

- (1992a). "Historia del español en el Uruguay". In C. Hernández (comp.) *Historia y presente del español en América*. Madrid: Junta de Castilla y León & Pabecal:743-58.
- (1992b). "El español actual en el Uruguay". In C. Hernández Alonso (comp.) *Historia y presente del español en América*. Madrid: Junta de Castilla y León & Pabecal:759-74.
- , Marias Malcuori and Virginia Bertolotti (1997). *El español en la Banda Oriental del siglo XVIII*. Montevideo: Facultad de Humanidades y Ciencias de la Educación/ Universidad de la República.
- (1997). "Los tratamientos *voseantes* y *tuteantes* en el español hablado en la Banda Oriental (siglo XVIII)". In Luis Behares and Oribe Cures (eds.) *Sociedad y cultura en el Montevideo colonial: 159-69*. Recopilation of the papers presented in the Seminar: "Sociedad y cultura en el Montevideo del siglo XVIII". Cabildo de Montevideo. October 8th-10th 1996. Facultad de Humanidades y Ciencias de la Educación/ Intendencia de Montevideo.
- Fontanella de Weinberg, María Beatriz (1977). "La constitución del paradigma pronominal del *voseo*". *Thesaurus* XXXII: 225-239.
- (1989). *El voseo bonaerense. Visión diacrónica*. Bahía Blanca: Departamento de Humanidades/ Universidad Nacional del Sur.
- (1994). "La incorporación del *voseo* en la lengua escrita bonaerense". In Carlos Hipogrosso y Alma Pedretti (comps.) *La escritura del español: 5-18*. Montevideo: Facultad de Humanidades y Ciencias de la Educación/ Universidad de la República.
- (1999). "Sistemas pronominales de tratamiento usados en el mundo hispánico" en Bosque, I., V. Demonte (eds). *Gramática Descriptiva de la lengua española. Entre la oración y el discurso*. Morfología. Colección Nebrija y Bello. Real Academia Española. Madrid: Espasa.
- Gabbiani, Beatriz (2000). *Escuela, lenguaje y poder*. Montevideo: Facultad de Humanidades y Ciencias de la Educación/ Universidad de la República.
- Kany, Charles (1969). *Sintaxis hispanoamericana*. Madrid: Gredos.
- Moyna, Irene (1996). *The Evolution of Verbal Voseo in Río de la Plata Spanish: Evidence from Plays and Popular Songs Between 1880 and 1930*. Master Thesis. University of Florida, Gainesville.
- Rigatuso, Elizabeth M (1992). "Fórmulas de tratamiento sociales en el español bonaerense de mediados del siglo XIX". *Cuadernos del Sur* 20-21(1992): 65-93.
- Rizos Jiménez, Carlos Angel (2000). "Rasgos coloquiales en la correspondencia familiar uruguaya entre 1800 y 1840". *Estudios Filológicos* 35 (2000): 105-123.
- Rona, José Pedro (1967). *Geografía y morfología del voseo*. Porto Alegre: Pontificia Universidad Católica do Rio Grande do Sul.
- Rosell, A. (1975). *El lenguaje en Florencia Sánchez*. Montevideo: Comisión Nacional de Homenaje del Sesquicentenario de los Hechos Históricos de 1825.

Diachronic Shifts and Register Variation with the "Lexical Subject of Infinitive" Construction (*Para yo hacerlo*)

Mark Davies
Illinois State University

0. Introduction

The "lexical subject of infinitive" construction in Spanish (1a-b) has been the subject of interest for at least three reasons.

- (1a) Juan se fue antes [**de** [yo verlo]]
'John left before my seeing him'
(1b) Es difícil [**para** [**ellas entender** tales cosas]]
'it's hard for them to understand such things'

First, researchers have been interested in several aspects of the underlying structure of LSI, such as how case is assigned to the subject. In Spanish the subject receives nominative rather than objective case, as happens in English (2a-b) (see Suñer 1986, Fernández-Lagunilla 1987, Lipski 1991, Yoon and Bonet-Farran 1991):

- (2a) John left [without [**us** seeing him]]
(2b) It's hard [for [**them** to understand such things]]

Second, a number of researchers have compared the Spanish construction to the "inflected infinitive" in Portuguese, and have questioned whether Spanish is in some way moving towards something like the Portuguese construction (e.g. Brakel 1980, Suñer 1986, Lipski 1991, Painter 1991).

- (3a) o João foi-se [sem [nos o vermos]]
'John left without us seeing him'
(3b) é difícil [para [elas compreenderem tais coisas]]
'it's hard for them to understand such things'

Third, a number of researchers have studied the construction from a more sociolinguistic perspective, and questioned whether or not it is a useful feature to distinguish Caribbean Spanish (where it is supposedly more common) from other dialects of Spanish (e.g. Morales 1986, Bentivoglio 1987, Morales 1988, Lipski 1991, DeMello 1995a, DeMello 1995b).

What is strangely lacking in previous research is hard data on several aspects of the LSI construction. The only truly data-oriented studies are DeMello (1995a, 1995b), which deals with spoken Spanish from the Habla Culta corpus. In no previous study, however, has there been an investigation of LSI in the written register, which would be a very good indication of the degree to which it is becoming part of “standard Spanish”. Our study will focus on LSI in a 35 million word corpus of written Spanish, representing at least one million words from each of twenty one Spanish-speaking countries.

