Linguistic variation in the subjuntivo imperfecto in Spanish America in the 16th century

Mircea Branza and Vincent J. van Heuven
Universiteit Leiden Centre for Linguistics

1. Introduction

It is generally accepted that the Indicativo and the Subjuntivo moods are in complementary distribution in Spanish, since the choice between the two is not arbitrary; it depends, among other things, on the meaning of the governing verb such that the two moods occur in contexts that suggest a difference in meaning.

(1) Le digo que *canta* mejor que antes. (*canta* is Indicativo)
him tell-1SG that sing-3SG-Ind better than before.
'I tell him that he sings better than before'

(2) Le digo que *cante* mejor que antes. (*cante* is Subjuntivo)
him tell-1SG that sing-3SG-SUB better than before.
'I tell him that he should sing better than before'

In (2) the indirect object of *digo* and the subject of *cante* are coreferential; this is not a requirement in (1).

In principle, the Subjuntivo occurs in embedded sentences only, and whether it is used depends on the meaning of the governing verb.[1] If this verb expresses the meaning of wish, command, doubt, uncertainty, feeling or emotion, as in (4) but not in (3), then the Subjuntivo is, according to the norm, the only option.

(3) Quiero que *canten*.
want-1SG that sing-3PL-SUB
'I want them to sing'

(4)   Dudaba   que *cantasen/cantaran* tan bien.
       doubt-1SG-PAST that sing-3PL-PAST-SUB so   well
       'I doubted whether they sang so well'

The Subjuntive mood in Spanish has four relatively frequent forms, viz. the Presente *cante*, the Perfecto *haya cantado*, the Imperfecto *cantara/cantase* and the Pluscuamperfecto *hubiera cantado/hubiese cantado*. Also, there are two highly infrequent forms which are only used in certain dialects and in legal documents, i.e. the Futuro *cantare* and the Futuro perfecto *hubiere cantado*. In this article we will only be concerned with Subjuntivo Imperfecto.

In present-day Spanish the Subjuntivo Imperfecto signals an event that is related to the past, present or future, and whose temporal demarcation is irrelevant. As a result, this form corresponds to three separate Indicativo forms (cf. Gili y Gaya 1958: 157). In (5), for instance, the event of *cantasen/cantaran* has its point of orientation in the past that is expressed by *quería*, and extents into the future.

(5)   Quería   que *cantasen/cantaran* mejor.
       wish-1SG-PAST that sing-3PL-PAST-SUB better
       'I wished that they would sing better (from then onwards)'

As is apparent from (4) and (5), Spanish has two forms of the Subjuntivo Imperfecto; one ends on *-se*, the other on *-ra*. These two forms have attracted considerable attention from linguists for at least three reasons.

The first reason lies in the fact that both forms are the result of language change. The original form of the Subjuntivo Imperfecto in Latin, e.g. *cantarem* 'I sing past subjunctive', disappeared in Spanish, as it did in practically all other Romance languages. Its position was taken over in Spanish by *cantase* (originally the form for the Subjuntivo pluscuamperfecto, *cantāvissem*, in Latin). The competing form *cantara* was the Indicativo pluscuamperfecto in Latin (*cantāveram*).

The change from *-se* to *-ra* started in the Middle Ages, and the first attested form that could be interpreted as an example of a *-ra* form functioning as a Subjuntivo dates back to the 13th century (Pérez 1998):

> *Conviene qe fablemos en la nuestra privanza*
> *del pleito de mi duelo, la de mi malandanza*
> *cómo sufrí martirio sin gladio e sin lanza,*
> *si Dios nos **ayudara**, fer una remembranza* (Berceo, DV, 44d)

> We ought to talk in intimacy
> about the cause of my struggle, that of my misfortune,

> about my martyrdom without sable and without lance,
> if God would help us remember

The second reason why these forms are interesting is that some linguists consider the use of these two verb forms a prime example of free variation, whilst others claim that the two forms signal a difference in meaning, be it ever so subtle and difficult to describe.

