

Towards a history of register in Spanish

Christopher J. Pountain

Although the significance of many other dimensions of variation in the data of Spanish historical linguistics is well recognised, the importance of studying variation in register has been underestimated and its feasibility questioned. This is in striking contrast to English historical linguistics, in which the study of register on the basis of electronic corpora is comparatively far advanced. This paper is a small-scale investigation of a 15th-century Spanish text, *Arcipreste de Talavera o Corbacho* (hereinafter referred to as *Corbacho*), whose author is clearly making an attempt to represent, perhaps stereotypically, different contemporary registers. It shows how, through a combination of statistical analysis and philological sensitivity, register-based linguistic variables can be recovered from a relatively short, multi-register text.

Keyword: Spanish, historical linguistics, variation, corpus linguistics, register, genre, style

1. Introduction

The term 'register' is usually considered to have been coined by Reid (1956) and is broadly used in sociolinguistics to refer to variation according to situational context (as opposed, for example, to variation which is dependent upon the place of origin, social class, age or sex of the speaker). Situational context may include quite a large number of factors such as the role and nature of the participants in a speech situation and the relations amongst them; the physical and temporal context of the discourse; its subject-matter and speakers' attitudes towards it, and the medium through which the discourse is expressed. For a comprehensive list, see Biber (1988: 30–31). Halliday's (1978: 31–35) parameters of field (subject-matter), tenor (relations between participants) and mode (medium) are much used in discussions of register. I follow Ferguson (1994: 21–3) in regarding 'register' as variation according to the situational

Spanish in Context 3:1 (2006), 5–24.

ISSN 1571-0718 / E-ISSN 1571-0726 © John Benjamins Publishing Company

© 2006. John Benjamins Publishing Company
All rights reserved

use of language and 'genre' as an overall message type (sermon, chronicle, epic poem, amongst others); while genre may be a factor in the situational context of the discourse, therefore, it is possible, and indeed usual, for a genre to contain examples of many different registers. It is more difficult, I think, to separate register rigorously from social variation, since social class, age and sex will inevitably affect relations between the participants in a speech situation and, from the linguistic point of view, there may be expected to be covariation between class and register, since more educated speakers will control a wider range of registers, especially written registers.

2. Register in historical linguistics

There has been a long-standing recognition of the importance of variation in the data of Spanish historical linguistics, though its analysis has often been rather marginalised, and it has often been implied that the language of a particular century or a particular author is somehow homogeneous. An interesting example of this admission of importance but lack of rigorous analysis is Keniston's (1937) survey of sixteenth-century Spanish syntax, a remarkably innovative work in its day and still used as a fundamental reference source. While Keniston goes to a great deal of trouble in his introduction to point out the diachronic, stylistic and regional variation in his corpus, he is still able to maintain that "the language of these texts represents a fair cross-section of the literary language of the sixteenth century" (p. xxi), as if the notion of 'literary language' were a self-evidently satisfactory concept. As far as the dimension of register or genre variation is concerned (Keniston would not, of course, have referred to 'register' as such), he categorises his texts into the genres of 'dramatic dialogue', 'fiction', 'didactic or religious exposition', 'didactic dialogue', 'letters', 'historical or biographical exposition', 'autobiographies' and 'miscellaneous collections of anecdotes'; more particularly, he claims that some texts contain representations of 'actual conversation'. On another possible variational dimension, namely, social class as defined by educational level, he notes that, while some authors are erudite, others have no formal education. Yet his analysis of the data consists most often of the objective statement of the frequency of different syntactic variants with no consistent attempt at correlation with register. Indeed, his view is that "within all these varying styles, whether those of individuals or classes, whether vulgar or learned, there lies a common core of materials. That is the subject matter of syntax" (1937:5). However, from time to time, some intriguing asides point the way to the possibility of more

rigorous variationist analysis. Discussing the position of qualifying adjectives, for example, Keniston gives more detailed statistics for ten texts, concluding as follows:

“Now this variation in practice corresponds in a rough way to the character of the works themselves. Of the six texts in which there is a predominance of postposition, two are narratives, in letter form, three are expositions (two of them in dialogue form), and one is a novel, in autobiographical form. Of the other four, two are plays and two are autobiographies (one of them in nov-
elesque form). It would therefore appear that if the approach to a subject is objective, as in narrative or exposition, that attitude is revealed by a predominance of objective adjective position. If, on the other hand, a work is primarily a revelation of personality, as in the theater or the autobiography, the subjective adjective position will predominate. It may be objected that the form of both the *Lazarillo de Tormes* and the *Guzmán de Alfarache* is the same — a picaresque novel in autobiographical form. But it is obvious that the character of the two works is totally different: the former is wholly devoted to Lázaro’s account of his own activities; the latter is largely comprised of the author’s comment on life and society.” (Keniston 1937: 299–300)