Another shortcoming of previous research is that there is little sense of where LSI is coming from, or where it is going, in a diachronic sense. The historical development of LSI has never been the focus of a previous study. Thus very little is known of its frequency before the mid-1900s, except that it may have been somewhat common in the 1500s (see DeMello 1995a: 826). Our study will investigate the historical evolution of the construction, based on an 80 million word corpus of more than 4000 texts from the 1200s to the 1800s. In summary, the modest goal of the present study is to provide useful data on the historical development and register variation of LSI in Spanish, which will hopefully be of benefit to future studies.

1. Diachronic

As has been mentioned, there has been very little research on diachronic shifts with the LSI construction. While there have been one or two passing references to the construction in historical grammars, it has never been the main focus of any study. More than eighty years ago, Beardsley (1921:236) simply noted the sporadic existence of the construction in Old Spanish, and gave two or three isolated examples. But there was no discussion of its overall frequency, or context in which it most commonly occurred.

Keniston (1937:550) is the most complete description of the construction for the pre-Modern Spanish period, but here again it is mostly mentioned just in passing. He states that “throughout the sixteenth century there is a fairly strong tendency to place the subject before the infinitive, an order which emphasizes the clausal character of the construction”, and he then provides a handful of examples from the 1500s, which are scattered throughout the chapter dealing with infinitival constructions, e.g.:

- (4a) se habían movido **sin él habérselo** mandado (Keniston 1937: 550)
 ‘they had moved away without him having ordered them to do so’
 (4b) sino que **al sol salir** o antes media hora vayan al trabajo (Keniston 1937:545)
 ‘rather, at the rising of the sun or within a half hour, they go to work’
 (4c) ¿Por quién he **de yo penar**? (Keniston 1937:541)
 ‘who should I feel sorrow for?’

Because of the lack of previous research, we are still without answers to some very basic questions regarding the LSI construction in earlier periods of Spanish. These include the following issues, which will form the basis of the discussion in the sections that follow:

- How common was LSI before the 1500s? Keniston shows isolated cases of LSI in the 1500s, and Beardsley shows two or three cases from Old Spanish. Was there a gradual increase in use from the 1200s-1400s, or did it increase sharply in the 1400s and 1500s?
- How common was LSI in the 1500s? Which were the prepositions that allowed it the most?
- Did LSI continue on after the 1500s, throughout the 1600s-1800s and into Modern Spanish?
- What are some of the motivations that might have contributed to historical shifts?

The data for the historical development of the LSI construction comes from the “Corpus of Historical and Modern Spanish” that I have created with funding from the National Endowment for the Humanities, and which is now online at www.corpusdelespanol.org¹. The historical portion of the corpus contains 80 million words in more than 4000 texts from the 1200s-1800s, and it allows a much wider range of queries than any other corpus of Spanish. Unlike CORDE, the corpus of historical Spanish that is funded by the Real Academia Española, the “Corpus of Historical and Modern Spanish” allows searches by grammatical category, lemma, synonyms, frequency in different historical periods and registers of Modern Spanish, as well as by several other features, as seen in (5). This ability to search by grammatical category (as in 5a) is especially useful for the present study because it allows us to find all of the relevant cases of LSI in each of the centuries from the 1200s-1800s.

- (5a) *.prep *.pn_subj *.v_inf [POS]
 (de yo saber, para él decir, sin ellos tener)
 (5b) *.pn_obj !querer.* *.v_inf [POS / SYN / LEMMA]
 (lo quiero ver, la deseaba comprar, se determinó tentar)
 (5c) * gracia 1800s>2 1900s=0 [FREQ]
 (en gracia, tanta gracia)
 (5d) estar.* !ir.v_ndo 1800s=0 1900s>3 [POS / LEMMA / FREQ]
 (estoy yendo, estaba saliendo)

In spite of the fact that the historical portion of the study is based on a very large corpus, there are still relatively few cases of LSI in the corpus. The following table summarizes the cases of LSI in each of the centuries from the 1200s-1800s, as well as indicating the number of words of text in each period, the number of cases of LSI with *por/para/pora* and the number of cases with *de/sin*.

Table 1. Historical distribution of LSI

	# words	para / pora / por		de	
		raw total	# / million	raw total	# / million
1200s	6,500,000	1	0.15	--	--
1300s	2,700,000	3	1.11	--	--
1400s	9,200,000	17	2.28	4	4
1500s	18,245,000	12	0.88	4	4
1600s	10,936,000	4	0.37	--	--
1700s	8,079,000	3	0.37	--	--
1800s	25,929,000	20+	±1.00	2	2

Let us first examine the frequency and use of LSI in Old Spanish. The data show that in the 1200s (6) the only cases of LSI are with the preposition *por*, and this holds for the 1300s as well (7a-c):

