(Dis)continuity in language change: *ser* and *estar* + age in Latin-American Spanish

Bob de Jonge

0. Introduction

Until now, scholars basically have assumed that language change is a gradual process, in which a new linguistic unity penetrates in a certain context and from there, gradually expands until the new form has expelled the old one completely (see for instance Kroch 1989:199-202). As also stated in Bailey (1973:77), in Allan (1987:141-2) it is pointed out that

> Lexical diffusion changes are divided into three stages: an initial stage where only a few items change; a middle stage where the rate of change accelerates rapidly (this is the most productive period); and a late stage where the rate of change recedes and eventually ceases, leaving some items unaffected by the change. When the rate of change is plotted against time the result is an S shaped graph:
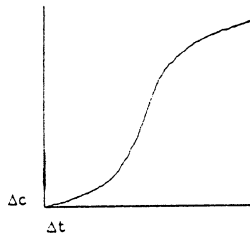

Figure 1: Representation of language change after Allan (1987:142)

Radford (1976), who discusses a change in causal structures in Romance languages over the last centuries, speaks in terms of a 'gradual historical process of 'fusion'', a 'gradual tendency towards increasing *matiness*[1]', and finally an 'ongoing process of fusion' (1976:92-94); all these descriptions match the flowing

---

[1] Described by Radford as 'degree of belongingness to the same constituent' (1976:91).

character that is attributed to language change in the existing literature on the subject.

In many processes of language change this image appears to be an adequate representation of events. The appearance of the particle *otros* 'others' with *nos* 'we' and *vos* 'you (pl)' in Medieval Spanish, leading to the modern forms *nosotros* 'we' and *vosotros* 'you (pl)' was such a gradual process that took place between the 13th and 16th centuries (García et al. 1990:63-101). The loss of *v*-with the oblique form *vos* 'you (dative/accusative, sg and pl)', on the other hand, which cannot be seen as a phonetic change, but has to be viewed as an optimalisation of the pronominal paradigm of Spanish, can hardly be represented as a process with a gradual character: the texts that show variation can all be dated around 1500 (García et al. 1990: 112-113). This gives rise to a suspicion that in the case of the oblique form *vos*, it should be viewed as a sudden switch from one form to the other, and not as a gradual process, a suspicion that is confirmed by the apparent randomness of the distribution of the alternating forms during the process of change.

In addition, we are confronted with problems of language change in which the process takes place in different stages and every now and then stops, sometimes for long periods of time. For example, in Pountain (1982) it is described how in different Romance languages the verb *stare* expels the verb **essere* in a number of specific contexts. In early Medieval Spanish, *estar* takes over the most important locative uses of *ser*, for instance when the subject is animate (1982:155). In the 16th century the use of *estar* with past participles is becoming general (1982:156), again at the expense of *ser*, after which time its use with adjectives also becomes more common. Hengeveld (1991) shows this image of different stages in the expansion of *estar* over different functions of *ser* even more clearly: he distinguishes, on the basis of language universals observed in a number of genetically different languages, a hierarchy of four types of contexts, namely locatives, adjectives, nouns and possessive contructions. In the Ibero-Romance languages, *estar* first encroached on *ser* in locative contexts, then to some extent on adjectives and finally, for example in Portuguese, *estar* invaded even the context with noun predicates. All of these changes took place without, however, putting a complete ban on the uses of *ser* in the particular contexts.

From a synchronic point of view, one could (and still can) speak of variation in the uses of *ser* and *estar* within the contexts discussed above. However, from a diachronic standpoint we also have to conclude that for example in the use of *ser* and *estar* with adjectives we can speak of a relatively stable situation: only slight changes can be observed during the last centuries, whereas the data presented by Pountain (1982) and Hengeveld (1991), who observe important shifts on a macro-historical scale, seem to suggest that we are in the middle of a process of language change. In this article I will focus on a particular context in the uses of

ser and *estar* with adjectives and see how diachrony is reflected in the synchronic states of two varieties of Latin-American Spanish.