The third reason is that the token frequency of the verb form in *-se* has steadily decreased since the 13th century, with the result that it has virtually disappeared in all forms of Spanish as spoken in Latin America today. As a case in point, Marín (1980) studied stage plays that were written in Spanish America in the period between 1920 and 1975. His corpus contains a type of text that is intended to be spoken, and which mimics the language use of everyday life. Marín (1980: 219) shows that the *-se* form occurs in less that 2% of all past tenses in the Subjuntivo. In Iberian Spanish, i.e. in Europe, the loss of the *-se* forms has not progressed to the extent that is has in Spanish America. According to Marín (1980: 205) the token frequency of the *-se* form in the Subjuntivo Imperfecto in Spain is 13% of all past forms of the Subjuntivo. On the basis of the above we ask the following questions:

1. In so far as the *-se* form still exists, are the two forms (*-se* and *-ra*) fully equivalent, and therefore exchangeable in all contexts, in present-day Spanish as spoken in Latin America?
2. If (1), did the equivalence exist in earlier stages of the language, or is this a recent development in present-day Spanish?
3. If the latter, then when and how did the neutralisation come about which apparently has led to the existence of free variation in present-day Spanish?

In the present paper we will not be concerned with the first question, although it part of the first author's research program. Instead we will concentrate on the second and third questions, and analyse the use of the two forms in an earlier stage of the language, i.e. the 16th century, which can be shown to be a critical stage in the shift from forms on *-se* to forms on *-ra*.

2. Approach

Since no sound recordings are available from the period before, say, 1900 and given that explicit descriptions of language in its spoken form (as opposed to written language) are virtually non-existent, we have looked for a corpus of

texts that might be taken as the written equivalent of informal, spoken language. We believe that we have been able to locate such a source of informal language use in the corpus we have used for our quantitative analysis. The corpus of texts dates back to the 16th century. Although we would have liked to access even earlier stages in the development of Spanish, our corpus is probably the oldest collection of informal language use in Spanish America. Since the loss of the *-se* form in favour of *-ra* has been more pervasive in Latin America than in Europe, the choice for our corpus seems well motivated.

The corpus comprises 648 letters, which were written in a range of geographic regions in Spanish America in the period between 1540 and 1616.[2] The letters were sent to Spain by 527 different senders. All letters were written by first-generation immigrants. The letters accompanied shipments of money that were sent back to the Old World, or were attempts to persuade next of kin and close acquaintances in Spain to also emigrate to America.

Since the corpus contains personal letters rather than official or legal documents or literature, and because writers and addressees were close relatives and friends, we argue that the type of language used closely approximates what can be called informal, colloquial language. This impression is supported by our observation that the writers do not heed the prescriptions of normative Spanish grammar in the 16th century, which would advocate the use of subordinating conjunctions. What we find instead is large-scale use of paratactic constructions, i.e. sequences of main clauses coordinated by repeated use of *y* 'and'.

Of the 527 authors 476 were men and 51 were women. The senders belonged to the upper or middle social strata. Typical professions listed are medical doctors, lawyers, priests, military and merchants. As a result the corpus is fairly representative of informal language use during the target period. The letters have been translitterated and published as a book *Cartas Privadas de emigrantes a Indias* (Private letters by emigrants to America), 1540–1616 (Otte 1988).

For our research it was necessary to collect every instance of verb forms in the corpus that end in *-ra* en *-se*.[3] To facilitate the search we decided to produce a machine-readable version of the corpus. To this effect every printed page in the *Cartas Privadas* was digitally scanned and converted to a text document using Optical Character Recognition technology. The text documents are available in MS-Word format, and were automatically spellchecked.[4] All remaining OCR errors were corrected manually. Every letter in the *Cartas* was then analysed by hand, and relevant information on name, gender and geographic

location of writer and addressee, as well as the date the letter was written, was extracted and entered in a database of meta-information on the *Cartas*.

