Towards subject-predicate agreement in Vera’a (Oceanic)

We investigate the choice between pronoun and zero anaphor for the expression of subjects in narrative texts from Vera’a (Oceanic; Schnell 2011, 2016). We relate this choice to a putative process of (re-)grammaticalisation of subject-predicate agreement typical of Oceanic languages (Ross 2004), whereby subject pronouns would eventually become the default referential choice, turning into agreement markers (Corbett 2003). This process interacts with older agreement morphology that has been reanalysed as exponents of tense, aspect, mood, polarity (TAMP) categories that have since deteriorated in morphological substance (François 2009). While 1st and 2nd person subjects are categorically expressed by a pronoun (Schnell 2018), we focus here on the alternation in 3rd person subjects. We exclude from this investigation full NPs, assuming that their use is relatively well explained in terms of accessibility theory (Ariel 1990) and related approaches to discourse structure.

Non-lexical 3rd person subjects are predominantly expressed by a pronoun too, but zero form is still attested in 26% of cases. A mixed-effect generalized linear regression model (Table 1) identifies six significant factors for the use of a pronoun, namely anaphoric distance and antecedent function, animacy and number, and form and agreement properties of co-present tense, aspect, mood, polarity (TAMP) morphology within the verbal predicate. Antecedent distance and function are relevant in restricting zero subjects largely to same-subject clause chains. Within this context, we find zero subjects to be most likely with singular subjects and prospective aspect, where subject agreement is co-expressed (unlike in any other TAMP category) by a formally unreduced marker, ex (1), in contrast to the respective plural form, ex (2), and many other high-frequency TAMP markers, ex (3).

Although our synchronic corpus investigation cannot provide any conclusive evidence for or against specific diachronic developments, our findings do square with the hypothesis of re-grammaticalizing subject-predicate agreement, so that pronouns are practically the default non-lexical form of expression, leaving zero anaphor only for those context where older agreement is co-present and formally fully transparent. Our findings suggest that a putative process of re-grammaticalizing subject agreement is spurred primarily by purely morphological changes (Barth & Kapatsinski 2017; Bybee & Thompson 2007) rather than functionally motivated (Givón 1976; Ariel 2000).
EXAMPLES

(1)  
gōsuwō  ne    kalraka    ne    rōw  lē =n    nanara  lumasag [...]  

rat    PROSP:3SG    get.up    PROSP:3SG    jump    LOC=ART    tree.sp    ontop

ne    PROSP:3SG    jump    ontop    high    PROSP:3SG    sit    DEL

‘Rat got up, jumped ontop of the nanara (trunk), jumped up ontop and sat down there.’

GAQG.028

(2)  
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

(3)  
di =m    sag    sur    suwō    di =m    kur  sa  =n    gako  wova’al ē

3SG    =PRF    sit    down    downwards    3SG    =PRF    gnaw    EMPH    =ART    stalk    pawpaw    DEM3

‘He sat down and gnawed (through) that very pawpaw stalk.’

GAQG.078

TABLE 1. Results of mixed-effects generalized linear regression

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.17</td>
<td>0.41</td>
<td>-0.42</td>
<td>0.67</td>
</tr>
<tr>
<td>Antecedent distance = 2+ clauses</td>
<td>1.74</td>
<td>0.31</td>
<td>5.63</td>
<td>0.00</td>
</tr>
<tr>
<td>Function of antecedent = object</td>
<td>0.10</td>
<td>0.37</td>
<td>0.27</td>
<td>0.79</td>
</tr>
<tr>
<td>Function of antecedent = other</td>
<td>1.51</td>
<td>0.33</td>
<td>4.55</td>
<td>0.00</td>
</tr>
<tr>
<td>Number = non-singular</td>
<td>1.25</td>
<td>0.29</td>
<td>4.35</td>
<td>0.00</td>
</tr>
<tr>
<td>Animacy = human</td>
<td>-0.83</td>
<td>0.34</td>
<td>-2.45</td>
<td>0.01</td>
</tr>
<tr>
<td>Animacy = inanimate</td>
<td>0.69</td>
<td>0.30</td>
<td>2.28</td>
<td>0.02</td>
</tr>
<tr>
<td>TAMP form = particle</td>
<td>0.54</td>
<td>0.22</td>
<td>2.44</td>
<td>0.01</td>
</tr>
<tr>
<td>TAMP person = no inflection</td>
<td>-1.24</td>
<td>0.37</td>
<td>-3.31</td>
<td>0.00</td>
</tr>
<tr>
<td>TAMP person marking by TAMP form = no inflection by particles</td>
<td>1.42</td>
<td>0.46</td>
<td>3.07</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Note: Positive coefficients are associated with higher pronoun expression

REFERENCES

Barth, Danielle & Vsevolod Kapatsinski. (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.