1. An example

In a specific group of adjectives a stable situation of alternation between *ser* and *estar* has been observed, namely those that indicate age (cf. De Jonge 1987). In these constructions, as in

(1)   Juan *es/está* joven/viejo.
       'John *is* [*ser/estar*] young/old'

both in Mexico City Spanish as in Caracas Spanish frequently the use of *estar* is observed, whereas in Peninsular Spanish *ser* would be normal. If the uses in Mexico City are compared with those in Caracas, however, it appears that these two differ considerably as well. In table 1, the distribution of both copular verbs in age expressions are compared for two equivalent corpora of educated speakers from Mexico City and Caracas, the México Culto (MC) corpus and the Caracas Culto (CC) corpus.[2][3]

Elsewhere (De Jonge 1991a) it has been argued that the observed difference in overall use of *estar* (vs. *ser*) is the result of a difference in the stage of process of language change in which Mexican and Caracas Spanish find themselves respectively, an argumentation that is supported by quantitative data.

---

[2] The data of the corpora have been collected from Lope Blanch (1971) and Rosenblat (1979), and personal consultation of unpublished material at the Centro de Estudios Hispánicos of the Universidad Nacional Autónoma de México (UNAM) and in the Instituto de Filología Andrés Bello (IFAB) of the Universidad Central de Venezuela (UCV). I am indebted to these institutions for their generous help in allowing me to consult these materials. The consultation of the unpublished parts was made possible thanks to grants awarded by the Netherlands Foundation for the Advancement of Tropical Research (WOTRO) and the Ministry of Education.

[3] In the left top cell of the tables presented in this article the results of some statistical tests are given in order to indicate the significance and strength of the presented skewings. The X -coefficient and corresponding probability (p<...) indicate the probability that the presented skewing is the result of chance; a p<0,05 is generally considered to be statistically significant (cf. Butler 1985:112-127). The Odds-Ratio is an indication of the strength of a skewing without taking into account the influence of total numbers (N). An odds-ratio of 1 indicates that there is no skewing at all, the higher the ratio, the stronger the skewing (cf. Kruskal & Tanur 1978:1133). Since to my knowledge there is no general standard as to the lowest level of interesting odds-ratio's, I take 2 as the lowest odds-ratio of interest, as a result of my own experience.

Table 1: distribution of ser vs. estar in age expressions in México Culto and Caracas Culto

X =10,5; p<0,01 Odds Ratio =2,0	Total N	*ser*	*estar*
MC	159	64%	36%
CC	192	47%	53%

It becomes even more interesting when it appears that, apart from the observation that both communities appear to represent different stages in the development process, both stages most probably have remained unaltered during the last 10 to 20 years. The data for MC were gathered by the end of the sixties and, according to the results of a questionnaire submitted in 1989, have remained virtually unchanged ever since.[4] [5] The data for CC were gathered in the beginning of the seventies, and comparison with data from the eighties, although not absolutely comparable, makes it highly plausible that here no changes have taken place as well. The results of this comparison are given in table 2.[6]

As can be expected, the presented skewing is statistically insignificant (p<0,5) and the odds-ratio close to 1 indicates that the observed difference of 5% in the skewing is not to be given much importance.[7]

---

[4] The questionnaire was submitted to 60 Mexican students of the Universidad Nacional Autónoma de México, divided into two groups, in order to check the parameters investigated in De Jonge (1990). In the questionnaire, 15 stimuli are comparable to the data presented in this article, which rendered 433 valid answers. Of these, 124 contain *estar* (29%), and the other 309 contain *ser* (71%). Moreover, the only statistically significant parameter turns out to be the one presented in table 3 below. For full details of the results, see De Jonge (1990:169-172). Because of the inherent incomparability of the results of the questionnaire and the data presented in this article, I do not present them together in a table nor do I apply any statistical test. The questionnaire data will not be used in the rest of this article; with the exception of table 2, in all other tables only data from the corpora as indicated in fn. 2 are presented.

[5] It is worth while mentioning in this context that in Juan Rulfo's *Pedro Páramo* y *El Llano en Llamas*, a Mexican novel and a collection of short stories of which the first editions appeared in 1955 and 1953 respectively, in 44% of a total of 25 age expressions *estar* is used.

[6] The data for the corpus of the eighties were provided by three MA-students of the Instituto de Filología Andrés Bello of the Universidad Central de Venezuela. I am indebted to Irania Malaver, one of the students, and to Paola Bentivoglio for giving me access to these data.

