clauses with their antecedent being the modified noun.[8] Table 6 shows the distribution of objects by their form and the distance of the antecedent.

Discourse interruption. Objects were coded for presence or absence of an interruption or disfluency in the text or a shift of levels of narrative representation, for example switching to direct speech in a text (see Lichtenberk, 1996). Having an interruption could theoretically reset the conditioning factors, in particular lowering the antecedent's accessibility. Table 7 shows the frequencies of interruptions and object forms.

Speaker. Individuals may make their own contributions to variable patterns of object expression. It may be the case that some people simply have tendencies to express objects as pronouns or zeros. Therefore, we also include speaker (n = 14) as a variable.

TABLE 5. *Object form by function of antecedent (n = 450)*

	Form		
Antecedent function	Pronoun	Zero	Total
Object	56 (25)	162 (74)	218
Subject	92 (73)	34 (27)	126
Left-dislocated	13 (26)	37 (74)	50
Other	30 (54)	26 (46)	56

Note: Percentages appear in parentheses.

TABLE 6. *Object form by distance of antecedent (n = 449)**

Antecedent distance	Pronoun	Zero	Total
Same clause	23 (42)	32 (58)	55
One clause	119 (43)	156 (57)	275
Two clauses	22 (49)	23 (51)	45
Three clauses	24 (47)	27 (53)	51
Relative clause	3 (13)	20 (87)	23

Note: Percentages appear in parentheses. *There was one token of discourse-deictic use, where the object anaphor refers back to a stretch of preceding discourse.

Statistical analysis

While an apparent possible approach to testing the predictors is logistic regression, it is not ideal for the investigation of our data, since animacy and discourse topicality are expected to show multicollinearity due to the inclusion of narrative texts, where human referents are also topical ones. It is the inclusion of the Vera'a plant and fish stories that allows for the testing of animacy and discourse topicality against each other, so we also need a statistical method that can test these predictors simultaneously. With recursive partitioning, we are able to test predictors and eliminate those that do not contribute to significant effects in the data without doing a stepwise regression procedure. Suitability of the recursive partitioning for our particular data and research question also lies in testing interactions. Some of our predictors have several levels: function of the antecedent has four levels; distance from the antecedent has six. When testing for interactions, it is easy to see that crossing these kinds of predictors (here what would be a six by four matrix) can lead to missing test cases (empty cells), which will cause a typical logistic regression to fail to converge.

Two types of recursive partitioning analyses are described in the results section, a binary classification tree using the partykit package and a variable importance assessment of a random forest using the party package (Hothorn, Buehlmann, Dudoit, Molinaro, & Van Der Laan, 2006; Hothorn & Zeileis, 2015; Strobl, Boulesteix, Zeileis, & Hothorn, 2007; Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008) in *R* (*R* Core Team, 2015). A binary classification tree divides the

TABLE 7. *Object form by discourse interruption*

Discourse interruption	Pronoun	Zero	Total
Interruption	15 (48)	16 (52)	31
No interruption	176 (42)	243 (58)	419

Note: Percentages appear in parentheses.

data into two sections based on which data points are most different from each other, using the given variables. An initial split is shown across levels of the independent variable that provides the best differentiation of the data's dependent variable and the data are divided into two nodes. Under each node (or branch of the tree), the data are then split into another two sections based on which data points are most different from each other using the remaining independent variables and remaining levels of independent variables from further up the tree. Because each set of data under a node is looked at anew, the same variable can be used again in a lower level of the tree, with different levels. This splitting continues recursively until the data under each final branch is relatively homogenous. The data in a node is considered homogenous enough when there are no significant differences (using a *p*-value) between the data points, given the independent variables. This allows *p*-values to be provided for each split (cf. Hothorn et al., 2006). The algorithm for creating the tree model will not necessarily use all independent variables listed in the model specification. If there are independent variables that would not make a significant split in the data, they go unused.

We also present a conditional random forest analysis. When many classification trees are averaged, factors can be ranked by their importance, determined by which factors most often make a significant split in the data, especially at higher levels in the tree. The analysis resamples across subsections of the data and uses different samples of the independent variables (here four out of seven possible ones) to increase the variation in the possible trees. Slightly different data subsets and variable combinations will result in potentially different variables performing well (sometimes animacy of the referent, sometimes function or form of the antecedent, etc.), but after looking at many cases ($n = 1000$), one will generally be the best factor for predicting the outcome of the dependent variable. This one factor will be ranked higher than the other in a variable importance ranking. A conditional permutation of the variable importance ranking, although computationally expensive, ensures that the evaluation of a variable's importance takes into consideration its behavior in relation to other variables in its ranking (Strobl et al., 2008; Tagliamonte & Baayen, 2012).

Potential multicollinearity and interactions

Each pair of predictors was checked for collinearity. Only one pair of predictors had *r* above .5 (or below − .5). Animacy and discourse topicality are collinear with

$r = .84$. In an analysis type such as generalized linear regression, we would not be able to felicitously include both these predictors.[9] However, in a classification tree, we can see which predictor makes the first (best) split of the data and how the predictors interact. Furthermore, as these kinds of recursive partitioning analyses can handle nonmonotonic effects, we can detect whether the predictor effects the whole dataset or just a portion of it. Because we are interested in the strength of animacy versus discourse topicality, using a classification tree is advantageous, allowing us to compare these collinear predictors directly against each other.

If structural priming is at work in the conditioning of the object form, we should see an interaction in the form and the function of the antecedent. If there is priming, we would expect there to be a higher rate of pronominal object forms when the antecedent is a pronominal object and a higher rate of zero object forms when the antecedent is a zero object. Higher rates of pronominal objects when the antecedent is a pronominal subject (or higher rates of zero objects when the antecedent is a zero subject) would not be due to structural priming. Table 8 shows the rates of zero/pronominal object forms by antecedent form and function. In Table 8, it should be clear that objects with pronominal object antecedents are more likely to be pronouns, but that when the antecedent is a pronominal subject, the object is also more likely to be a pronoun. Objects with zero antecedents are more likely to have a zero, but so are objects with lexical NP objects. The results section shows the significant interaction of these predictors, given the other predictors. As is indicated by the numbers in Table 8, structural priming cannot be the (sole) factor at play in the matching of antecedent and target forms.

