Variation and grammaticisation: The emergence of an aspectual opposition

Rena Torres Cacoullos
University of New Mexico

Using multivariate analysis, this study tracks the configuration of factors conditioning variation between the Spanish Progressive (estar “to be (located)” + Gerund (Verb-ndo)) and the simple Present, in 15th, 17th, and 19th century data. While the direction of effect remains stable, change is manifested in shifts in magnitude of effect. The Progressive begins as a locative construction with a presentative function, as shown by strong early effects of co-occurring locatives and postverbal full NP subjects. Over time locative meaning weakens and aspectual meaning, which is an implication of the locative construction, strengthens. The aspectual opposition between the Progressive and simple Present then extends to stative predicates. Thus, the changing relative magnitude of effect of the factor groups in the multivariate analyses shows the gradual emergence of the progressive – nonprogressive opposition. In grammaticisation, new constructions gradually evolve from patterns of variation, in the course of speakers’ recurrent choices in discourse.

Keywords: Variation, grammaticisation, progressive aspect, stativity, neutralization in discourse, multivariate analysis, Spanish.

1. Variation and grammaticisation of progressives

The variationist focus on language change as the spread of new usage across groups of speakers (cf. Weinreich, Labov & Herzog 1968) has been criticised for neglecting innovation. For example, Croft (2000:54), upholding a distinction between innovation and propagation in language change, poses the problem thus:

The central problem of the sociolinguistic theory of language change, for all its empirical success in explaining propagation, is that it always presupposes the existence of multiple variants. […] the question remains how specific variants arise and become part of the variable linguistic system.

(my underlining)

In addressing this question, speaker motivations for innovation commonly put forward are various definitions of expressiveness and interactional maxims; for example, linguistic change has been said to arise (non-teleologically) as speakers talk in such ways, for example, as to draw attention (cf. Keller 1990/1994:101, Haspelmath 1999:1057, Croft 2000:74). This study of the Spanish progressive examines how variants arise in grammaticisation.

The distinction between progressive and nonprogressive forms is obligatory in English whereas in Spanish use of the progressive form is optional (Comrie 1976:33; but see Torres Cacoullos 2000) and some languages, including earlier varieties of English and Spanish, do without a progressive form altogether. How do grammars get progressives? In usage-based theory, the major source for new grammatical morphemes is grammaticisation, whereby existing constructions with particular lexical items gain frequency and become new constructions, following cross-linguistic evolutionary paths (Bybee 2006:719-721). Thus, the Progressive (PROG), composed of estar “be (located)” and a verb in gerund (-ndo) form, has developed from a locative, as in the 15th c. (Old Spanish) example in (1), to a progressive
construction, as in the 20th c. example in (2a) (Bybee, Perkins & Pagliuca 1994: 127-133, Torres Cacoullos 2000:71-88).1 In contrast to the Progressive in (2a), the simple Present (PRES) in (2b) expresses habitual aspect.

(1) unos están en su casa folgando  
“people are in their home resting” (PROG)  

(15th c., Corbacho, IV/I)

(2) a. […] ¡¿Está durmiendo Pablo?!  
“[…] Pablo is sleeping (PROG)⁈”  

“[…] Pablo – after eating he goes to sleep (PRES) […]”  

(20th c., COREC, ACON023A)

Grammaticisation studies have examined the structural and semantic changes accompanying new constructions, including phonological reduction with changes in constituency and loss of specific lexical meaning with conventionalisation of contextual implications (e.g. Hopper & Traugott 2003). Less studied are changes in the patterns of variation with existing grammatical expressions, in this case the variation between the Progressive and the simple Present, as in (3), where both variants, está devaneando “is raving” and devanea “raves”, express a situation in progress at speech time.

(3) a. Está devaneando entre sueños.  
“He is raving (PROG) in his sleep.”  