[7] And what is more, this difference goes counter to the general prediction that *estar* would become more frequent over time!

Table 2: distribution of ser *vs.* estar *in two not contemporary Caracas corpora.*

X =0,7; p<0,5 Odds Ratio =1,2	Total N	*ser*	*estar*
CC (early 70s)	192	47%	53%
C (late 80s)	92	52%	48%

If we, notwithstanding these data, maintain the (according to Pountain (1982) and Hengeveld (1991) correct) presupposition that the use of *ser* and *estar* finds itself in the middle of a process of language change, the following questions arise:

i)   How can it be that a process of language change comes to a complete standstill for a shorter or longer period?
ii) If the process of language change does come to a standstill, can we still speak of a 'gradual' process, or should this vision be adjusted?

2. A new hypothesis for language change.

From the preceding it may become clear that the supposed gradualness of language change is to be doubted. Thus I propose the following:

Hypothesis: *The process of language change should not be represented by an S-shaped graph which is of a basically gradual nature, but as a staircase-formed graph that is characterised by horizontal and vertical phases.*

In figure 2 this hypothesis is rendered graphically.
    In this graph I have attempted to render the following. According to the proposed hypothesis, a new linguistic form 'conquers' pieces of territory of the older form on various different occasions in the process of language change. These conquests of pieces of territory are represented in the graph as the vertical lines and each one of them should be interpreted as a conquest of one particular type of context. Each piece of territory is conquered in a relatively short time, but the amount of time that passes between different conquests may vary. These varying periods of time are represented by the horizontal lines, which therefore have different lengths in the graph. In this period of time, a clue has to be found in order to conquer the next piece of territory/type of context. If a clue has been found, nothing can stop the conquest of the next group of contexts, which will then take place within a relatively short period of time.
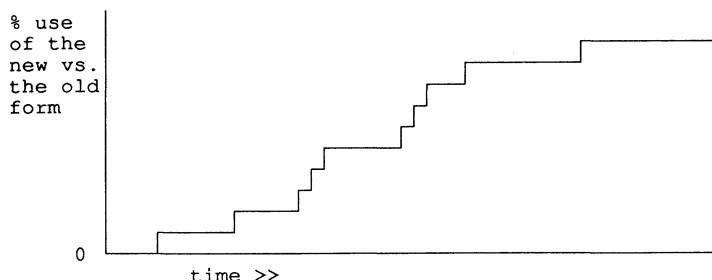
```
% use
of the
new vs.
the old
form



   0
        time >>
```

Figure 2: graphic representation of the proposed hypothesis on language change.

I shall try to illustrate how this process takes place by means of the innovative use of *estar* at the cost of *ser* in age expressions, as rendered in (1).

3. The data.

In De Jonge (1987) it has been argued that the innovative use of *estar* in age expressions is a consequence of the chronological character of some of these age expressions. In other words, the process of growing older is relevant for the wider context. An example of this use is shown in (2) (Lope Blanch 1971:316):

(2)   Enc. - [...] es muy fino su...
       A. - Sí.
       Enc. - ... su teatro.
       B. - Bueno, sí, pero se me hace que ella ya *está* grande para determinados papeles, como que no le quedan.
       'Int. - [...] it is very elegant, her...
       A. - Yes.
       Int. - ... her way of acting.
       B. - Well, yes, but it seems to me that she *is* [*estar*] already old for certain parts, as if they don't fit her.'

In accordance with Falk (1979), the relevance of the circumstances to explain the use of *estar* is very clear: the informant is not stating that she is just old, but says that her advanced age makes it impossible for her to play certain parts. Therefore, it is not surprising that *estar* has entered this type of age expression.

   In De Jonge (1991b) it is argued that linguistic innovations develop as a result of pragmatic motivation, i.e., because the speaker in a specific context might feel an urge to express himself more clearly than he usually does. The motivation for the expansion and the subsequent grammaticalisation of the new form, however, should be found in an increasing identification with recognizable elements in the

direct context. In (2) the time adverb *ya* 'already' is such an element: this word can be easily connected with the chronological view on age as far as its meaning is concerned. It is therefore more likely that a hearer will associate the use of *estar* with the word *ya* than with a phenomenon such as 'chronological view on age' which is difficult to be separated from the context. This identification with *ya* should be taken as a first step on the way to grammaticalisation. Therefore, in MC the first context to be conquered by *estar* could have been the one containing the adverb *ya*. Table 3 shows the distribution of *ser* and *estar* in age expressions with and without the time adverb *ya*.