We then automatically located all instances of the target verb forms from the letters, which were coded for a large number of potentially important intralinguistic and extralinguistic variables. This information was stored in a second, much larger database, using a relational management system such that textual information and meta-data on the letters could be accessed together. For statistical processing relevant information from both databases was later transferred to SPSS.

3.   Database[5]

Every verb form in the *Cartas* ending on *-ra or -se* was extracted from its context and entered as a record in the database. The verbs were coded in terms of a number of intralinguistic (textual) variables and extralinguistic (social) variables. We will now list and discuss both types of variables.

3.1   Intralinguistic variables

Mood. Since verb forms ending in *-ra* may signal the meaning of the Condicional (modal verbs only), Indicativo or Subjuntivo, all verbs were coded for mood. The decision which of the three moods was signalled by the verb form, was based on contextual information. The decisions were made by two experts on Spanish, i.e., the first author and an assistant (see also note 5). The coders based their decision on the norms that apply to both classical and present-day Spanish as formulated by the Royal Spanish Academy (Real Academia Española 1974). The coders made their decisions independently of each other. In a second stage codes were compared. Only in 1 out of the total of 2527 forms did the codes not converge; this case was then coded as 'undecided'.

Clause type. This variable coded whether the target verb occurred in the main clause or in an embedded clause. Subclauses were subdivided into a large number of types following a traditional classification system. These decisions, which are based on traditional intuitive parsing of the sentences, were made by the assistant and spot-checked by the first author. The subtypes we distinguished are listed in Table 1. The table also specifies key words that function as cues that help make the decision, plus some English near-equivalents.

Table 1. Types of dependent clauses coded in database, plus typical hypotactic indicators in Spanish (and English).

Type of clause	Subclause typically introduced by	English
Non-hypothetical clauses		
1. subject	*que*	*who, what*
2. object	*que, como, si, qué, cuándo, dónde, a quién(es)*	*that, what*
3. prep. object	preposition + *que*	–
4. predicative	form of *ser* + *que*	–
5. relative	*que* or some relative pronoun	*that*
6. instrumental	*con* (+ definite article) + *que*	*with which*
7. comparative	*cual, como, cuanto, que*	*more .. than*
8. temporal	*ahora que, a la vez que, a medida que, …*	*before, after*
9. locative	*donde, de donde, hacia donde, por donde, …*	*where*
10. manner	*como, sin que, con que, según*	*like*
11. consecutive	*a que, para que, a fin de que, con el fin de que, ..*	*in order to*
12. causal	*porque, pues, puesto que, ya que, como*	*because*
13. effect	*luego, conque, así es que, por lo tanto, pues, …*	*so that*
Hypothetical subclauses		
14. conditional	*si, como, en el caso de que, a condición de que*	*if*
15. concessive	*aunque, a pesar de que, aun cuando, así*	*though*
16. modal	*cual si, como si*	*as if*

The sixteen types could later be reduced to two supercategories, i.e., a small set of so-called hypothetical clauses (types 14–16) and a much larger set of non-hypotheticals that comprises all other types.

Event. Independently of the coding of the subclause, the two forms of the Subjuntivo Imperfecto were coded in terms of the type of event they expressed. The event could either be hypothetical or non-hypothetical. This latter type was not subdivided. The hypotheticals, however, were subdivided into those that suggest the event will probably occur as opposed to events that were presented as highly unlikely to become reality. This part of the coding (which applied to a relatively small subset of the target forms) was performed by both experts independently. Conflicting codes were resolved through discussion between the experts. When no agreement could be reached, the case was coded as undecided, and omitted from further analysis.

Governing verb. When the target verb occurred in a subclause, the governing verb in the main clause was located and separately coded for the choice of ending, i.e., whether the governing verb did or did not end in *-ra*. This, of course, is

a purely formal decision that does not rely on linguistic intuition. The variable was included in order to check our intuition that Spanish has a predilection to use the *-ra* form in the subclause if the governing verb in the matrix sentence also ends on *-ra*.