RESULTS

Recursive partitioning results

The binary classification tree (Figure 1) and random forest variable importance ranking (Figure 2) show consistent results in modelling the prediction of speakers' choice to express objects as pronouns or zeros.[10] The classification tree has splits and nodes. The splits show the predictors that make a significant difference to the classification of pronominal versus zero object pronouns for each subsection of the data. The nodes show the proportion of zero or pronominal expression for the final partitions of the data. This classification tree has 4 splits and 5 nodes. Figure 1 shows that the first split in the data comes from the discourse topicality of the referent. Under split 1, the left branch shows that nontopical referents favor zero object expression, as there is a higher proportion of dark gray than light gray in nodes 3 and 4. Under the right branch of split 1, topical referents (humans, spirits, and anthropomorphized animals in narratives, fish and plants under discussion in descriptive texts) favor pronoun object expression, as seen by the higher proportion of light gray in nodes 7, 8, and 9.

TABLE 8. *Object form by antecedent (n = 450)*

		Antecedent form		
Form	Antecedent function	Pronoun	Zero	Lexical NP
Pronoun	Object	43	5	8
	Subject	71	9	12
	Left-dislocated	1	0	12
	Other	22	1	7
Zero	Object	6	79	77
	Subject	13	6	15
	Left-dislocated	2	0	35
	Other	5	0	21

Form of the antecedent results in a further split on both sides of the tree. Under split 2 on the left side of the tree, we see that when an antecedent is a pronoun, the object is more likely to be a pronoun than when the antecedent is a lexical noun phrase or zero. Generally nontopical referents are highly likely to be zeros, but this pattern is attenuated when the antecedent is a pronoun. When the antecedent is a lexical NP or zero, then a nontopical referent is highly unlikely to be expressed pronominally (only two tokens, both with a lexical antecedent). Under split 5 on the right side of the tree, we see the same pattern. Generally topical referents are likely to be expressed pronominally, and this tendency is increased when the antecedent is also a pronoun and slightly less likely when the antecedent is a zero or a lexical NP. A split based on the same configuration of levels for one variable in multiple parts of the tree means that this effect is found throughout the data. There is no interaction with the discourse topicality of the referent, because the pattern remains the same in both splits: zeros and lexical NPs favor zeros to a greater degree and pronouns favor pronouns to a greater degree, in both subsections of the data.

The function of the antecedent results in the last partition of the data under split 6. When the antecedent was an object or a left-dislocated topic (dislocated), the object was more likely to be expressed as a zero. When the function was a subject, or had another function, the object was more likely to be expressed as a pronoun. This shows that there is likely priming by zero object antecedents on zero object targets. Additional testing showed that this split was still significant after removing lexical NP objects.

Despite split 6, which indicates structural priming for zero objects, the form of the antecedent influences the form of the target, whether or not the antecedent is an object. That is, there is form matching also when a pronominal antecedent is a subject, indicating that structural priming is not the sole underlying motivation for matching. We take this point up in the discussion.

Other predictors (animacy, discourse interruption, and distance of the antecedent) do not appear in the tree in Figure 1, because they provided no significant splits in the data. Figure 2 shows a conditional variable importance ranking for a 1000-tree random forest of classification trees. This forest shows
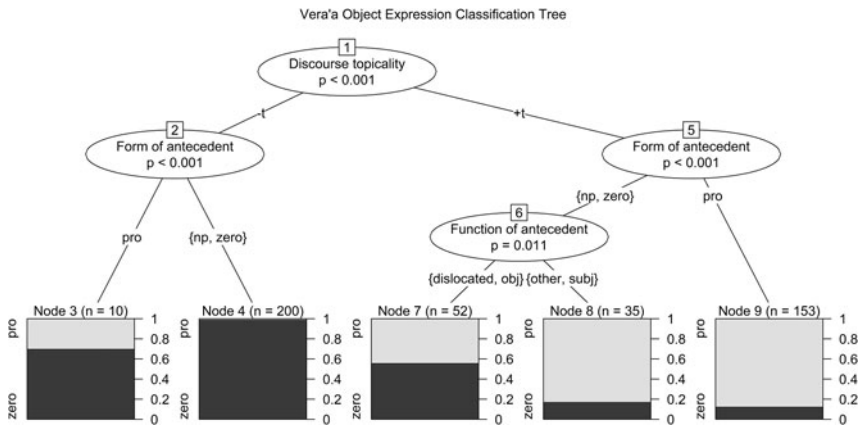
FIGURE 1. Classification tree of Vera'a object expression.

that the three variables that appeared in the tree in Figure 1 are indeed the variables most likely to have an effect, even when taking the other variables into account. According to Strobl, Malley, and Tutz (2009:21), the "variables whose importance is negative, zero, or has a small positive value that lies in the same range as the negative values, can be excluded from further exploration." For the ranking in Figure 2, this means that distance of the antecedent, discourse interruption, and animacy are uninformative for predicting zero or pronominal expression of the object, when the effects of the other predictors are taken into account.

Taken together, the tree and forest show a united picture, which is not always a foregone conclusion with these kinds of analyses. For Vera'a object expression, the discourse topicality of the referent is the top contributing factor, showing us that when a referent is topical, it is much more likely to be expressed as a pronoun than a non-discourse-topical one (versus zero). The form of the antecedent influences the form of the object. An object with a subject antecedent is more likely to be expressed as a pronoun, and one with an object antecedent is more likely to be expressed as a zero.