Table 3: distribution of ser vs. estar in age expressions with presence/absence of the time adverb ya 'already' in MC.

X =14,5; p<0,001 Odds Ratio =4,8	Total N	*ser*	*estar*
+ *ya*	30	33%	67%
- *ya*	129	71%	29%

Table 3 shows clearly that in the great majority of the cases with *ya*, *estar* is used.

De Jonge (1991b) shows in addition to these data how in CC the conquest of a new group of contexts takes place, namely those with the conjunction/time adverb *cuando* 'when': use of *estar* with *ya* supposedly made an associaton possible with another adverb with temporal implications. In (3) an example of this new context for *estar* in CC is given (Rosenblat 1979:660):

> (3) Inf. - ... no, pero nosotros sabíamos que no debíamos pasar...
> Enc. - ... ¿Sabían qué?
> Inf. - ... porque ya sabíamos... Entonces *cuando estaba* muchacha decían:
> "Esa es la calle de las mu... de las mujeres malas."
> 'Inf. - ... no, but we knew that we should not pass...
> Int. - ... Knew what?
> Inf. - ... because we knew... In those days *when* I *was* [*estar*] little, they
> said: "That is the street of the wo... of the bad women."'

In this example we cannot speak any more of the relevance of circumstances; the age expression is used as an indication of a moment in the past (cf. De Jonge 1987).

Table 4 renders the distribution of *ser* and *estar* in contexts where *ya* and/or *cuando* occur in the CC corpus.

Table 4: distribution of ser *vs.* estar *in age expressions with presence. of* ya *and/or* cuando *in CC.*

X =14,6; p<0,001 Odds Ratio =3,2	Total N	*ser*	*estar*
+ *ya/cuando*	78	31%	66%
- *ya/cuando*	114	59%	41%

As appears from table 4, the great majority of age expressions with the presence of *ya* and/or *cuando* show *estar* as the preferred copula (66%), whereas in the other cases *ser* is still the preferred form (59% *ser* vs. 41% *estar*).[8] CC seems to behave with *ya/cuando* like MC with *ya*, especially when we consider the distribution shown for CC in table 4, but now for MC in table 5:

Table 5: distribution of ser *vs.* estar *in age expressions with presence of* ya *and/or* cuando *in MC.*

X =0,7; p<0,5 Odds Ratio =1,3	Total N	*ser*	*estar*
+ *ya/cuando*	60	60%	40%
- *ya/cuando*	99	67%	33%

From table 5 it is clear that the observed preference of *estar* for contexts with *ya* and/or *cuando* in CC cannot be observed in MC. In MC the majority of the contexts with *ya* and/or *cuando* is still with *ser*, so the small rise of *estar* in comparison with the contexts in which *ya/cuando* are absent (33% to 40%) undoubtedly is due to the contexts with only *ya* as illustrated in table 3. And moreover, all the statistical tests indicate that the observed differences are not to be taken as significant. In other words, in MC, the marked contexts for use of *estar* were **only** those with the presence of *ya*; in CC, the contexts with *cuando* have been added to this group.

In view of the results presented in tables 3 to 5, two questions can be raised:

---

[8] In my view, the fact that the use of *estar* in some contexts is relatively higher than in other contexts of MC (cf. table 3) has to do with the fact that in CC *estar* is no longer the numerically marked form; the general tendency to use *estar* in case of doubt must become greater. At this time, we do not know how this factor may influence the process of language change as a whole.

a)   Supposing that the innovative uses of *estar* in age expressions in MC and CC have the same historical origin, why is it so that CC appears to be one stage ahead of MC?

b)   Why did the process of language change come to a standstill both in MC and CC, and, especially, why did it do so at different moments?