3.2 Extralinguistic variables

These variables serve two purposes. The names of authors and addressees are irrelevant as such, but may indicate social status (not coded in our data) and with only very few exceptions allow us to determine the gender of the individuals. The names then are needed to determine authorship, i.e. to decide whether two texts were written by the same or different individuals. Gender, both of sender and of addressee, may have consequences for the choice between the competing verb forms. It has been claimed, and experimentally corroborated, that in situations of language change women generally lead the change, at least when the change is 'from below' (Labov 2001). Changes from below are subtle changes the linguistic community at large is not aware of. The linguistic variable concerned is not subject to explicit norms. By extension of the argument, we would expect communications between a female sender and a female recipient to be the most progressive type, i.e., to provide the social context that is most conducive to the emergence of the new phenomenon, in our case verb forms ending on *-ra*.

Geographic information on sender and addressee are potential sources of information on the type of dialect spoken by the author. Since all authors were very recent immigrants it is believed that the place where the letters were written will not be very useful in pinpointing the variety of Spanish spoken by the author. Since the authors and their addressees were always members of the same family or circle of close friends, we will assume that the geographic location of the recipients of the letter is indicative of the Spanish dialect spoken by the author.

4. Results

Intralinguistic factors. In Table 3, we crosstabulate absolute and relative frequency of verb forms ending on *-se* and *-ra* in subclauses whose governing verb in the main clause does or does not end in *-ra*. The table is split, such that Table 3A presents the data for those cases in which the event portrayed in the

Table 3. Absolute and relative frequencies of verb forms ending on *-se* and *-ra* in subclause crosstabulated against governing verbs in main clause that do and do not end in *-ra*, broken down by occurrences in non-hypothetical (panel A) and hypothetical subclauses (panel B).

A. Non-hypothetical subclause	Verb form in subclause ends in					
	-se		*-ra*		Total	
Gov. verb does not end in *-ra*	902	84.5%	166	15.5%	1068	100%
Governing verb ends in *-ra*	13	19.1%	55	80.9%	68	100%
Total	915	80.6%	221	19.4%	1136	100%
B. Hypothetical subclause	**Verb form in subclause ends in**					
	-se		*-ra*		Total	
Gov. verb does not end in *-ra*	175	69.4%	77	30.6%	252	100%
Governing verb ends in *-ra*	4	2.1%	190	97.9%	194	100%
Total	179	40.1%	267	59.9%	446	100%

dependent clause is non-hypothetical, whilst the presentation is hypothetical in Table 3B.

A comparison of the numbers in Tables 3A and B reveals, first of all, that target verbs occur much more often in non-hypothetical clauses (1136) than in clauses that present some hypothetical event (446). Next, there is a large shift in the proportion of *-se* versus *-ra* forms in the subclause, depending on the type of event that is portrayed in the subclause. If the subclause is non-hypothetical, forms on *-se* are roughly four times as frequent as the competing forms on *-ra*. However, when the subclause presents a hypothetical event, forms on *-ra* are more frequent than *-se*-forms, in a 60/40 ratio. Clearly, then, the use of the *-ra* form is associated with the presentation of the event as hypothetical, i.e. a condition that promotes the use of the Subjuntivo mood. Third, for both panels A and B we observe that the likelihood of finding a *-ra*-form in the subclause increases substantially if the governing verb in the main clause also ends in *-ra*: from 16 to 81% in panel A and from 31 to 98% in panel B. We may draw a provisional conclusion here that in the 16th century the change from *-se* to *-ra* has already reached full completion in one specific set of conditions, viz. in hypothetical subclauses governed by a verb in the main clause which itself ends in *-ra*. A second interim conclusion would be that, obviously, the choice between forms on *-se* and *-ra* is not a matter of free variation. Since the choice is largely determined by syntactic/semantic differences in the context, we must assume that there was complementary distribution even as early as the 16th century.