(15th c., Celestina, VIII)

b. Hijo, déxala dezir, que devanea;  
“Son, let her talk, she is raving;” (PRES) [literally: raves]  

(15th c., Celestina, IX)

A fundamental principle for the empirical foundations of a theory of language change is “normal heterogeneity”: in a speech community there are normally “alternate […] ways of saying ‘the same thing’” (Labov 1982: 17-18). In the study of grammaticisation variability has been recognised as “layering,” the availability of different constructions to serve “similar or even identical functions” as newer layers emerge without displacing older ones within a functional domain (Hopper 1991:22-24). The notion of layering privileges a view of variation as the result of grammaticisation. Here we adopt the view that inherent variation contributes to grammaticisation rather than merely resulting from it.

Grammar arises from language use, as “structure emerges from the repetition of many local events” (Bybee 2006:714). The “local events” that we will examine are speakers’ choices in discourse between the two variants, the Progressive and the simple Present illustrated in (3). We discover the patterns of these recurrent choices by multivariate analysis (Sankoff 1988). In underscoring the gradualness of change, the results support the variationist

1 I capitalise the first letter of language-specific forms, to distinguish them from cross-linguistic categories or functions, following Comrie (1976: 10).
hypothesis of change through neutralisation in discourse of functional distinctions among different constructions (Sankoff 1988): in the course of the recurrent choices which speakers make between different ways of saying the same thing, the variants themselves evolve.

2. Data

We consider data from three time periods, the 12th-15th (Old Spanish), 17th, and 19th centuries, from different genres including chronicles, didactic stories, novels, and plays, mostly in prose, yielding a corpus of approximately 1.5 million words for Old Spanish and 500,000-word corpora for the 17th and 19th centuries each. Progressive Present occurrences (N=374) were exhaustively extracted; the normalised token frequency ranges from 1 occurrence per 10,000 words in the 15th to 2 in the 17th and 4 in the 19th century. A sample (N=1648) of the much more frequent simple Present was taken by extracting two tokens in the vicinity of each Progressive, usually the one immediately preceding and following, and Present occurrences of lexical types appearing in the Progressive in a given text.

The variable context – the context(s) in which speakers have a choice of forms – is defined broadly here as present temporal reference (cf. Walker 2001:14-16). Excluded from the quantitative analyses were occurrences of the simple Present with future or past reference, modal periphrases (for example, poder “can”-plus-Infinitive), proverbs, and constructions that appear invariably with the simple Present, including discourse routines in the first or second person, for example, digo “I say”, ya ve(s) “you see”.

3. Hypotheses and coding of tokens

All Progressive and simple Present tokens were coded for features of the linguistic environment in factor groups, or independent variables, which operationalise hypotheses about variant choice. These are stativity, aspect, co-occurring locatives, co-occurring temporal adverbials, subject form and position, transitivity, and polarity-mode (lexical frequency was included but will not be discussed here).

Stativity tests the hypothesis that progressives deriving from a locative source are first restricted to dynamic predicates (e.g. Bybee et al 1994:133, 139-140; cf. Comrie 1976:35). The prediction is that statives such as “fear” (4b, 5b) will disfavour the Progressive construction more strongly in earlier than in later periods. Aspect (Comrie 1976:3) was coded independently of the inherent aspextual properties of dynamic and stative verbs (Comrie 1976:41) by distinguishing between situations of unlimited and limited duration. “Unlimited duration” subsumes habitual aspect, which describes customarily repeated situations (cf. Comrie 1976:27-8), as in (4a), and states without temporal limits, which begin before speech time and continue indefinitely, as in (4b). In contrast, “limited duration” includes progressive actions occurring simultaneously with speech time (5a) and states circumscribed to a period near speech time (5b); also included is continuous aspect (Comrie 1976:26, 33).

(4) a. Que a quien más quieren, peor hablan
   “To those they love best they speak (PRES) worst”
   (15th c., Celestina, XI)
b. cómo, siendo el principio de la sabiduría el temor de Dios, tú, que temes más a un lagarto que a Él, sabes tanto

“how, fear of God being the beginning of wisdom, you, who fears (PRES) a lizard more than Him, know so much”

(17th c., Quijote II, XX)

(5) a. escucha, que hablan quedito

“listen, they are speaking (PRES) softly”

(15th c., Celestina, XII)

b. estoy temiendo que, entre la lengua y la garganta, se ha de atravesar el riguroso cordel que me amenaza

“I fear (PROG) [literally: am fearing] that the severe noose threatening me must tighten between my tongue and my throat”