- (6) tan grant verguença al menor dellos commo al mayor del mundo por ende sennor **por yo ffazer** seruicio a cassio (Alfonso X / General estoria VI)
- (7a) 'such embarrassment for both great and small, for my having helped C.'
me han por loco **por yo vestir** tales paños commo estos (Anónimo / Libro del caballero Zifar)
- (7b) 'they ridicule me because I wear poor clothes like this'
que nunca lo dexara **por yo saber** morir (López de Ayala / Libro rimado de Palacio)
- (7c) 'that I hadn't let him die, as far as I know'
Sy amj padre matasse // **Por yo matar** mj enemigo en buena lid en campo (Anónimo / Mocedades de Rodrigo)
'if he killed my father, because I killed my enemy in battle'

Although there are relatively few examples from the 1200-1300s, the number of examples per million does increase from the 1200s to the 1300s. The biggest increase, however, is in the 1400s, where there are seventeen cases with *por/para*, exemplified by the following:

- (8a) & non sotiles **por yo atar** las dubdas que se siguieron a su opjnjon (Moises ben Maimon / Mostrador y enseñador de los turbados)
'and not subtle, because I put to rest the doubts caused by comments'
- (8b) entención toda **para yo fazer** mj boda (Alfonso de Baena / Cancionero)
'all of the efforts for me to carry out my wedding'
- (8c) pluguiesse alos dioses que **por yo ferir** avos otras (Juan de Mena / Coronación)
'that it might please the gods that, because I had wounded others on your behalf'

- (8d) & **por yo tener** mas aquien se faz injuria (Alonso de Cartagena / De los officios)
'and because I had someone that might be wounded'

More importantly, it is in the 1400s that for the first time we see cases of LSI with other prepositions such as *de*, in which the construction resembles that seen in certain dialects of Modern Spanish:

- (9a) Resçibo pena solamente en maginar **de yo querer** escuchar vuestra obra (Alfonso de Baena / Cancionero)
'I am saddened simply by the thought of wanting to hear your work'
- (9b) ni me atreua **de yo ingerir** para la virtud vn caullero de sangre tan alta (Martínez de Ampiés / Tratado de Roma)
'nor dare I question the power of such an honorable gentleman'
- (9c) despues le dixo queo que pensaste **de tu ser** bueno & sin pecado (Moises ben Maimon / Mostrador y enseñador de los turbados)
'afterwards, he told him how you considered yourself to be so chaste and without sin'
- (9d) No es tiempo **de yo bivar** (Fernando de Rojas / La Celestina)
'this is no time for me to be alive'

Although Keniston showed a high rate of LSI in the 1500s, our data show quite conclusively that the large increase was actually in the 1400s. There were actually two to three times as many cases (per million words) in the 1400s as there were in the 1500s. This is an important point that we will return to in a subsequent section, where we discuss possible motivations for the increase in LSI.

As may have been predicted from the data in Keniston, however, we do find that LSI remains fairly frequent in the 1500s as well. As before, the majority of the examples are with *por/para*:

- (10a) O muy bendita muger **por tu ser** tan piadosa eres tu la mas dichosa (Juan de la Encina / Teatro completo)
'oh blessed woman, because you are so righteous you are the most fortunate'
- (10b) que es cosa maravillosa de ver del arte como salen por mitad de los ríos, y **para ellos dar** gracias a Dios nuestro señor (Cieza de León / Crónica del Perú)
'it's wonderful to see how they go into the middle of the river, and for them to thank God'
- (10c) y aun es poco **para yo vivir** sin afrenta (Cortés / Cartas de relación)
'it is a small thing for me to live in peace'
- (10d) parece que bastaría **para yo conseguir** victoria (Pérez de Oliva / Obra selecta)
'it seems that simply being victorious would be enough for me'

As with the 1400s, however, there are still some cases with other prepositions such as *de*:

- (11a) pues como padre, que huelga **de yo acertar**, me avisa de lo que me conviene (Luis de Granada / Historia de Sor María de la Visitación)
 (11b) D'esso me guarde Dios -dixo Polendos- **de yo fazer** tal yerro (Anónimo / Primaleón)
 'May God not allow me - said P. - to make such a mistake'
 (11c) y la causa **de yo venir** por su tierra de esta suerte ha sido ... (Juan de Timoneda / El patrañuelo)
 'and the reason that I have come into your lands this way is that ...'
 (11d) porque de otra manera ningún remedio había **de yo vivir**. (Núñez Cabeza de Vaca / Naufragios)
 'because otherwise there was no way that I could live'

We have seen that LSI did increase markedly in the 1400s, and we know that it is also quite common in certain dialects of Modern Spanish. The question, then, is whether it remained common throughout the intervening period (the 1600s-1800s), or whether it more or less disappeared for two or three hundred years, only to reappear in Modern Spanish. The data suggest that the latter scenario is correct. In contrast to the 1400s-1500s, there are only isolated cases of LSI in the 1600s-1700s and these occur only with *por/para*. Shown below are three cases from the 1500s (12) and three more from the 1600s (13):