Extralinguistic factors. In this section we will examine the effects of three extra-linguistic variables on the choice between verb forms on -*se* and -*ra*. First we will address the effect of region on the choice of verb. We speculated that settlers in the colonies in South America would be less conservative in their language use as the contacts with the country of origin, i.e. Spain, were more intense and frequent. A crude but potentially effective way of testing this hypothesis would be to assume that the geographic distance between the colony and Spain would strongly determine the intensity of the contacts, i.e. the longer the geographic distance, the sparser the contact. Therefore we predict that the north-eastern colonies, which lie closer to Spain, will be more advanced in the language shift than the western and southern colonies. The more remote, southern colonies are Chile, Peru, Quito (Equador), Rio de la Plata (Uruguay and Argentina), while all other regions were lumped together as proximate: Central America, Venezuela, Antilles, Nueva España (Mexico) and Nuevo Reino de Granada (Columbia). Table 4 presents the division of Subjuntivo foms on -*se* and -*ra*, collapsed over all conditions but broken down by the proximate (north-eastern) versus remote (western/southern) orientation towards Spain.

Table 4. Relative frequency of Subjuntivo forms on -*se* and -*ra* broken down by authors living in regions relatively close to versus remote from Spain.

Distance from Spain	Subjuntivo	
	-ra	*-se*
Proximate	34%	66%
Remote	28%	72%

The data support our hypothesis that writers from the remote colonies were more conservative in their use of the verb forms than writers from regions that are closer to Spain. The interpretation of this finding is not without difficulties. We would have to assume that the change from -*se* to -*ra* started in continental Spain, somewhere in the 13th century, and had begun to gather momentum in the 16th century, when the first wave of emigration to the Americas took place. As was shown in our introduction, the change in Spain never reached the full completion that it got in South America, so that we must assume that Spain led the change in the 16th century but the trend was taken over, and completed more successfully, by settlers in the colonies.

The second prediction we made is that women are more progressive in language change than men, at least when the linguistic innovation is a change from below, as it should be in the case at hand. Since we are looking at a language change in progress, we should also be able to see a shift in the distribution of

verb forms on -*se* towards a preponderance of forms on -*ra* over time. In order to test this second prediction we have simply dichotomised the letters in our database, such that verb forms found in letters written between 1541 and 1580 are in the early part of the target period, and those written between 1581 and 1620 in the later part. Table 5 crosstabulates the relative frequency of Subjuntivo forms on -*ra* by male and female writers in the first versus second half of the target period.

Table 5. Relative frequency of Subjuntivo forms on -*ra* used by male versus female writers in the first versus second half of the target period.

	Gender of author	
years	male	female
1541–1580	36%	45%
1581–1620	49%	58%

The results show two clear effects. First, it is the case that the relative frequency of Subjuntivo forms on -*ra* increases from the first to the second half of the target period with 13 percentage points, irrespective of the gender of the authors. Second, female writers use the innovative feature more often than men (9 percentage points). We take these results as support for both our hypotheses.

5.   Conclusions and discussion

Although the variation between Subjuntivo forms on -*se* and -*ra* started in the 13th century (see introduction), we now have reason to believe that the critical stage in the development, at least in Spanish America, lies in the 16th century. When we apply a simple temporal split in the data between those found in the first half (between 1541 and 1580) of our target period and those in the second half (between 1581 and 1620), the results show that the new form (-*ra*) was found in less that 50% of the cases in the first half but more than 50% in the second half. The general pattern for language changes is that they begin rather reluctantly with a few incidental cases, but accelerate as the change gathers momentum, and then decelerate as the change is almost complete. In this slow-quick-slow pattern the change is fastest when the competing forms occur in equal proportions, i.e., 50/50% (see, for example, Denison 2003). In the case of the change from Subjuntivo on -*se* to -*ra* the fastest change is located exactly in the middle of our target period. We do not know when the change reached its

completion, but for Latin America at least, the final stage of the change must lie later than the 16th century.