(17th c., Quijote II, LXIII)

In grammaticsation theory grammatical morphemes have inherent semantic content deriving from their lexical source construction (Bybee et al 1994:138,148) (whereas in structuralist approaches grammatical morphemes form maximal contrasts based on the presence versus absence of a semantic feature). The prediction following from the hypothesis of retention (Bybee et al 1994:16) or persistence (Hopper 1991) of locative meaning is that a co-occurring locative, such as en la sala “in the drawing room” in (6), will favour choice of the Progressive over the simple Present. Furthermore, if new grammatical resources express more specific meanings than those of existing grammatical morphemes (Bybee et al 1994:133), co-occurring temporal expressions, such as ya “already, now” in (6), should also favour the Progressive.

(6) de boda estoy vestida; ya me están aguardando en la sala don Fernando [] y mi padre [], con otros testigos

“I’m dressed to be married; don Fernando and my father are already waiting (PROG) for me in the drawing room, with other witnesses”

(17th c., Quijote I, XXVII)

Two more factor groups operationalise retention and the richer meaning of the newer construction. Consonant with the locative origins of the Progressive is a presentative function, since locative expressions, like motion verbs, can be used to introduce a new referent (cf. Bentivoglio 1993:222, n.9, Dumont 2006:295). The prediction following from the hypothesis of a presentative function is that the Progressive should be favoured by subjects that are full NPs (the form in which new-information mentions are typically introduced (Bentivoglio 1993)) and appear postverbally (the position in which new lexical NP referents are introduced (Ocampo 1993)), for example don Fernando in (6). From the original function of giving the location of the subject, the construction may also imply that ‘the subject is in the midst of doing something’ (at that location), which conveys the active involvement of the subject in the activity (Bybee et al 1994:133-137). The notion of subject involvement is operationalised with transitivity, since transitive predicates are more likely to be under the control of the subject (Hopper & Thompson 1980:265), with the prediction that the presence of an object will favour the Progressive more than intransitive uses.
Finally, in considering *polarity and mode*, we predict a disfavouring effect of negatives and interrogatives, which tend to be conservative contexts (Givón 1979:121; Torres Cacoullos 2000:184). The multivariate analyses ahead confirm a consistent constraint on Progressive and simple Present variation. Under negation, situations in progress appear in the simple Present, as in (7), as do questions about situations in progress such as the formulaic ¿qué haces? “what are you doing?” [literally: what do you do?].

(7) […] no es posible sino que adivinase lo que nos está sucediendo.
   Pero, señora, si no sucede nada.
   “It is not possible but that he should guess what is happening (PROG) to us.”
   “But, madam, nothing is happening (PRES).” [literally: happens]
   (19th c., Sí de las niñas, I/V)

4. Tracking the conditioning of variation over time

We are interested in two lines of evidence from the multivariate analyses, the direction of effect and the strength of effect (cf. Tagliamonte 2006:235-245). First, the *direction of effect* is instantiated in the order of the factors within a factor group from more to less favourable as indicated by the probability or factor weights, which are assigned such that the closer to 1, the more likely, the closer to 0, the less likely that the Progressive will be chosen in the given environment. Second, a factor group may not achieve significance when considered simultaneously with other factor groups; for the significant factor groups, relative *magnitude of effect* is indicated by the *range*, the difference between the highest and lowest factor weight in the group. Table 1 shows the results of three independent variable rule analyses (Sankoff, Tagliamonte & Smith 2005) of the Old Spanish, the 17th c. and the 19th c. data.

The strong favouring effect of a co-occurring locative (.72) in the Old Spanish data supports the hypothesis of the locative origins of the Progressive. Consonant with these origins is the presentative function indicated by the favouring effect of postverbal full NP subjects (.70) and the non-aspectual nuance of a higher degree of subject involvement indicated by the transitivity effect (.61). As anticipated, negative polarity and interrogative mode constitute unfavourable contexts. Finally, an aspectual effect is in place with situations of limited duration favouring the Progressive (.62). Note, though, that its magnitude is not greater than that of the other factor groups, as indicated by the close ranges (22-24).