- (12a) aunque no lo es **para ellos tornar** armas, siempre que se les proporcione ocasión (Díaz de Guzmán / La Argentina)
 'although it is not strange for them to take up arms, since they do so whenever they have the chance'
 (12b) Porque a la verdad **era para ellos** ver hombres a caballo ... tan asidos y trabados (Palafox y Mendoza / Memorales)
 'truly it was something to see them on horseback, so nicely arrayed'
 (12c) Dióle motivo **para él ver** la disposición de cara y talle de su compañero (Castillo Solórzano / Aventuras del Bachiller Trapaza)
 'this was enough to make him sit up and take notice of his companion'
 (13a) Porque, en fin, era tentarle **para yo lograr** ocasión para manifestar lo poco que sabe (Forner / Los gramáticos...)
 'because it was tempting for me to show how little he knew'
 (13b) se hubieron por contentos **por yo quedar** libre e restituído en la sucesión ... (Martínez Marina / Teoría de la Cortes...)
 'they were made glad at my having been made free and been restored to the throne'
 (13c) sino que sería conocida ventaja **para él limpiar** el Estado enteramente de ellas (Campillo y Cosío / Nuevo sistema)
 'it would be a great victory for him to remove all of them from the country'

In the 1800s, however, LSI increases to approximately the same level as the 1400s-1500s. There are numerous examples with *por/para* (14) and there are also sporadic cases again with other prepositions such as *de* (15):

- (14a) era más importante **para ella hablar** con Nepomuceno que andar por allí dando saltos (Alas / Su único hijo)
 'it was much more important for her to talk with N. than for her to simply waste time there'
 (14b) Elvira dio a conocer cuán costoso era **para ella salir** de su oscuridad Navarro Villoslada / El ante-cristo)
 'E. let them know just how hard it was for her to leave this darkness'
 (14c) ¿Cómo era posible **para ella vivir** sin que se inclinaran mil frentes en su presencia? (Castelar / La Hermana de la Caridad)
 'how could she live, unless a thousand heads should bow in her presence?'
 (14d) Sería más importante **para ellas participar** de la prerrogativa (Nervo / La lengua y la literatura)
 'it would be more important for them to take advantage of the prerogative'
 (15a) además **de yo haber** gastado todo cuanto yo tenía (García Icazbalceta / Colección de documentos...)
 'in addition to my having spent all I had'
 (15b) cuya muerte he de pagar al tiempo **de yo parir** (Caballero / Genio e ingenio del pueblo andaluz)
 'whose death I shall avenge when I give birth'

In summary, the data show that LSI has always been possible in Spanish since at least the 1200s. There was a significant increase in the 1400s-1500s, followed by a decrease in the 1600s-1700s. By the 1800s, LSI was on the increase once again. Finally, in the periods in which LSI was relatively uncommon, it occurs only with *por/para*, whereas in the periods of the greatest overall frequency it extends to other prepositions such as *de* as well.

These data naturally lead to a number of questions. First, why did LSI start with *por/para*, and extend from there to other prepositions? Second, how did this extension take place? Third, why did this extension and increase in LSI take place in the 1400s-1500s? Fourth, why was there a decrease after the 1500s?

Regarding the first two questions, it seems uncontroversial that the origins of LSI are to be found in simple prepositional phrases like (16), in which the NP is an evaluator.

- (16) [PP para Juan] ... sería difícil comer sin pagar
 '[for John] ... it would be hard to eat without paying

The PP can also of course be placed before the infinitive, but with the NP still acting as the object of the preposition:

- (17) sería difícil . . . [PP para Juan] . . . comer sin pagar

As Fischer (1995) discusses, once the PP is in the same position as COMP + SUBJ, it is then possible for it to be re-analyzed as such:

- (18) sería difícil [COMP para [SUBJ Juan] comer sin pagar]

And once the NP is re-analyzed as a subject, it is then possible for it to occur with other prepositions:

- (19) pudimos comer allí [PREP sin [SUBJ yo] tener que pagar nada]

Following the model that Fischer proposes, in the 1200s-1300s Spanish allowed constructions like (17). At the point that a reanalysis occurred, along the lines of (18), it then became possible to have sentences like (19).

In addition to the model proposed by Fischer, there are a number of other features of Spanish that may have led to an increase in LSI. First, because of its relatively free word order, Spanish regularly allows SUBJ + INF constructions with other governing elements such as auxiliary verbs:

- (20) bien **puedo yo pasar** entre tanta buena gente (Cervantes / Don Quijote)
'I can easily associate with such illustrious people'

In Old Spanish there was also another PREP + PRON + INF construction that could have potentially led to a re-analysis for the LSI construction. During this time the forms *nos* and *uso/vos* could be either subject or object pronouns. Therefore in cases like the following, there is ambiguity in which the pronoun either may or may not be seen as the subject of the clause:

- (21a) sseredes bien consssejado **de vos tornar** mjo vasallo (1300s / Anónimo / Crónica de veinte reyes)
'you would be well advised to become [+REFL] my servant'

- 'it would be well if you were made [-REFL] my servant'
(21b) nos es menester **de nos detener** de non enbjar por el (1300s / Anónimo / Libro del caballero Zifar)
'it's not necessary for us to hold back [+REFL] from sending for him'

'it's not necessary for them to hold us back [-REFL] from sending for him'