Methodological discussion and an alternative approach

Using conditional inference trees and random forests, instead of logistic or linear regression with sociolinguistic data, is still relatively new (cf. Tagliamonte & Baayen, 2012). There are some clear disadvantages to using recursive partitioning. One disadvantage with the partykit (Hothorn & Zeileis, 2015) or party (Hothorn et al., 2006) implementation is that it does not provide an easy means to assess variation at the speaker level (i.e., a random intercept and slope in a mixed-effects regression framework). Including speaker as a variable is possible in a random forest, but in a conditional inferencing tree it can cause confusing splits because there are so many levels in the variable. Although we
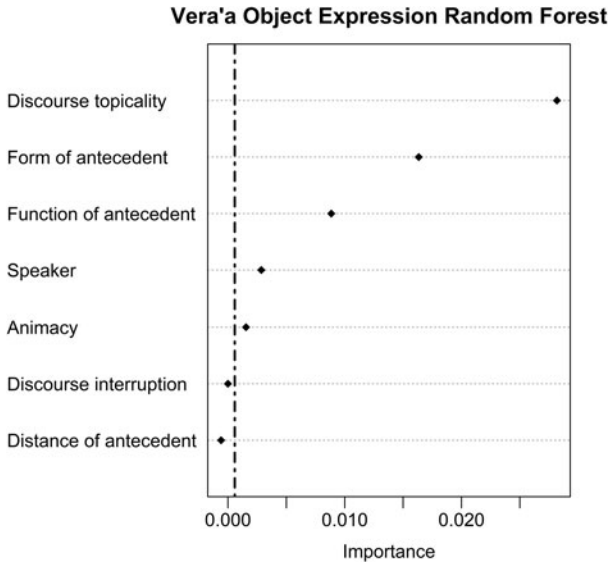
FIGURE 2. Variable importance ranking from random forest for Vera'a object expression.

were able to include speaker in the tree displayed in Figure 1, and see that it did not contribute to any significant splits due to low interspeaker variance, one would often not include speaker as a variable in a conditional inferencing tree, just as one would usually not include it as a fixed-effect in a regression model. Another disadvantage is the lack of coefficients in a classification tree figure. In a regression model, we can compute coefficients that show the contribution of each predictor to the fit of the model. The structure of a classification tree model is fundamentally different, and because predictors do not always affect every data point, one cannot provide coefficient values. Finally, some see recursive partitioning as "data dredging." To address this last point, we have clearly stated predictors that have theoretical motivation. Testing these predictors, and eliminating the predictors that do not have strong predictive contributions simultaneously, is an advantage in our eyes.

As an alternative, we built two generalized linear (logistic) mixed-effects models, one including discourse topicality (Table A1), form of the antecedent, and function of the antecedent as fixed effects and speaker as a random intercept with random slopes on discourse topicality. The other has animacy in place of discourse topicality (Table A2). The models show that either discourse topicality or animacy will be significant if it appears in a model without the other. However, the Akaike information criterion (AIC) (Akaike, 1973) for the model with discourse topicality is 279.1, as opposed to 321.1 for the model with animacy, meaning we should prefer the former, and the latter has essentially no support as the difference between the models is $\Delta > 10$ (Burnham & Anderson, 2002). The recursive partitioning shows this same result in a more elegant

manner, showing that animacy makes no significant contribution when discourse topicality is available as a predictor (see the appendix for the generalized linear [logistic] mixed-effects model results).

DISCUSSION

The first thing to note is that our findings do not lend support to AT (Ariel, 1988, 1990), at least not when considering antecedent distance as its most central factor: antecedent distance is not a relevant factor in our analysis. This is illustrated by numerous examples of long-distance zero anaphors in our Vera'a corpus, like the following (cf. Table 6 for figures):

(12) a. | *alē*   *di*   *ne*   *le*   =*n*   *mēdēr*
               alright 3SG AOR:3SG take =ART rock
          | *di*   *ne*   *burbur*   *gōr*   *sa*   [=*n* *'erē* *qoro- liē*]$_k$
               3SG AOR:3SG DISTR:shut block EMPH =ART PL   hole- cave
          | *di*   *ne*   *van qēl*   *ma*
               3SG AOR:3SG go descend hither
          | *di*   *ne*   *le*   =*n* *mēdēr ne*   *vōwal* | Ø *ne*   *bur gōr* Ø$_k$
               3SG AOR:3SG take =ART rock   NUM.ART one   AOR:3SG shut block

        'Alright, then she took rocks and she shut all the (openings of) **the cave**. She came down here, she took a rock and shut (**it**).' HHAK.044–045 [narrative]

    b. **wōñarñar**$_k$ *(0.9) kamam me 'o'o'og lē -n (0.9)'e   vōwal asenē*
       almond.sp   1PL.EX FUT IPFV-stay LOC =ART year one DEM.MAN1.A
      | *si* =*n* *lañ*   *vus* *m*   *vus kamam (0.6)*
        if =ART wind hit TAMP1 hit 1PL.EX
      | *ne gitag*   *go-mam*   *ēn gengen*
        ? not.exist POSS.EAT-1PL.EX ART food
      | *alē* *kamam me* *van* | Ø *me* *le*   Ø$_k$ *(0.7)*
        allez 1PL.EX FUT go   FUT take

       'The **wōñarñar**, as we live here any given year like this, if a hurricane hits **us** and we don't have any food to eat, alright, then we go and get (**it**).'
                    PANR.002–006 [plant]

    c. *di*$_j$ *ne*   *dōñ*   **[*ēn naw*]**$_k$ | Ø$_k$ *ne*   *kal* *ba'a mē diē*$_j$ |
      3SG AOR:3SG swallow ART saltwater   AOR:3SG enter into DAT 3SG
      ga: margō (0.2) | *di*$_j$ *ne*   *lē'u*   *lu*   *kēl* Ø$_k$ | *ga* *misin*
      STAT quick   3SG AOR:3SG release through back   STAT long.time

       'When it swallows **sea water**, for (**it**) to enter it takes no time, but when it releases (**it**) again, that takes a long time.'   Fish15.004–006 [fish]

In all examples in (12), the antecedent of a zero object in each final clause is an NP occurring three clause units away. Obviously, the intended patient or theme referent is identifiable through clues coming from the verbal semantics, world knowledge, and discourse context. Thus, the higher distance as such does not trigger the use of the more explicit form of a pronoun. Conversely, object pronouns are frequently (cf. Table 6) found in contexts with very low antecedent distance, as in (13), where their use would appear not only redundant, but in fact misleading from an AT point of view since it would trigger addressees to recover the appropriate antecedent of the object pronoun from a distance larger than the immediately preceding clause.

(13) | di_k =m   wotoqtoqo | e   =n   maru-gi anē   =m   vus **die**$_k$
       3SG =TAMP1 pregnant   DISC =ART uncle-3SG DEM1.A =TAMP1 kill 3SG
       | Ø ’ō   **di**$_k$ lē =n   ’ōnō- ’ev’ē   voruo’ | Ø vus   **diē**$_k$
         carry 3SG LOC =ART place- throw garbage   kill 3SG

  ‘When she got pregnant, her uncle killed **her**. Took **her** to the litter place and killed **her**.’   ANV.010–012 [narrative]

Contrary to current models of referential choice, like AT, the choice between pronoun and zero in objects does thus not relate to instructions of antecedent retrieval, at least not if distance is taken as the major factor. Similarly, discourse interruptions do not show any effect in our data, a finding likewise not easily accountable in terms of accessibility.