While there is no change in the direction of effect, there are clear shifts in magnitude of effect in the 17th century. First, aspect is evidently the most important linguistic feature in the variation, with a range which is 50% (65:42 = 1.5 times) greater than the next largest, that of the polarity-mode factor group. Second, stativity now appears as a statistically significant constraint. Third, neither subject form and position nor transitivity is significant and the locative factor group is relegated to fourth position, dropping to non significance in the 19th century. In this third period the range of the aspect factor group is 80% (1.8 times) greater than the next largest and, furthermore, the temporal adverbial effect achieves significance for the first time.
Table 1. Factors contributing to the choice of the Progressive over the simple Present
(Non-significant factors within [ ])

<table>
<thead>
<tr>
<th></th>
<th>Old Spanish (N = 82/493)</th>
<th>17th century (N = 107/676)</th>
<th>19th century (N = 185/853)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locative co-occurrence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>.72 42% 36</td>
<td>.71 43% 47</td>
<td>[.56] 33% 69</td>
</tr>
<tr>
<td>Absent</td>
<td>.48 15% 457</td>
<td>.48 14% 622</td>
<td>[.49] 21% 767</td>
</tr>
<tr>
<td>Range</td>
<td>24</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Aspect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited duration</td>
<td>.62 28% 152</td>
<td>.79 33% 269</td>
<td>.69 41% 391</td>
</tr>
<tr>
<td>Extended duration</td>
<td>.38 13% 152</td>
<td>.14 2% 191</td>
<td>.17 6% 192</td>
</tr>
<tr>
<td>Range</td>
<td>24</td>
<td>65</td>
<td>52</td>
</tr>
<tr>
<td>Polarity - Mode</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affirmative declarative</td>
<td>.54 19% 410</td>
<td>.57 18% 577</td>
<td>.56 25% 684</td>
</tr>
<tr>
<td>Negative, Interrogative</td>
<td>.31 7% 83</td>
<td>.15 2% 99</td>
<td>.27 10% 169</td>
</tr>
<tr>
<td>Range</td>
<td>23</td>
<td>42</td>
<td>29</td>
</tr>
<tr>
<td>Subject form and position</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postverbal full NP</td>
<td>.70 32% 41</td>
<td>[.60] 20% 46</td>
<td>[.60] 36% 58</td>
</tr>
<tr>
<td>All others</td>
<td>.48 15% 452</td>
<td>[.49] 16% 630</td>
<td>[.49] 21% 795</td>
</tr>
<tr>
<td>Range</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transitivity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transitive</td>
<td>.61 27% 134</td>
<td>[.55] 19% 203</td>
<td>[.52] 23% 339</td>
</tr>
<tr>
<td>Intransitive</td>
<td>.45 14% 290</td>
<td>[.47] 14% 378</td>
<td>[.49] 21% 407</td>
</tr>
<tr>
<td>Range</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stativity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamic predicate</td>
<td>[.53] 19% 378</td>
<td>.60 20% 475</td>
<td>.56 26% 636</td>
</tr>
<tr>
<td>Stative predicate</td>
<td>[.40] 10% 115</td>
<td>.27 6% 201</td>
<td>.33 8% 217</td>
</tr>
<tr>
<td>Range</td>
<td>33</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>Temporal co-occurrence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>[.61] 21% 62</td>
<td>[.59] 23% 83</td>
<td>.64 30% 116</td>
</tr>
<tr>
<td>Absent</td>
<td>[.48] 16% 427</td>
<td>[.49] 15% 590</td>
<td>.48 21% 727</td>
</tr>
<tr>
<td>Range</td>
<td>16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Also included: Lexical frequency (significant in all analyses; low frequency verb types favour the Progressive)

Figure 1 summarises the changes. Polarity-mode persists as a constraint throughout the three time periods, with negatives and interrogatives disfavouring the Progressive. Weakening over time are the effects of co-occurring locative, subject form and position, as well as transitivity, which operationalise meanings that follow from the locative origins of the construction, as predicted. Loss of specific features of meaning, or semantic bleaching, of a construction and concomitant generalisation of the contexts in which it is used is exactly what we expect in grammaticisation (Bybee et al 1994:6). On the other hand, the strengthening of the effects of aspect and co-occurring temporal adverbial indicates the gradual emergence of an aspectual opposition between the Progressive and simple Present: the Progressive increasingly indicates limited duration (progressive, continuous) in contrast to the extended duration meaning of the simple Present (habitual, existing state).