In addition to the potential ambiguity with *nos/vos*, it is also quite common for there to be ambiguity regarding the element that looks like an infinitive. In some cases it may simply be a verb that has been nominalized, with a preceding possessive:

- (22a) & como dixo el fisico **por tu temer** a dios & conplir sus mandamjentos (Anónimo / Libro del caballero Zifar)
'and as the physician said, because of your [fearing / fear of] God and your keeping the commandments'

- (22b) porque **de tu padecer** yo traygo titulo de legitimo heredero (Juan de Flores / Grimalte y Gradissa)
'because of your suffering (V / N), I grant to you the title of true heir

- (22c) por que sy yo no te honrrara nunca la virtud **de tu perdonar** mostrar se pudiera (Diego de San Pedro / Arnalte y Lucenda)
'because if I never honored you for the goodness of your [forgiveness / foregoing], it would be shown'

In summary, there are a number of structural reasons that could have led to an increase in LSI, including a reanalysis of the PP as COMP+SUBJ (18), the ambiguity of constructions with *nos/vos*+ INF (21), and the possible ambiguity with POSS + N (22).

The logical question is why this reanalysis happened precisely in the 1400s. It is unlikely that we will ever know for sure, although there is one possible explanation. As is widely known, there was a strong latinizing tendency in Spanish, which started in the mid-1400s and which lasted until then 1600s, which is the exact period in which we the largest increase in LSI. One of the major features of this latinizing tendency in syntax was an increase in the use of infinitival constructions where earlier and later stages of Spanish would have preferred a finite clause. Examples include the following:

- (23a) quien del real lo vio salir, cierto, **creyó ser** algún indio (Cieza de León / Guerras civiles peruanas)
'whoever saw him leave, thought that it was an Indian'

- (23b) los bandos entre los Cocomes -que **decían ser** echados injustamente (Diego de Landa / Relación de las cosas de Yucatán)
'the groups of C. - which they said were unjustly evicted'

- (23c) (la principal del oficio sacerdotal) porque [depende **de ella**] ser uno bueno y perfecto sacerdote (Muñoz / Vida y virtudes del venerable varón)
'because being a good abd perfect priest depends on that'

- (23d) Di, religioso: ¿no has fiado tu ánima de tu Superior? ¿Pues cómo no fias **de él** [para (omitted)] **estudiar** gramática o artes (González Dávila / Pláticas)
'How is it that you don't trust in him in order to study grammar or arts?'

If in fact the tendency to adopt Latin-like infinitival clauses in the 1400s-1500s is a partial motivation for the increase in LSI in the 1400s, then it may also help to motivate the decrease in LSI after the 1500s, when the stylistics of the language changes and there was less of a latinizing influence on the language.

2. Synchronic

Let us now turn to the distribution of LSI in Modern Spanish. DeMello (1995a: 825-27) provides a very good overview of previous studies and their claims about the existence and frequency of LSI in different dialects and registers, and there is little need to repeat that summary here. The general consensus of the more recent articles is that LSI is found throughout the entire Spanish-speaking world, although it is much more common in Caribbean Spanish. DeMello's extensive data on LSI in spoken Spanish bears this out. It shows that the vast majority of the cases of LSI come from the Caracas and San Juan PR components of the Habla Culta corpus, and that there are only scattered examples from other dialects. For example, while there are 23 examples in Caracas and 29 in San Juan PR, in none of the other ten cities are there more than four cases of LSI.

In spite of the value of DeMello's studies, some important questions still remain unanswered. First, the Habla Culta corpus contains approximately 2,500,000 words of text from eleven countries. Are these corpora representative of the other countries that are not represented in the Habla Culta corpus? How do these data compare with the frequency of LSI in other corpora of spoken Spanish? Second (and most importantly) what is the frequency of LSI in written Spanish? Answering this last question may provide additional insight into the degree to which LSI is becoming accepted in normative Spanish.

In order to answer these questions, we will investigate the frequency of LSI in a 35 million word corpus of written Modern Spanish, taken from more than sixty online newspapers². This includes 15 million words from seven web-based newspapers from Spain and 20 million words from fifty-five web-based newspapers from Latin America, including one million words in twenty different countries. As a result, we are able to look at LSI in a large written corpus from every country in which Spanish is the dominant language, rather than just selected countries, as in the Habla Culta corpus. In addition, in order to verify the data on spoken Spanish in previous studies, we have also used 1,000,000 words from the *Corpus oral de referencia de la lengua española contemporánea*, which represents several genres of spoken Spanish from Spain.

The following table summarizes the number of cases of LSI with three different prepositions in the various sub-corpora of Modern Spanish. The first two rows summarize the results of DeMello (1995a). The third row comes from the 1,000,000 *Corpus oral de referencia de la lengua española contemporánea* from Spain. The last three rows are from the 35,000,000 words from web-based newspapers. In addition, the numbers in parentheses on the rightmost columns

represent the number of tokens that come from reported speech. For example, there are eight cases of LSI with *para* in newspapers from the Caribbean region, and three of these eight tokens are from reported speech.