Turning now to the three most relevant factors, we first consider switch-function and same-function objects. Our analysis indicates that where an object's antecedent is not itself an object, but say a subject of a preceding clause or some other function, pronominal realization is more likely; conversely, where the antecedent is also an object, resulting in a chain of same-object reference, there is a preference for zero realization (cf. Table 5). The following examples illustrate this pattern:

(14) a. **dir** =m   van kal   sar   ma
        3SG =TAMP1   go move.up   bushwards hither
        Tumeren   *ne*   *’ēn* **dirē**
        pers.name AOR:3SG see   3PL

  ‘As **they** were walking up the shore, Tumeren spotted **them**.’
                                                  JJQ.270 [narrative]

  b. |^1 e   duru^i =k   mom sur suwō [=n ’erē qebusu-n e   ruwa]^k
       DISC 3DU =AOR put down thither =ART PL finger-CS ART two.people
       anē |^2 Ø^i =k   ’ē-’er   Ø ^{k}|^3 dir^i =ēk   ’ēr’ēr   Ø^k
       DEM1.A   =AOR   RED-peck   3PL   =AOR   RED:peck

  ‘Then the two (birds) put down **the fingers of the two** (dead girls) and started to peck at (**them**). They kept pecking (**them**).’ ANV.092–093 [narrative]

Examples like (13) or (15), where a pronoun is used in same-function contexts, are much less common than zero objects in this context (cf. Table 5):

(15) *nik me 'ēn **di** sa 'ekēnē'*
      2SG FUT see 3SG EMPH LOC.DEM1
      *nik me go' m̄as **diē***
      2SG ABIL1 hook ABIL2 3SG
      *gidē ga go-go' **diē***
      1PL STAT RED-hook 3SG

'[It forages here. When you go down to the sea to hook for it,] you will find **it** right there. You can hook **it.** We hook **it**.'   fish01_009–010 [fish]

Similarly rare are examples of zero objects with a nonobject antecedent, for instance a subject:

(16) | *di ga manman ēn 'añsara*
        3SG STAT RED:poisonous ART person
      | *'aklē- 'ekē **di** ga wiē (0.2)*
        some- place 3SG STAT good
      | *rekso a Vera'a gēnē (0.3) gidē me big **Ø***
        like LOC.SP *place.name* here 1PL.IN FUT eat

'It is poisonous for humans. But in some places, it is ok. Like here at Vera'a, we eat (**it**).'   fish13.006–008 [fish]

The greater likelihood of objects to be expressed by a pronoun where their antecedent is a subject may have two plausible explanations. For one thing, this may suggest that a switch in function triggers the use of a pronoun in object function due to considerations of accessibility. While this interpretation is in principle justified, we should also consider the more global tendency of subjects to often have discourse-topical referents. This tendency is very pronounced in the narrative texts in our data, where approximately 85% of all subjects have human discourse-topical referents. In fish descriptions, this proportion is somewhat lower, namely 70%. Only in plant descriptions do we find a proportion of only 40% of subjects with discourse-topical referents. Given that narrative texts account for a much larger proportion of our data, this means that the third split to the left under node 3, relating to antecedent function, may in fact merely be a by-product of the overall most relevant factor of discourse topicality: discourse topical status triggers expression of objects by a pronoun, and at the same time that same referent is likely—for the same reason—to be a subject in the preceding clause.

As for discourse topicality, our analysis shows that all humans and those nonhuman referents that are also the discourse topic (fish and plants under discussion in descriptive texts) tend to be realized as an object pronoun, hence the split under node 1. Objects with nonhuman, animate referents that are not also the discourse topic tend to be realized as zero rather than pronoun (cf. Table 3). This is the case with animals referred to in narratives, as shown in (17),

but also with reference to fish that are not the topic of the description at hand, as shown by the comparative examples in (18).

(17) Ø_j *le*   *suw*   *sa*   [=**n** **raw**]_k   *ē*   *mē-n* [*e* *Dōl*]_l   *va'anē*
      transfer thither EMPH =ART intrasex.pig DEM3 DAT-CS ART *pers.name* TEMP.DEM1

    Ø_j *le*   *kēl*   *suwō*   **Ø_k** [...]
       transfer   back   thither

    [*e*   *Dōl*]_l   *ē* [...]   *ne*   *kalak*   *anē*   Ø_l *le*   *kēl* **Ø_k**
    ART *pers.name*   DEM3   TAM2:3SG get.up DEM1.A   transfer   back

'And then (the two) give that **hermaphrodite pig** to Dōl, give (**it**) back to him. [Then the two stay there for a little while longer until he says: "Alright, you two guys better go home back home now." Alright, and so they go home.] And Dōl, he gets up and takes (**it**) again.'   1.AS.108–114 [narrative]

(18) a. *si nik ē*   *kar*   *qal* **diē** *di* *ne*   *nen*
       if 2SG AOR:2SG shoot hit 3SG 3SG TAM2:3SG puff.up

     'When you hit **it** [the puffer fish], it will puff up.'   fish15_030 [fish]

    b. *di*   *ga*   *kur* [**ēn** *'erē mes a* =**n**   *ga*   *lu-luwō* *ga*   *tiktik*]_k
       3SG STAT eat ART PL fish REL =ART STAT RED-big   STAT   RED:small

     di   *ne*   *qal* *gōr*   **Ø_k**
     3SG AOR:3SG hit   secure

     'It [a shark species] eats **all kinds of fish, large and small ones**, as it finds (**them**).'   fish15_011 [fish]

Hence, the most relevant factor here is not ontological class, in particular animacy, as such, but rather the discourse topicality of an object referent. Even inanimate plants tend to be referred to with object pronouns where they are the topic of a description (cf. Table 3), as in (19a), whereas other inanimate referents that are not the discourse topic tend to take zero reference, as in (19b):

(19) a. **qa'aga** *(1.2) diē (0.8) kamam* [...] *ga le*   **diē**
      tree.sp   3SG   1PL.EX   STAT transfer 3SG

     'Qa'aga, it, we take **it**, [like it's good for hoise building].'   MVQ.002 [plant]

    b. *nik ē*   *lō'*   *mulumlum* [**ēn**   **wōvinqa**   **anē**]_k *(0.3)*
      2SG AOR:2SG smash soft   ART   coconut.shell   DEM1.A

     van 'ō'   **Ø_k** *lē*   =**n**   *niñ*   *'a-'ar(0.5)*
     go   carry   LOC =ART house RED-pray

     'You smash the coconut shell to become soft and take (**it**) to church.'
                            GAS.011–012 [plant]

Nonetheless, humans and anthropomorphized living beings are still more likely to be pronominal than fish and plant species under discussion (cf. Table 3). Our analysis does not suggest an overall explanation for this finding. We take this point up in connection with first and second person objects.