There is surely much to qualify in Keniston’s ‘psychological’ account of adjective position, which probably cannot be reduced to such a straightforward dichotomy, based on ‘objectivity’ and ‘subjectivity’, as he suggests; but the hypothesis that there is a correlation between adjective position and register or genre¹ is *prima facie* likely and deserving of further investigation. In Pountain (1998a), I presented some statistical evidence to show that register is a consistently significant factor in adjective position in the history of Spanish, the anteposition of qualificative adjectives being strongly associated with cultured writing and almost totally absent from spontaneous speech. Also of great significance is Keniston’s observation that works which apparently belong to the same genre do not necessarily exhibit similar statistics regarding adjective position. This, in fact, anticipates the view that an individual text is not necessarily written consistently in a particular register, and will contain much variation, especially a literary text whose author may consciously or subconsciously manipulate variation for artistic purposes. Indeed, a brief examination of the anteposed adjectives in the *Lazarillo* reveals that the anonymous author is in all probability using this device for stylistic effect. For example, he refers to his book as “esta nonada, que en este *grosero* estilo escribo” (pp.8–9): while it is true that the novel for the most part has superficially the linguistic features of a straightforward narrative redolent of direct speech, it is in fact anything but crude, with its sly verbal humour and understatement; the preposing of *grosero*

can thus be seen to be politely self-deprecating rather than an objectively presented property. Preposed adjectives are also used for oxymoronic effects such as “mi nuevo y viejo amo” (p.22, his first master, the blind man) and “aquel dulce y amargo jarro” (p.32, the pitcher which the blind man brings crashing down on his face to punish his stealing of the wine). It is impossible, therefore, to consider that the language of the *Lazarillo* is consistently characterisable as belonging to a picaresque genre which is definable in linguistic terms, though it is likely that it oscillates between a number of different registers of its time. The same is plainly true of some other genres, most obviously drama. A Lope de Rueda play, for example, typically represents the speech of upper and lower classes, as well as the conventionalised ‘stage dialects’ of gypsies and *negros*, and cannot possibly be thought of as representing a homogeneous spoken language (Pountain 2001: 147).

Today, variation is seen as crucial in understanding language change. As Milroy puts it, “As language is variable at all times, an adequate historical account must be an account of changing patterns of variation (1992: 221).” Penny (2000) is a most important application of variationist theory to the history of Spanish. As he points out, the main theoretical insights have been based on data taken from English and the Germanic languages and had not before his book been systematically applied to Spanish (p.ix), (though that is not to say that scholars had been unaware of variational phenomena). Oddly enough, however, register variation receives rather short shrift:

“Since it is the case that all languages observable today or in the recent past show all the kinds of variation discussed here, we are entitled to conclude that such variation must be true of all languages that have ever been spoken, in all places, at all times. This principle cannot be tested, since linguistic evidence from the past (except the very recent past) comes only in written form, and such written evidence is incapable of showing more than a small fraction of the range of variation we assume to have existed. In particular, each piece of written evidence will typically reflect the formal register (because written) of a particular user of the language concerned.” (Penny 2000: 7)

It is of course the case that we have no primary evidence of spontaneous spoken language before the advent of tape recorders and transcribed oral corpora (for Spanish-language examples of the latter, see ‘Corpora’ under ‘References’). But it is surely over-simplistic to assume that the only registers worth speaking of are ‘spoken’ and ‘written’ register: that would at the very least imply that our concept of register is limited to the Hallidayan dimension of mode (Halliday 1978: 31–5). That the distinction between speech and writing has long been considered a basic one in linguistics is not challenged here. Rather, it is

proposed that the idea that spoken and written language represent the two fundamentally discrete registers, with regard either to situation or to linguistic features, has now been conclusively discredited (Biber 1988: 24–5, and for Old Castilian, Oesterreicher 1998 quoted in Jacob 2001: 157). Even more fundamentally, written language cannot simply be equated with formal register, since, as we have already seen, and as I will go on further to demonstrate, attempts are often made to represent features of spoken registers in written form: explicitly so in the inclusion of direct speech in narrative text and in written dialogues of various kinds, but more subtly so in the manipulation of register by creative literary writers. Even within written language there are many different registers, since written language, like all language, varies according to situation and purpose of use. This is crucial for historical linguistic research as it means that the absence of transcribed spoken language should not be a hindrance to the study of register. However, another interesting possibility offers itself: since we know that today creative writers often represent speech stereotypically, such data may serve as a convenient shortcut to those features of speech that are or were considered to be distinctive of colloquial registers.