Table 2. LSI in Modern Spanish

	# words	para	de / sin
Habla Culta (+ Caribe: Ven, PR)	423,000	43	17
Habla Culta (-Caribe)	1,796,000	10	6
Spain - Spoken	1,000,000	4	2
LatAm (+Caribe) - News	3,000,000	8 (3)	2
LatAm (-Caribe) - News	17,000,000	6 (3)	2
Spain - Newspapers	15,000,000	6 (2) [+5 impers]	1 (1)

The first finding from the corpus is that the level of LSI in the *Corpus oral de referencia de la lengua española contemporánea* from Spain is roughly comparable to that in the non-Caribbean corpora from the Habla Culta, which included eight other cities from Latin America and two from Spain (Madrid and Sevilla). There are four examples with *para* and two more with *de*.

(24a) Si yo os he dicho antes que a pesar **de yo llevar** - porque, claro es que yo te diré a ti qué ya tenía - a - a la escuela (CCON022D)

'if I've told you before that in spite of my taking - because, of course I'll tell you what I already had - to school'

(24b) Luego, a la hora de **- de nosotros aplicar** esos criterios (AENT006C)

'afterwards, when we applied those criteria'

(25a) [las capitulaciones] Las Casas consigue una **para él colonizar**, bajo su dirección (BHUM029A)

'L.C. finds one to colonize, under his direction'

(25b) me decís lo que hay y me dais el presupuesto que va a ser, **para yo saber** cuánto dinero me va a costar antes de arreglarlo (CCON013B)

'you tell me what there is and you give me a budget, so I can know how much it's going to cost before I fix it'

(25c) Entonces fue cuando - cuando yo - me dio el - el - el banderazo de salida **para yo tener** que contar el trabajo que yo había estado haciendo durante dos años (BENT015C)

'that's when he gave me the notice for me to count up all of the work that I had been doing for two years'

(25d) O sea - véndeme **50 para yo vender** más tarde. (CTEC038A)

'in other words, give me fifty (of them) for me to sell later on'

These data validate the findings of DeMello (1995a), and show that there is not much difference in the level of LSI between Spain and non-Caribbean Latin America. In addition, the *Corpus Oral* contains a much wider range of spoken genres than the *Habla Culta*, but this apparently has little impact on the

frequency of LSI. Finally, we should note the interesting use of LSI in the relative clause in (25). According to Fischer (1995), this is a relatively advanced type of LSI, in terms of the typological cline and historical development of the construction. It is common in English (e.g. "I found a ball for you to play with"), but our data show that it is still quite rare in Spanish.

Turning now to the written register, we first examine the frequency of LSI in written Spanish from the Caribbean region. Although it is uncontroversial that LSI exists in nearly all levels of spoken Spanish in this region (e.g. Jiménez Sabater 1984: 69), there is less agreement on how common it is in written Spanish. Gili Gaya (1964:143) claims that it is not at all frequent in written Spanish, but this is disputed by Suñer (1986: 191).

This study is the first two look quantitatively at LSI in written Spanish from the Caribbean region, and it shows that although it is much less common than in spoken Spanish of the same region, it does exist. This study defines the Caribbean region as Cuba, the Dominican Republic, and Puerto Rico (as opposed to San Juan PR and Caracas in DeMello 1995a), and we find several cases of LSI in the three million words of web-based newspapers. The least surprising cases are the cases with *para* and which represent reported speech. One of these is even a translation from English (26a), which makes LSI even less surprising

- (26a) "y así es muy difícil **para él lanzar** cada día.", dijo el manager Mike Hargrove (DR: Ultima Hora)
- (26b) "Nosotros, por instrucciones del Señor Presidente, tenemos la responsabilidad de que todos los hospitales existentes queden totalmente saneados en los próximos dos años, **para él dejar** como legado un problema de extrema importancia resuelto a este pueblo" (DR: Ultima Hora)
- (26c) "Cuando yo termino de fregar y enjuago pues el busca mas pa' poder lavar, **pa' yo poder** lavar la casa" (PR: Noticentro)

We should briefly note sentence like (26a), in which the subject pronoun is most likely an object of the preposition *para*. Recall that in (16)-(18) above we showed how – at the surface level – a sentence with an object of a preposition has the same form as the true LSI construction – *es difícil* [*pp para Pepe*] *hacerlo* vs. *es difícil* [*comp para* [*subj Pepe*] *hacerlo*]. In fact, in historical terms this surface-level ambiguity was probably a factor in the reanalysis. Therefore in (26a) and in a number of subsequent examples (27a, 27c, 29a, 30a-e, 32a-b, 33a-e, and 34a-f) we will include cases where the subject pronoun may well be an object of a preposition, simply because it is the surface-level ambiguity that is important here.