While our cross-register investigation helps tease apart the semantic feature of animacy and the discourse factor of *discourse* topicality, we should also stress that the latter does not correlate with *sentence* topicality, as suggested by Dalrymple and Nikolaeva (2011). It was shown that topicalization constructions in Vera'a allow for both pronominal and zero resumption in situ. In fact, for all cases of topicalization, we find roughly the same distribution of both forms as predicted by discourse topicality—that is where the topicalized referent is also a discourse topic it will tend to get resumed by a pronoun, as in examples (20a) to (22a), but if it is not a discourse topic it will tend to be left zero, as in examples (20b) to (22b):

(20) a. *ekē   e   raga   anē   no   me vus **dir** siviē*
         oh.dear PERS.ART people DEM1.A 1SG FUT kill 3PL   how

         '"Oh dear, these people how am I ever going to kill **them**?"' JJQ.401 [narrative]

      b. *riar   mu-gi   […]   di   =m   mom   Ø'i suwei*
         k.o.bow POSS.GEN-3SG   3SG   =TAMP1 put   DEL down

         'His bow, he put (it) down, (he didn't climb with (**it**)).'
                                                                    ISWM.133–134 [narrative]

(21) a. *ba =n (0.3) =n   ge   'ew'i a   salsal   kēkē   anē*
         but =ART   =ART THINGY other REL   RED:float on.top   DEM1.A
         *gidē   ga   go-go'   **diē***
         1PL.IN STAT RED-hook 3SG

         'But that one that floats at the top (i.e., close to surface) of the saltwatrer, we hook **it**.'   fish04_CH.010 [fish]

      b. *e   =n   naw   anē   di   ne   le   lu   Ø*
         DISC =ART saltwater DEM1.A 3SG   AOR:3SG transfer out
         'a'ag   ēn   rekso =n   ēs   nō-gi
         according.to ART like   =ART life POSS.VAL-3SG

         '[It swallows seawater in order to puff up with. Alright, in order to get rid of the seawater, it takes a long time. It will float up towards the surface, and it will stay there for a while.] Now, this seawater, (it) takes it out following its ways [it knows how (to do it)].   fish15_MV.035–039 [fish]

(22) a. *qa'aga (1.2) diē (0.8) kamam […]   ga   le   **diē***
         tree.sp   3SG   1PL.EX   STAT transfer 3SG

         'The black palm, it, we use **it** [as posts in a house].'   MVQ.002 [plant]

      b. *rekso =n   dudu-gi kamam me   as Øs*
         like =ART leaf-3SG 1PL.EX FUT rub

         'Like its leaves, we crush (**them**).'   GAMO.012 [plant]

This suggests that discourse topicality and sentence topicality do not converge when it comes to the use of pronouns for objects in Vera'a, and discourse topicality is a factor in its own right, distinct from the pragmatic relation of topic within a sentence.

Lastly, we observed priming effect in the choice between pronoun and zero objects, so that a pronominal form will trigger the use of another pronoun in a latter mention of the same referent, and likewise for zeros. However, given the much stronger factor of discourse topicality, it seems likely that this effect is epiphenomenal of this overall most important factor: since discourse topicality does not vary for any given discourse referent in a given text, serial reference to this participant will necessarily lead to the use of the same form throughout the series. Such serial effects are in this way merely the effect of an association of referential choice with a text-stable feature. This may also explain why the effect is observed not only with pronominal object antecedents, but also for pronominal subject antecedents.

Before we proceed to draw some conclusions from our findings, we would like to briefly take up the behavior of speech-act participants with regard to the choice at hand: we excluded first and second person objects from the analysis because we essentially do not find any variation here in terms of their formal expression. That is to say, first and second person objects are (nearly)[11] categorically pronominal. The following are some illustrative examples:

(23)   a. *e   nik   'ovi'I vusvus **no***
           DISC 2SG   PROH   RED:hit 1SG

           '[And spider says:] "Don't kill me!"'   JJQ.067 [narrative]

       b. *di   me   man   m̄as   **nikē***
           3SG ABIL1 poison ABIL2   2SG

           'It can poison you.'   fish16_028 [fish]

       c. *si =n   lañ   vus   m   vus   **kamam**   (0.6)*
           if =ART wind   hit   TAMP1 hit   1PL.EX

           'If a hurricane strikes us [we will resort to this plant as staple].'
                                                                    PANR.004 [plant]

Note that second person singular pronouns can have generic reference, as in (23b). How can these observations be related to our findings regarding third person objects? There are essentially two possibilities: the first one is that we are dealing with two essentially autonomous subsystems, so that Vera'a grammar includes a rule to realize first and second person objects overtly, a rule that is absent from third person objects.

Another possibility is that first and second person objects share the same functional considerations with third person objects, but that these are much more pronounced with the former, so as to lead to the near-categorical use of pronouns. One functional explanation often discussed in the typological literature in regard to DOM (e.g., Comrie, 1989:128) is in terms of markedness reversal (see also Aissen, 2003): the default in natural communication is for humans to recount their acting upon other humans or—more typically— nonhuman objects; by extension, it is more natural for first and/or second persons to act upon third persons. Therefore, subjects typically correlate with positive semantic salience features, like being high in animacy, definiteness, and

specificity, whereas objects show the converse negative set, being typically low in animacy, definiteness, and specificity. The explanation for special marking of objects is in terms of a deviation from this more natural constellation, so that objects correlating with positive semantic saliency features tend to get marked off, iconically reflecting the markedness of conceptual event structure and its participants. Likewise, a first or second person object is unexpected, and although the effects of markedness reversals have been discussed mainly in regard to differential case marking (but see Schwenter [2006] on Spanish and Portuguese objects), similar considerations may be at work in the choice of first and second person pronouns for objects, marking their relative unexpectedness.