The English-speaking world, in which the concept of register was coined, has not been slow to develop the study of register in historical linguistics. Romaine (1982) strenuously opposes the Labovian view that written texts are not a proper source of data. She advances the ‘uniformitarian principle’ that the past may be expected to resemble the present in terms of variation and that consequently the style of written texts may be expected to parallel social variation in the same way as it does today, thus allowing sociolinguistic reconstruction (pp. 122–6). Nevalainen and Raumolin-Brunberg (2003) conclude that “external factors such as genre differences can indeed be reconstructed for the past and that they correlate with processes of linguistic change” (p.192). English linguists have also stressed the importance of electronic corpora in making the best use of data (Nevalainen and Raumolin-Brunberg, 2003: 26–7). On the subject of corpora, it is interesting to see that there have been moves towards the creation of single-genre corpora which might be expected to facilitate the task of identifying register-based variation. Meurman-Solin (2001) introduces the notion of ‘philological computing’, which insists on the importance of the authenticity of texts and a proper account of their genre, which is demonstrated to be a non-trivial problem (to this I will return).

In Spanish, a small number of single-genre corpora exist for the modern language, but the identification of genre in digitally-encoded historical texts is more problematic. Mark Davies’s Corpus del Español, by far the fastest and the most sophisticated in terms of searches for syntactic data, permits rudimentary

discrimination of genre (literary, oral, 'text') only in 20th-century sources. The CORDE corpus of the Real Academia Española permits the filtering of results according to five different 'medios' as in novel, newspaper, magazine, oral, 'miscellaneous,' and no fewer than 362 different 'temas'; there is an assumption, however, that each textual source can be placed within one of each of these categories, which, as I have suggested and will further demonstrate, is not necessarily always appropriate. Another impressive corpus project, the Biblioteca Virtual Miguel de Cervantes at the University of Alicante, has concentrated on more straightforward issues of markup in the digitisation of texts (Bia and Pedreño 2001).

My own interest in register in the history of Spanish has stemmed from a number of projects I have undertaken in the area of historical syntax, which began as purely internal investigations. However, my involvement in the projects made me increasingly aware of the importance of register in understanding syntactic innovations — or capitalizations, as I would now prefer to call them (see Pountain 2000) — which appear to have been the result of learned contact with Latin. I have already referred to my work on adjective position (Pountain 1998a), in which I have argued that the example of the anteposing of qualificative adjectives in Latin encouraged the exploitation of a similar strategy in Castilian (not, however, totally novel, because of the availability of anteposing with quantifiers and 'common' adjectives of size such as *grande* and *pequeño*). Something of a similar story is to be told concerning the 'accusative and infinitive' construction with verbs of saying and thinking (Pountain 1998b), except that this ultimately fails to be embedded in any register. The role of register in such learned syntactic borrowing seems to me to be fundamental, since these two phenomena 'enter the language' in very particular literary registers which are the preserve of an educated elite. Arising out of such work, I have more recently (Pountain forthcoming) called attention to the process of conversion from *Abstand* to *Ausbau* status (Kloss 1967) in the history of Castilian, which is closely associated with concerns of register. In Kloss's terminology, an *Abstand* language is "a linguistic unit which a linguist would have to call a language even if not a single word had ever been written in it" (p.29); an *Ausbau* language, by contrast, is a language which has been "reshaped [...] in order to become a standardized tool of literary expression" (p.29). While register variation almost certainly already occurs in the spoken form of *Abstand* languages (cf. Dorian 1994), the repertoire of written registers dramatically increases when a language is 'elaborated' in its passage to *Ausbau* status, a process clearly charted for Castilian by Cano Aguilar (1997: 194–204): in this way, a high degree of written register variation is one of the chief characteristics of

an *Ausbau* language. Several articles in Jacob and Kabatek (2001) have shown the importance of this dimension of linguistic analysis for an understanding of the history of Spanish. Cano Aguilar (2001), comparing the syntax of Berceo and the *Alexandre*, concludes that apparent syntactic change is to be construed as a process of “elaboración lingüística... que genera una tradición textual en castellano medieval, tradición inexistente hasta entonces, pero que iba a continuar” (2001:149), and that such elaboration is in all probability due to Latin or French models. Jacob (2001), in a study of the *haber* + past participle construction, shows the importance of relating apparent syntactic change to the study of text type, concluding:

“Hay que insistir en el hecho de que esta sucesión de ‘estadios’, en tan breve espacio de tiempo como aparece en los textos del siglo XIII, no es el espejo fiel de una evolución diacrónica, sino el reflejo de una situación en la cual la lengua española sólo estaba en vía de establecerse y afirmarse como lengua escrita, todavía en busca de los géneros y de las formas de expresión apropiados a las nuevas tareas textuales y pragmáticas...” (Jacob 2001:171)

It is the process of standardisation which inevitably accompanies an *Ausbau* language that ensures the possibility of the downward migration of features which are originally peculiar to the more specialised written registers. As the vehicle of a literary culture, an *Ausbau* language may also encourage literary writers to display versatility in the communicative competence with which they command a wide range of the available registers of the language, leading naturally to the creation of multi-register texts. The data of Romance historical linguistics has traditionally been the latter. While the stylistic heterogeneity of such texts has more recently led to the exploration of non-literary sources, and this is certainly likely to constitute a major direction for the investigation of register in the history of Spanish, it is probably the analysis of multi-register texts which gives the best chance of accessing the elusive spoken language, which is otherwise not overtly represented.