4. *One step forward in the process of language change.*

If the (almost) systematic use of *estar* with *ya* in MC is to be taken as a first step in the grammaticalisation of the process of language change, as suggested above, then the basis for the next step has to be found in that specific group of age expressions. Therefore, it is my supposition that in CC there must be a significant relationship between the age expressions with the presence of *ya* and those with the presence of *cuando* in order to be able to explain why in CC this group was the next context to receive use of *estar*. First we will discuss example (4) (Rosenblat 1976:475) and see how such a relation may come about. Informant A is the daughter of informant B:

(4)   Inf.A. - Él me hacía horrores a mí, era el enemigo malo, lo llamábamos [risas]: "¡Ahí viene mi enemigo malo!" [risas]. Él nos espiaba...
      Inf.B. - Sí, pero eso era *cuando ya estaban* más grandes, y les espiaba las conversaciones...
      'Inf.A. - He did horrible things to me, he was the bad enemy, we used to call him [laughter]: "There comes my bad enemy!" [laughter]. He spied on us...
      Inf.B. - Yes, but that was *when you [inf.A and her brothers and sisters] were [estar] already* older, and he spied on your [pl] conversations...'

In this example it is clear that from a qualitative point of view *ya* and *cuando* are easily related to one another: the reference point in the past indicated with *cuando* not only coincides with the moment of *already* being older, but in fact it cannot be separated from a temporal point of view. Moreover, in both the MC and CC corpora, both elements were always juxtaposed when occurring together, although not always in the same order. This fact of course facilitates an association of the two elements.

In order to check if this qualitative relation between *ya* and *cuando* also exists quantitatively, all contexts of MC and CC showing the presence of *ya* were taken and controlled for the presence of the word *cuando*. It was my expectation that these two 'adverbs' should have a higher correlation in CC than in MC to such an extent that it could explain the contamination of *cuando* with *estar* in CC where

this would have been very unlikely in MC. The results of this test are given in table 6.

Table 6: presence of cuando in contexts with presence of ya in MC and CC.

X =1,4; p<0,5 Odds Ratio =3,2	Total N	only *ya*	*ya + cuando*
MC	30	88%	12%
CC	15	67%	33%

Although the total numbers are low and the results are therefore not statistically significant, an odds-ratio of 3,2 indicates a considerable skewing. Moreover, the results do not run counter to the expected tendency: in CC *ya* and *cuando* co-occur almost three times more often than in MC, namely in 33% vs. 12% of all cases where *ya* is present. In practice, this comes down to one case of *cuando* for each three *ya*'s in CC, whereas in MC this is the case in only one out of eight *ya*'s. In spite of the statistical insignificance of these results, it can be concluded that it is at least plausible that a frequency of one out of three cases can be taken as a sufficient basis for a contamination of *estar* in the contexts with *cuando* through its prototypical context with *ya*, as explained above.

Having concluded this, then the answer to question b) above is also given: apparently a frequency of one *cuando* to eight *ya*'s is not sufficient for a contamination of the contexts presenting *cuando* with *estar* in MC, and so the process of language change comes to a standstill. In much the same manner, use of *estar* cannot expand any further in CC until a new candidate in the contexts with *estar* comes around so that a new contamination may take place.

5. Conclusion: pitch sticks.

In this article I have argued that processes of language change are not necessarily of a gradual character, but can be better represented as a process that is essentially discontinuous. This representation offers an important advantage over the general representation of language change thusfar. The proposed hypothesis can explain why some processes of language change can come to a standstill for a shorter or longer period of time. The MC and CC data presented here show that the continuation of the process depends on the circumstances within which the new form appears. If an element occurs often enough in particular contexts (as with *cuando* in CC), a contamination takes place so that the train of language change can move ahead to the next station. If there is no new element at hand that can bridge the gap to the next station, the train holds still. How long does the

train hold still? At this time, the question remains unanswered. It appears that some (casual?) change in frequency of use of the elements that surround the innovative form in its contexts must arise in order to make a new contamination possible.