The choice between *-se* and *-ra* forms in the 16th century is certainly not a matter of free variation. It can be safely concluded from the effects of syntactic and semantic properties of the contexts in which the new *-ra* forms were preferred over *-se* forms in the Subjuntivo Imperfecto, that the competing forms were in complementary distribution. Two linguistic factors were shown to be highly influential in the choice between the two forms. Using the *-ra* form proved the preferred option in subclauses in which the event was represented as hypothetical, while the same form was clearly less preferred if the event was represented as factual or situated in the future. Moreover, the choice for *-ra* in the embedded clause was clearly co-dependent on the occurrence of the same form in the governing verb.

The quantitative analysis of the data shows that one condition is especially conducive to the use of the *-ra* forms, viz. if an event is presented in the subclause as hypothetical, while the governing verb also has *-ra*. In this condition the change from *-se* to *-ra* seems to be completed: here we find the *-ra* form in 98% of the cases. From this we argue that this specific context is most likely the origin of the change; in the next stages of the change the *-ra* forms spread to non-hypothetical subclauses, and no longer needed licensing by a governing verb in the same form.

Finally, the change seems to qualify as a change from below. If it is, the language users were not aware that a change was going on, or even that a choice could be made between competing forms. Under such circumstances women typically lead the change, and the linguistic community is more open to change as there is less pressure to maintain conservative standards. Although the effects found in the database are relatively small, the data reflect these general properties of a regular change from below. Women appear to be the more progressive language users as they use the innovative *-ra* forms more often than the men (8% more), while the regions in Spanish America that are most remote from the home country, Spain, are also more conservative in the change (6% less).

Notes

1. Throughout this article we will refer to the Spanish moods and tenses by the terms that are traditionally used in Spanish grammars. Suitable English terms would be 'past subjunctive' for Subjuntivo imperfecto, 'pluperfect subjunctive' and 'pluperfect indicative' for Subjuntivo pluscuamferfecto and Indicativo pluscuamferfecto, respectively.

2. Two more letters were written by authors in the Philippines. These were included in the database but will not be considered in the present study.

3. We also located and coded all instances of verbs ending in *-ría*. These forms will be studied in future investigations on the distribution of *-ría* versus -ra in non-Subjuntivo contexts, but will be ignored in the present article.

4. In the edition of the Cartas by Otte (1988), modern Spanish orthography was used, so that a spellchecker for present-day Spanish could be used. Note that the syntax and morphology in the edition still reflects that of the original texts.

5. The database was a Microsoft Access application developed by Ing. P.M. Hijzelendoorn (programmer at the Universiteit Leiden Faculty of Arts). The application checked input formats for context coding and allowed fast visual retrieval of relevant verb forms. Fast communication between the database and the collection of word documents containing the textual materials was implemented through the use of word indices. All the scanning and OCR correction as well as the brunt of the coding of verbs and contextual information were done by drs. Saskia Vandenbroeck (part-time student assistant at ULCL).

References

Denison, D. (2003) 'Log(ist)ic and simplistic S-curves'. In R. Hickey, ed., *Motives for language change*. Cambridge University Press, Cambridge, 54–71.

Gili y Gaya, S. (1958) *Curso superior de sintaxis española*. Vox, Barcelona.

Labov, W. (2001) *Principles of linguistic change. Social factors*. Blackwell, Oxford/Malden.

Marín, D. (1980) 'El uso moderno de las formas en -ra y -se del subjuntivo', *Boletín de la Real Academia Española*, LX (220), Madrid, 197–230

Otte, Enrique (1988) *Cartas Privadas de emigrantes a Indias, 1540–1616*, V Centenario, Consejería de Cultura, Junta de Andalucía, Escuela de estudiso Hispano Americanos de Sevilla.

Pérez, M.R. (1998) *El sistema verbal en Gonzalo de Berceo — Las formas de irrealidad*, Colección Lucus Lingua, Anexos de *Moenia, Revista Lucense de Lingüística & Literatura*, 4, Universidad de Santiago de Compostela.

Real Academia Española (1974) *Esbozo de una nueva gramática de la lengua española*, Espasa-Calpe, S.A., Madrid.