Yet another possible functional explanation could be in terms of what could be called "communicative salience," that is superordinate to discourse topicality: since all communication among human beings takes place as relevant for their experience, human referents are universally salient in discourse due to a construed equivalence with speaker and addressee. This could mean that even where fish and plants are the topics of descriptive texts, human beings are simply more salient globally. In other words, despite their topical status, fish and plants are not by themselves the communicative ends of these descriptions, and speakers are nonetheless more concerned and involved when it comes to humans, and in particular themselves or the Vera'a community, referred to by the first and second person pronouns. Appealing to communicative salience may also explain the differences between human and nonhuman discourse topics: although fish and plants follow the trend to be expressed by a pronoun when they are discourse topics, this effect may be much stronger with human discourse topics since these are of yet more concern for producer and audience of these texts than the fish and plant species under discussion.

Our analysis does not suggest any clear answer to these questions, and we do not commit to any of these fairly speculative proposals here. It is still possible that we are dealing with an idiosyncratic split between first and second person objects, and third person objects. But even under the latter hypothesis, considerations of markedness or communicative salience may nonetheless be involved in the diachronic development of such a person split system.

CONCLUSIONS

We conclude that discourse topicality is overall the decisive factor triggering the use of a pronoun in contrast to zero in third person objects in Vera'a. Crucially, discourse topicality can be shown to be a factor clearly distinct from semantic salience feature of high animacy. Given the notorious convergence of these two features, it seems methodologically highly relevant that we were able to tease them apart only through systematic investigation of a varied text corpus, taking into account differences in text registers, and the content and topics of classes of texts. In this sense, our study adds to the more recent findings of typological and theoretical studies in the differential treatment of objects that claim topicality to

be primary over semantic salience (Dalrymple & Nikolaeva, 2011; Iemmolo, 2010; Iemmolo & Klumpp, 2014 and contributions therein; Schwenter, 2006, 2014). However, while Dalrymple and Nikolaeva, for instance, identified sentence topicality as the relevant factor, which they saw as resulting from discourse topicality, our findings suggest discourse topicality as a factor in its own right, distinct not only from more or less correlating semantic features of animacy, but also from sentence topicality. Since we treat discourse topicality differently here, our findings are not immediately comparable to those of Schwenter (2006), who found a correlation between topic continuity in the sense of Givón (1983a) and the use of pronominal forms in objects in Portuguese and South American varieties of Spanish. However, it seems that our findings are sufficiently similar to his, and that both Schwenter's and our analysis corroborate the view that more global topicality on discourse level is more important a factor than sentence topicality or animacy.

Given the relatively high effort of our study in terms of corpus design and systematic analysis, it is not impossible that hitherto identified animacy (humanness) splits in other languages, and for other phenomena such as case-marking or agreement DOM, are indeed also a matter of discourse topicality, a possibility that could only be ascertained by extensive examination of the kind undertaken here.

The identification of discourse topicality as the best predictor for the use of object pronouns bears implications for diachronic accounts of the emergence of object agreement, since the increased use of pronouns rather than zero anaphor reflects an initial stage of such a development (Siewierska, 2004): According to Dalrymple and Nikolaeva (2011:18), grammaticalization of object agreement will start out from a tendency to mark topical nonsubjects, in particular objects, and later expand the use of pronominal agreement markers to nontopical objects that have features generally typical of topics, namely human or at least animate objects. The second possibility they mention is narrowing of topical marking, where marked objects are only those topical objects that bear specific semantic features. While we cannot provide a full-fledged alternative account at this point, it is clear that our findings do not square with either scenario: the use of pronouns in Vera'a encompasses all human objects, and discourse-topical nonhuman ones, but excludes sentence-topical ones where they are not human and not discourse topical. This means that the scenarios proposed by Dalrymple and Nikolaeva (2011) are at least not universally attested. Working out the relevance of this conclusion for accounts of grammaticalization of object agreement in Oceanic languages and beyond will be left for future research in this area.

Finally, our findings provide ample counterevidence to the universal relevance of accessibility and activation, suggesting that at least the choice between pronoun and zero for objects is not accountable for in terms of AT and similar frameworks concerned with discourse structure. From an activation point of view, these two forms of reference appear to be too similar to mark significant differences. As suggested already by Schwenter's (2006, 2014) work, it seems generally not unlikely that such similarity may result in operationalizing this choice for other

purposes, like the marking of discourse topicality, not only in Vera'a, but possibly in yet other languages with a choice between pronoun and zero for objects.

NOTES

1. Morpheme-by-morpheme glossing follows the Leipzig Glossing Rules (Comrie, Haspelmath, & Bickel, 2015); abbreviation for glosses are listed in the appendix.

2. The common article as well as two TAMP allomorphs are what we call detached clitics, a term adapted from Bickel and Nichols (2007:176). Thus, these are categorically unrestricted bound formatives that occupy the first syntactic slot in their functional phrase and attach phonologically to whatever word precedes them, similar to determiners, and subject and object markers in Kwakw'ala, discussed by Anderson (1992).

3. These constructions are nonfinite and do not take any TAMP marking. These properties are not further relevant for the purposes of the current paper. See Schnell (in press) for some discussion.

4. One anonymous reviewer remarks that zero as opposed to pronominal form of nonlexical objects may also be lexically determined, for instances with verbs like eat, drink, or read. According to our definition, zero objects are referentially zeroes—that is to say that there is a specific referent recoverable from the discourse context. A verb like eat in English does not permit a zero object of this kind. The alternation that the reviewer is alluding to is not treated as an instance of zero objects here, but as a constructional variation between a transitive and an intransitive frame that this verb is compatible with.

5. Note that nonhuman NPs are usually not plural-marked, and anaphoric expressions do not agree with them in number.

6. Thus, Neeleman and Szendröi (2007) correlated the parameter of pro-drop with the morphological makeup of pronominal forms across individual languages, and Bickel (2003) found a correlation between certain case-marking patterns in complement clause constructions and (higher) degrees of referential density across languages.