Finally, the proposed hypothesis also explains those processes of language change where the innovative form rises suddenly from 0% to 100%. This change has to be described either as a big leap into one group of contexts or can be characterised as having so many small steps that *grosso modo* it looks like the graph rendered in figure 1.[9]

It seems that we have come across an important missing link in language change as described by Keller (1989) in the 'invisible-hand theory'. What is needed in the theory of language change is an invisible, inevitable force that can account for the continuation of language change, a force that can explain the shift of a new form from one context to the next. The invisibility of the force we have discussed here is that it does not lie with the copula themselves, but with a casual co-occurrence of two elements that have to do with time and apparently nothing with age or *estar*. The natural relation between age and time should explain why this co-occurrence could ultimately influence the use of *estar* in CC. Thus, the more frequent co-occurrence of *ya* and *cuando* in CC as compared to MC should explain why the process of language change is one step ahead in CC: an identification of *cuando* with *estar* could take place only in CC and not in MC. Thus, the main conclusion is: *pitch sticks*, but only if there is enough of it.

References

Aitchison, J. (1991) *Language change: progress or decay?*, Cambridge University Press, Cambridge.

Allan, W. Scott (1987) 'Lightfoot noch einmal', in *Diachronica* IV, 123-157.

Bailey, Ch.-J. (1973) *Variation and linguistic theory*, Center for Applied Linguistics, Washington, DC.

Butler, C. (1985) *Statistics in linguistics*, Basil Blackwell, Oxford/New York.

Falk, J. (1979) 'Visión de norma general vs visión de norma individual', in *Studia Neophilologica* 51, 275-293.

García, E.C., R. de Jonge, D. Nieuwenhuijsen, and C. Lechner (1990) '(V)os-(otros): ¿dos y el mismo cambio?', in *Nueva Revista de Filología Hispánica* XXXVII, 63-132.

Hengeveld, K. (1991) 'Tipología, diacronía, sincronía', in *Foro Hispánico* 2, 81-94.

de Jonge, R. (1987) 'Estar comes of age', in F. Beukema and P. Coopmans, eds., *Linguistics in the Netherlands 1987*, Foris, Dordrecht, 101-110.

de Jonge, R. (1990) *Sprekersstrategieën en taalverandering: ser en estar in leeftijdsuitdrukkingen*, dissertation, Leiden University.

---

[9] This idea is much alike the 'slow-quick-quick-slow' representation of sound change as described in Aitchison (1991:76-88). She supposes that syntactic changes also 'seem to follow the typical S-curve slow-quick-quick-slow pattern associated with sound change' (1991:98), but gives no further evidence. I am indebted to Olga Fischer for drawing my attention to Aitchison's vision on language change.

de Jonge, R. (1991a) 'Verschuivingen in het gebruik van ser en estar in Latijns-Amerika', in C. van Esch, and M. Steenmeijer, eds., *Spaans in onderwijs onderzoek en bedrijfsleven* 2, KUN, Nijmegen, 50-61.

de Jonge, R. (1991b) 'La cosa (no) es como está', in C. Hernández, G.P. Granda, C. Hoyos, V. Fernández, D. Dietrick and Y. Caballera, eds., *El español de América* I, Actas del III Congreso Internacional de El Español de América, Junta de Castilla y León, Salamanca, 495-504.

Keller, R. (1989) 'Invisible-hand theory and language evolution', in *Lingua* 77, 113-127.

Kroch, A.S. (1989) 'Reflexes of grammar in patterns of language change', in *Language variation and change* 1, 199-244.

Kruskal, W.H. and J.M. Tanur, eds. (1978) *International Encyclopedia of Statistics* Vol.2, MacMillan Publishing Co. Inc., New York/London.

Lope Blanch, J.M. (1971) *El habla de la ciudad de México: materiales para su estudio*, Universidad Nacional Autónoma de México, México D.F.

Pountain, C. (1982) '*Essere/stare as a romance phenomenon', in N. Vincent and M. Harris, eds., *Studies in the Romance Verb*, Cremona Vol, London, 139-160.

Radford, A. (1976) 'On the non-transformational nature of syntax: synchronic and diachronic evidence from Romance causatives', in M.B. Harris, ed., *Romance syntax: synchronic and diachronic perspectives*, University of Salford, Manchester, 69-95.

Rosenblat, A. (1979) *El habla culta de Caracas: materiales para su estudio*, Ediciones de la Facultad de Humanidades y Educación, Instituto de Filología Andrés Bello, Universidad Central de Venezuela, Caracas.