Returning to the data at hand, we find that in addition to possible cases of LSI with reported speech, there are seven other cases in written segments of the corpus, including five cases with *para* (27) and two with *de* (28):

- (27a) Señaló el orgullo que representaba **para ella estar** en una asamblea con tantos jóvenes valiosos (Cuba 3)
- (27b) parecen decir ahorren energías para vender más al extranjero y consuman menos **para yo tener** más (Cuba 1)
- (27c) Así lo manifestó el jardinero izquierdo de los Cachorros de Chicago al preguntársele sobre la que ha significado **para él estar** junto a Sosa en la presente campaña (DR: Ultima Hora)
- (27d) El reo quiere conocer al ciudadano español que se interesó en pagar la multa **para él salir** de la cárcel (DR: Ultima Hora)
- (28a) cuando se cumplen 35 años **de ellos haber** jugado juntos (DR: Listin Digital)
- (28b) La señora Carmen Ramirez . . . refirió que después **de ella conseguir** intención de visas por el gobierno español (Cuba: CubaNet)

What are perhaps more interesting are the cases of LSI in Latin American newspapers from outside of the Caribbean region. The frequency in the 17,000,000 words of text from these other seventeen countries is only about one-sixth that of the Caribbean region, but there are still scattered examples. Three of the six examples with *para* are from reported speech (29), and one of these is another translation from English (29a).

- (29a) "Al igual que su padre, Griève es un gran bateador. Debe ser especial **para él haber** roto un juego sin hits", dijo Clemens al diario USAToday (Venezuela: El Universal)
- (29b) "Creo que la posición **para yo triunfar** es ser volante de marca" (Panama: El Siglo)
- (29c) "Para yo hablar de política tengo que estar en Colombia" (Ecuador: El Vistazo)

Yet there are still five other examples with *para* in written segments of the corpus, although some of these have an impersonal reading:

- (30a) Es difícil **para él olvidar** y seguir viviendo igual (Guatemala: La Hora)
- (30b) No fue fácil **para ella vivir** el misterio de su Hijo (México: Diario Yucatán)
- (30c) cuatro de cada 10 asambleístas dejan el trabajo a sus suplentes **para ellos poder** terciar en las próximas elecciones de diputados (Ecuador: El Vistazo)
- (30d) ¿Cómo han tomado los niños y jóvenes esta experiencia? --**Para ellos tocar** el repertorio que se les está enseñando es un reto (Guatemala: Prensa Libre)
- (30e) Cuando le digo, señalándole a los niños que juegan, que **para ellos vivir** así, en este confinamiento y tensión, entre armas, pedreas, estallidos e incertidumbre, será terrible (Perú: Caretas)

Finally, there are two other examples with the preposition *de*, which of course represent the more advanced type of LSI:

- (31a) porque su belleza es como un imán que atrae a los hombres y los entloquece, **sin ella querer**, porque ella no es una chica fácil (Pánama: El Siglo)
- (31b) ¿Podrá Cerpa Cartolini aceptar que su destino se determine en La Habana, **sin él participar**, cual lejano actor en ensombrecido segundo plano? (Perú: Caretas)

Finally, we turn to LSI in the 15,000,000 word corpus of newspapers from Spain. As with the corpus of spoken Spanish from Spain (as compared to the *Habla Culta* corpora from Latin America), there is again very little difference between LSI in the written Spanish from Spain and the written Spanish from non-Caribbean Latin America. There are eight examples of LSI with *para*. Two of these are in reported speech:

- (32a) “entiendo que sea difícil **para él abandonar** el Real Madrid” (El Diario)
- (32b) “ya es muy duro **para él dar** marcha atrás” (ABC)

The other eleven examples are from written Spanish. Five of these, however, have a left-adjoined prepositional phrase that has a highly impersonal reading (33). The other six examples are more representative of true LSI (34):

- (33a) **Para ella ser** necesaria significa ser amada (Vanguardia)
- (33b) **Para ella, actuar** con sus ídolos ya es toda una experiencia (Periódico)
- (33c) **Para ellos estar** aquí ya es un premio (El Diario)
- (33d) “A partir de ahí, todo será **para él coser** y cantar” (El Mundo)
- (33e) **Para él ser** alcalde es casi un oficio (Sur)
- (34a) es ya una costumbre **para ella bajar** de 10.80 (El Mundo)
- (34b) [Monica Lewinsky] así que sería mejor **para ella salir** de Washington. Clinto podría buscarle un empleo... (El Periódico)
- (34c) hubiera sido más fácil **para ellos recurrir** a viejas glorias de siempre (El Mundo)
- (34d) buscando que los ayuntamientos contraten los planes de empleo **para ellos subcontratar** y sacar su beneficio (Comercio)
- (34e) lo importante que era **para él poder** jugar en el vigente campeón de Europa (El Diario)
- (34f) admitió que no había sido fácil **para él prescindir** de Clemente (El Mundo)

Finally, there is just one case with a preposition other than *para*, and this comes from reported speech:

- (35) El centrocampista madrileño quiso dejar claro que “toda persona tiene derecho a elegir su futuro. No es justo que algunos me colocaran en otros equipos **sin yo dar** el visto bueno” (El Diario)

Although there are no cases of LSI in non-reported speech, with prepositions other than *para* – as opposed to the two cases from non-Caribbean Latin American text – there are still too few examples to conclude that LSI is less common or less advanced in Spain. Even with a 35 million word corpus, we sometimes do not have all of the data that we would like.