7. Our understanding of discourse topic is thus similar to Nichols' (1985) "theme," or what Lichtenberk (1996) captured under the label "thematic prominence." According to Lichtenberk (1996), a thematically most prominent discourse entity does not necessarily need to be mentioned more frequently than a less prominent one.

8. Note that the relativized function in Vera'a can be a pronoun or zero. There are neither specialized relative pronouns nor gapping as specific relativization strategies, hence the same alternation pertains here.

9. When collinear predictors are present in a regression model, because they are making similar contributions to model fit, it is difficult to determine the exact contribution per predictor. In a classification tree, a series of decisions is made to determine the best predictor for the first split, and then each partition of the data is looked at anew. Because of this, the presence of multicollinearity does not affect p-values, where the splitting in a tree occurs, nor the ranking of the strength of the predictors in the random forest analysis.

10. The classification tree split criterion was set at .95, resulting in splitting only when there are differences at the .5 p-value level. The tree was not pruned; all significant splits are shown.

11. Only 3 of 93 first or second person objects are zero in our corpus.

REFERENCES

Aissen, Judith. (2003). Differential object marking: Iconicity versus economy. *Natural Language and Linguistic Theory* 21(3):435–483.

Akaike, Hirotogu. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Cáski (eds.), *Second International Symposium on Information Theory*. Budapest: Akademiai Kaidó. 267–281.

Anderson, Stephen. (1992). *A-morphous morphology*. Cambridge: Cambridge University Press.

Ariel, Mira. (1988). Referring and accessibility. *Journal of Linguistics* 24(1):65–87.

———. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

———. (2000). The development of person agreement markers: From pronouns to higher accessibility markers. In M. Barlow & S. Kemmer (eds.), *Usage-based models of language*. Stanford: CSLI Publications. 197–260.

———. (2008). *Pragmatics and grammar*. Cambridge: 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, Maechler, 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.

Biber, Douglas, & Conrad, Susan. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.

Bickel, Balthasar. (2003). Referential density in discourse and syntactic typology. *Language* 79:708–736.

Bickel, Balthasar, & Nichols, Joanna. (2007). Inflectional morphology. In T. Shopen (ed.), *Language typology and syntactic description*. Vol. 3. *Grammatical categories and the lexicon*. Cambridge: Cambridge University Press. 169–240.

Burnham, Kenneth P., & Anderson, David R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. 2nd ed. New York: Springer.

Cameron, Richard, & Flores-Ferrán, Nydia. (2004). Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context* 1:41–65.

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.

———. (1994). *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: The University of Chicago Press.

Chung, Sandra. (1984). Identifiability and null objects in Chamorro. *Proceedings of the Annual Meeting of the Berkley Linguistic Society* 10:116–130.

Clancy, Patricia M. (1980). Referential choice in English and Japanese discourse. In W. Chafe (ed.), *The Pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood: ABLEX Publishing. 127–202.

Comrie, Bernard. (1989). *Language universals and linguistic typology*. 2nd ed. Chicago: The University of Chicago Press.

Comrie, Bernard, Haspelmath, Martin, & Bickel, Balthasar. (2015). *The Leipzig glossing rules: Conventions for interlinear morpheme-by-morpheme glosses*. Leipzig: Max-Planck Institute for Evolutionary Anthropology.

Corbett, Greville C. (2003). Agreement: Canonical instances and the extent of the phenomenon. In G. Booij, J. DeCesaris, A. Ralli, & S. Scalise (eds.), *Topics in morphology: Selected papers from the Third Mediterranean Morphology Meeting (Barcelona, Sep 20–22, 2001)*. Barcelona: Universitat Pompeu Fabra. 109–128.

Dalrymple, Mary, & Nikolaeva, Irina. (2011). *Objects and information structure*. Cambridge: Cambridge University Press.

Dowty, David. (1982). Grammatical relations and Montague grammar. In P. Jackson & G. K. Pullum (eds.), *The nature of syntactic representation*. Reidel: Dordrecht. 79–130.

DuBois, John W. (1987). The discourse basis of ergativity. *Language* 63:805–855. doi: 10.2307/415719.

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, & William J. Ashby (eds.), *Preferred argument structure: Grammar as architecture for function*. Amsterdam: Benjamins. 197–223.

Givón, Talmy. (1976). Topic, pronoun, and grammatical agreement. In C. N. Li (ed.), *Subjects and topics*. New York: Academic Press. 149–188.

———. (1983a). Topic continuity in discourse: An introduction. In T. Givón (ed.), *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam: John Benjamins. 1–41.

———. (ed.) (1983b). *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam: John Benjamins.

———. (ed.) (1998). The functional approach to language. In M. Tomasello (ed.), *The new psychology of language: Cognitive and functional approaches to language structure*. Hillsdale: Erlbaum. 38–62.

Haspelmath, Martin. (2013). Argument indexing: A conceptual framework for the syntactic status of bound person forms. In D. Bakker & M. Haspelmath (eds.), *Language across boundaries: Studies in memory of Anna Siewierska*. Berlin: De Gruyter Mouton. 197–226.

Hothorn, Torsten, Buehlmann, Peter, Dudoit, Sandrine, Molinaro, Annette, & Van Der Laan, Mark. (2006). Survival ensembles. *Biostatistcs* 7(3):355–373.

Hothorn, Torsten, & Zeileis, Achim. (2015). partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research* 16:3905–3909.

Huang, C.-T. James. (1984). On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15:531–574.

Huang, Yan. (1994). *The syntax and pragmatics of anaphora: A study with special reference to Chinese*. Cambridge: Cambridge University Press.

————. (2000). *Anaphora*. Cambridge: Cambridge University Press.

Iemmolo, Giorgio. (2010). Topicality and differential object marking: Evidence from Romance and beyond. *Studies in Language* 34:239–272.

Iemmolo, Giorgio, & Klumpp, Gerson. (2014). Introduction. In G. Iemmolo & G. Klumpp (eds.), *Differential object marking: Theoretical and empirical issues*. Special issue, *Journal of Linguistics* 52(2):271–279.

Krifka, Manfred. (2008). Basic notion of information structure. *Acta Linguistica Hungarica* 55 (3–4):243–276.