Polarity, mode = Persists
Co-occurring locative ← Weakens
Subject expression and position ← Weakens
Transitivity ← Weakens
Aspect → Strengthens
Co-occurring temporal → Strengthens
Stativity → Strengthens (between Old Span and 17th c.)

Figure 1. Changes in magnitude of effect of factor groups (linguistic environmental features) conditioning Progressive vs. simple Present variation
5. Stativity and progressivity

How does this emerging distinction between progressive and nonprogressive forms interact with stativity? The prediction was that stativity as a constraint would weaken over time as the Progressive generalises to stative predicates, but the stativity effect initially strengthens (between Old Spanish and the 17th century; cf. Figure 1).

The Progressive was in fact used early on with statives, albeit less than with dynamic predicates, but more as a locative construction, as illustrated in (8).

(8) a. les semeia q<ue> esta alli callando todo.
   “it seems to them that everything is there being quiet. (PROG)”
   (13th c., GE.I, fol. 126r)

b. Y en tierra está adorando a la más antigua [y] puta tierra
   “He is on his knees worshipping (PROG) the oldest whore in town”
   (15th c., Celestina, I)

The locative meaning of the construction with statives is confirmed in separate analyses of stative verbs, in which the one significant factor group in the Old Spanish data is that of co-occurring locative.

As locative meaning bleaches and the construction is increasingly associated with progressive aspect, the restriction on use with statives strengthens, achieving statistical significance in the 17th century (Table 1). That is, the restriction on use with statives develops as the aspectual function of the construction emerges. In the 17th century the Progressive is largely restricted to dynamic verbs (85% (78/92) of all Progressive tokens coded for aspect are of dynamic verbs in limited duration contexts).

The aspectual opposition first conventionalised with dynamic predicates then begins to generalise to statives, as illustrated in (9): the Progressive expresses a state circumscribed to speech time. In a separate analysis of 19th century statives, aspect and co-occurring temporal adverbial are now the two significant factor groups.

(9) Se me está figurando ahora que no soy tan malo como yo mismo creo.
   “It seems (PROG) [literally: is seeming] to me now that I am not as bad as myself believe.”
   (19th c., Doña Perfecta, XXVIII)

Thus, it is not so much that the Progressive construction extends to statives but rather, the aspectual opposition between the Progressive and the simple Present extends to statives.

6. Conclusion

Labov (1982: 84) concludes by urging an “evolutionary perspective”, reminding us that “we have indeed come to understand a thing when we understand how it came to be.” In tracking the configuration of factors conditioning variation between the Progressive and
simple Present we showed that while the direction of the effect remains stable, change is manifested in shifts in magnitude of effect. The Progressive begins as a locative construction with a presentative function, as shown by the strong early effects of co-occurring locatives and postverbal full NP subjects. Aspectual meaning is part of the locative construction, following from the implication that the subject is located somewhere doing something. Over time, as locative meaning weakens aspectual meaning strengthens. The changing relative magnitude of effect of the factor groups in the multivariate analyses shows the gradual emergence of the progressive – nonprogressive distinction.

Thus, the answer to the question of how “variants arise” (Croft 2000:54) in grammaticisation is given by the variationist hypothesis of neutralisation in discourse (Sankoff 1988:153-154), which states that while contexts can almost always be found in which different constructions have different meanings, there are alternations in which the full accompaniment of meaning distinctions is not pertinent for either the speaker or the interlocutor. As Sankoff (1988: 153) states, neutralisation-in-discourse “is the fundamental discursive mechanism of (nonphonological) variation and change”.

Innovation may well be linguists’ post hoc depiction of a gradual process. In grammaticisation at least, there is no ostensive innovation, since the variants are emergent from patterns of variation. In this study we have seen that in the course of speakers’ recurrent choices in discourse between alternate ways of saying the same thing, which are the repeated local events that define the structure of the variation, the variants themselves evolve.

References