3. Conclusions

This study provides data on the historical development of LSI and its distribution in written Spanish, which are two topics that have not been studied previously. The 80 million word corpus of historical Spanish shows that LSI has existed since Old Spanish, but that it did increase significantly in the 1400s-1500s. There are a number of structural factors that may have motivated this shift, including ambiguity in the surface form and the latinizing tendency in Spanish during this period. The data for written Modern Spanish is based on 35 million words from web-based newspapers. It shows that LSI in written Spanish occurs at about the same level in Spain as it does in non-Caribbean Latin America. As with the written register, LSI in written Spanish is much more common in Caribbean Spanish than in non-Caribbean Spanish. Most importantly, although LSI is much less common in written Spanish than in spoken Spanish, it is still found in registers like newspaper articles, which presumably have passed through one or two levels of revision. This may suggest that LSI is on the way to becoming an accepted part of standard Spanish in all regions and registers.

Notes

1. Although the historical portion of the Corpus del Español is now online (www.corpusdelespanol.org), queries of the corpus after the time that this article is published may yield different results than those contained in this paper. This is due to some changes in the corpus itself, but also to the different search algorithms employed in this study and in the online corpus.
2. The corpus containing the 35 million words of web-based newspapers was constructed in late 1998, and the original newspaper articles are in all cases from 1996-98. Although the full citation (including publication date and page number) is not given in this paper, the links for the newspapers listed in the citations can be found at <http://mdavies.for.ilstu.edu/corpus/news.asp>, and in many cases the newspaper websites contain searchable archives of past editions.

References

- Beardsley, Wilfred. 1921. *Infinitive constructions in Old Spanish*. New York: Columbia Univ. Press.
- Bentivoglio, Paola. 1987. "Clauses Introduced by the Preposition *para* in Spoken Spanish: An Analysis from a Discourse Viewpoint." *Studies in Romance Languages*, eds. Carol Neidle and Rafael Nunez Cedeno. 1-14. Dordrecht: Foris.
- Brakel, Arthur. 1980. "Infinitives, subjects, word order, and case in Portuguese and Spanish." *Hispania* 63:85-91.
- De Mello, George. 1995a. "Preposition + sujeto + infinitivo: 'Para yo hacerlo'." *Hispania* 78: 825-36.
- De Mello, George. 1995b. "Expresion impersonal + 'para' + persona + infinitivo: 'Es importante para Juan estudiar'." *Estudios Filológicos* 30: 57-68.
- Fernández-Lagunilla, Marina. 1987. "Los infinitivos con sujetos lexicos en español." *Sintaxis de las lenguas románicas*, eds. Violeta Demonte and Marina Fernandez Lagunilla. 125-47. Madrid: Arquero.
- Fischer, Olga. 1995 "The Distinction between to and Bare Infinitival Complements in Late Middle English." *Diachronica: International Journal for Historical Linguistics* 12(1):1-30
- Keniston, Hayward. 1937. *The syntax of castilian prose: the sixteenth century*. Chicago: Univ. of Chicago Press.
- Jiménez-Sabater, Max. 1984. *Más datos sobre el español de la República Dominicana*. Santo Domingo: Universidad Autónoma de Santo Domingo.
- Kany 1950. *American-Spanish syntax*. Chicago: Univ. of Chicago Press.
- Kirschner, Carl. 1992. "The Spanish Subjunctive and the Spanish-English Bilingual: A Semantically-Motivated Functional Shift." *Hispanic Linguistics* 5: 89-108.
- Lipski, John. 1991. "In search of the Spanish personal infinitive." *New Analyses in Romance Linguistics: Selected papers from the XVIII Linguistic Symposium on Romance Languages*, eds. Dieter Wanner and Douglas Kibbee. 201-20. Amsterdam: John Benjamins.
- Morales, Amparo. 1986. "Preposición *para* más infinitivo: Implicaciones en el español de Puerto Rico." *Actas del VII Congreso. Asociación de Lingüística y Filología de America Latina (ALFAL)*, Santo Domingo, 1984. 217-30. Santo Domingo: Filial Dominicana.
- Morales, Amparo. 1988. "Infinitivo con sujeto expreso en el español de Puerto Rico." *Studies in Caribbean Spanish Dialectology*, eds. Robert Hammond and Melvyn Resnick. 85-96. Washington, DC: Georgetown UP.
- Painter, Marília. 1991. "The Inflected Infinitive in Brazilian Portuguese." *Language Quarterly* 29: 1-46.
- Raposo, Eduardo. 1989. "Prepositional Infinitival Constructions in European Portuguese". *The Null Subject Parameter*, ed. Osvaldo Jaeggli and Kenneth Safir. 277-305. Dordrecht Kluwer.
- Suñer, Margarita. 1986. "Lexical subjects of infinitives in Caribbean Spanish." *Studies in Romance Linguistics*, ed. Osvaldo Jaeggli and Carmen Silva-Corvalán. 189-204. Dordrecht: Foris.
- Yoon and Bonet-Farran. 1991. "The Ambivalent Nature of Spanish Infinitives." *New Analyses in Romance Linguistics: Selected papers from the XVIII Linguistic Symposium on Romance Languages*, eds. Dieter Wanner and Douglas Kibbee. 201-20. Amsterdam: John Benjamins.
- Zamora, Juan and Jorge Guitart. 1988. *Dialectología hispanoamericana*. 2 ed. Salamanca: Almar.