Kuznetsova, Alexandra. (2014). lmerTest: Tests in linear mixed effects models. Version 2.0-20. Package for *R*. Available at: http://cran.r-project.org/web/packages/lmerTest/index.html. Accessed January 15, 2015.

Lambrecht, Knud. (1994). *Information structure and sentence form: Topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Lichtenberk, František. (1996). Patterns of anaphora in To'aba'ita narrative discourse. In B. Fox (ed.), *Patterns of anaphora*. Amsterdam: John Benjamins. 381–411.

Malau, Catriona. (2016). *A grammar of Vurës, Vanuatu*. Boston: de Gruyter Mouton.

Meyerhoff, Miriam. (2002). Formal and cultural constraints on optional objects in Bislama. *Language Variation and Change* 14:323–346.

Neeleman, Ad, & Szendröi, Kriszta. (2007). Radical pro drop and the morphology of pronouns. *Linguistic Inquiry* 38(4):671–714.

Nichols, Joanna. (1985). The grammatical marking of theme in literary Russian. In M. S. Flier & R. D. Brecht (eds.), *Issues in Russian morphosyntax*. Columbus: Slavica Publishers. 170–186.

Poplack, Shana. (1980). The notion of the plural in Puerto Rican Spanish: Competing constraints on /s/ deletion. In W. Labov (ed.), *Locating language in time and space*. New York: Academic Press. 55–67.

Prince, Ellen. (1981). Towards a new taxonomy of given and new. In P. Cole (ed.), *Radical pragmatics*. New York: Academic Press. 223–255.

————. (1992). The ZPG letter: Subjects, definiteness, and information status. In S. A. Thomson & W. C. Mann (eds.), *Discourse description: Diverse linguistic analyses of a fund raising text*. Philadelphia: John Benjamins. 295–325.

R Core Team. (2015). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org/. Accessed July 20, 2015.

Rizzi, Luigi. (1986). Null objects in Italian and the theory of *pro*. *Linguistic Inquiry* 17:501–557.

Scherre, Maria M. P., & Naro, Anthony J. (1991). Marking in discourse: "Birds of a feather". *Language Variation and Change* 3(1):23–32.

Schnell, Stefan. (2011). *A grammar of Vera'a*. Kiel: Kiel University.

————. (in press). Whence subject-verb agreement? Investigating the role of topicality, accessibility, and frequency in Vera'a texts. *Linguistics* 56(4).

Schwenter, Scott A. (2006). Two kinds of differential object marking in Portuguese and Spanish. In P. Amaral & A. M. Carvalho (eds.), *Portuguese-Spanish interfaces: Diachrony, synchrony, and contact*. Amsterdam: John Benjamins. 238–260.

————. (2014). Null objects across South America. In T. L. Face & C. A. Klee (eds.), *Selected proceedings of the 8th Hispanic Linguistics Symposium*. Somerville: Cascadila Proceedings Project. 23–36.

Siewierska, Anna. (2004). *Person*. Oxford: Oxford University Press.

Strobl, Carolin, Boulesteix, Anne-Laure, Kneib, Thomas, Augustin, Thomas, & Zeileis, Achim. (2008). Conditional variable importance for random forests. *BMC Bioinfomatics* 9:307. Available at: http://www.biomedcentral.com/1471-2105/9/307. Accessed November 12, 2012.

Strobl, Carolin, Boulesteix, Anne-Laure, Zeileis, Achim, & Hothorn, Torsten. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* 8:25. Available at: http://www.biomedcentral.com/1471-2105/8/25. Accessed November 12, 2012.

Strobl, Carolin, Malley, James, & Tutz, Gerhard. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods* 14(4):323–348.

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.

Torres Cacoullos, R. & Travis, Catherine E. (2013). Prosody, priming and particular constructions: The patterning of English first-person singular subject expression in conversation. *Journal of Pragmatics* 63:19–34.

Travis, Catherine E. (2007). Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change* 19:101–135.

APPENDIX

Abbreviations

1	first person
2	second person
3	third person
A	a-form of demonstrative
ABIL	ability
ABL	ablative
AOR	aorist
ART	common article
AT	Accessibility Theory
DAT	dative
DEM	demonstrative
DISC	discourse particle
DISTR	distributive
DU	dual
EAT	eat possession
EMPH	emphatic particle
EX	exclusive
FUT	future
GEN	general possession
HOUSE	house possession
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
PROH	prohibitive
QUOT	quotative
RED	reduplication
REL	relativizer
SAP	speech-act participant
SG	singular
SP	specific
STAT	stative
TAMP	tense-aspect-mood-polarity
VAL	valuable possession
VC	verb complex

Logistic regression models

Logistic regression models using the lme4 (Bates, Maechler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, 2014) packages in R (R Core Team, 2015).

TABLE A1. *Results of mixed-effects generalized linear regression predicting animacy with discourse topicality*

Log likelihood: −129.5
AIC: 279.1
BIC: 320.2
Speaker variance: .03 ± .17
Discourse topicality by speaker variance: .00 ± .06

	Estimate	SE	Z	p
Intercept	4.34	.63	6.91	.00
Discourse topicality = +t	−4.27	.59	−7.27	.00
Form of the antecedent = pro	−1.45	.39	−3.69	.00
Form of the antecedent = zero	.27	.50	.54	.59
Function of antecedent = s	−1.14	.38	−3.00	.00
Function of antecedent = dislocated	−.20	.54	−.37	.71
Function of antecedent = other	−1.06	.55	−1.93	.05

Note: Negative coefficients are associated with higher pronoun expression. BIC, Bayesian information criterion.

TABLE A2. *Results of mixed-effects generalized linear regression with animacy*

Log likelihood: −150.6
AIC: 321.2
BIC: 362.3
Speaker variance: 1.58 ± 1.26
Animacy by speaker variance: 2.69 ± 1.64

	Estimate	SE	Z	p
Intercept	3.62	.63	5.70	.00
Animacy = +anim	−3.11	.68	−4.54	.00
Form of the antecedent = pro	−1.95	.37	−5.28	.00
Form of the antecedent = zero	−.02	.46	−.05	.96
Function of antecedent = s	−1.04	.36	−2.89	.00
Function of antecedent = dislocated	−.92	.53	−1.75	.08
Function of antecedent = other	−.81	.49	−1.64	.10

Note: Negative coefficients are associated with higher pronoun expression.