3. Data and methodology

To show both the potential of a register-sensitive approach to historical data and some of the associated problems, I have examined the Second Part of *Corbacho*, some 18,712 words of a mid-15th-century text which is well known for its apparent representation of spoken language. Admittedly, this is not a large text, but it is sufficient to constitute a pilot study which will illustrate the issues I want to raise.

The text is relatively easy to mark up according to discourse types, and the following categories were distinguished:

<cap>	chapter numbers	28 words
<cit>	quotations from classical authors	277 words
<citbib>	quotations from the Bible	36 words
<citprov>	<i>Refranes</i>	48 words
<comm>	commentary, usually consisting of moral observations and generalisations	10,956 words
<dir>	direct speech not attributable to males or females	6 words
<dirfem>	direct speech of females	4,831 words
<dirmale>	direct speech of males	578 words
<end>	Ending	10 words
<heading>	chapter headings	146 words
<indirfem>	indirect female speech	5 words
<intro>	introduction to the Second Part	32 words
<narr>	narration of incidents and <i>exempla</i>	1,434 words
<quote>	quotations not attributed to any author	9 words
<rep>	introductions to direct speech	316 words

A brief word on the criteria used in markups like this is nevertheless appropriate since objectivity in such matters is a notorious problem in corpus linguistics. Although I was inspired by some of the standard markup tags of the Text Encoding Initiative (Sperberg-McQueen and Burnard 2004), it seemed to me that, from the point of view of this study, a greater degree of both delicacy in some respects and generalisation in others was appropriate. It is actually an interesting question as to whether ‘universal’ categories of discourse function could eventually be proposed, or whether discrimination is of necessity done on such an ad hoc basis. The general opinion seems to be that the latter is a more likely scenario:

“Neither text linguistics nor LSP [Language for Special Purposes] claims to have more than a very rough and ready classification of text or discourse types. Few people in those fields today hold high hopes that a rigorous typology will appear soon, and fewer still would insist that research cannot proceed without one.” (Beaugrande 1993: 16)

The discourse types identified in *Corbacho* are the following. Quotations (<cit>, <quote>) are clearly formally indicated with introductory material such as “dixo Salomón”, “dise el enxienplo bulgar”. Organizational features of the text, such as chapter numbers (<cap>), chapter headings (<heading>), introductory

material (<intro>) and end material (<end>) are also easily recognizable on formal grounds. Direct and indirect speech (<dir...> and <indir...>) have the usual formal linguistic indications. The only really subjective judgement I made was the classification of 'narrative' (<narr>) and 'commentary' (<comm>), though the distinction seems to relate quite obviously to the field of discourse, 'narrative' being essentially storytelling and 'commentary' being the author's generally moralising observations. Some other markup was made on the basis of a provisional assessment of linguistic relevance. It seemed desirable to distinguish the quoting of *refranes*, which often have an elliptical syntax and a rhyming structure (<citprov>), from the quoting of classical authors (<cit>), which may well owe something to their (usually Latin) original, and quotations from the Bible (<citbib>); yet such quotations, based on well-established intertextual reference, are quite unequivocal. I also distinguished introductions to direct speech (<rep>) from other narrative material, since I anticipated that not to do so could skew statistics concerning the frequency of verbs of saying, which are field-dependent, since these verbs occur in almost all introductions to direct speech, but are otherwise not so frequent in narrative.

The apparent ease of carrying out such an encoding should not, however, lead us to assume that such categories will ultimately be found to represent consistent examples of discrete registers. I want to concentrate on four of the discourse categories I have proposed: <comm>, <narr>, <dirfem> and <dirmale>. At first sight, <comm> and <narr> may be taken to represent written language shorn of the genre-specific material which constitutes <cap>, <quote>, <cit>, <citbib>, <citprov>, <heading>, <intro>, <end> and <rep>, while <dirfem> and <dirmale> are representations of the spoken language. The following questions would seem to be approachable: (a) to what extent are these discourse-determined categories distinguished in terms of linguistic features, and (b) which of these linguistic features may be taken as indicators of register.

The distinctiveness of the categories may immediately be seen by comparing some raw statistics for <comm> and <dirfem>, the most substantial samples. (Analysis was carried out on a digitised version of the González Muela edition of the text using WordSmith Tools.) They are discriminated by word-length, <dirfem> having a generally higher percentage of shorter words than <comm> (see Figure 1).

The average sentence-length of <comm> is 29.81 words as against 10.16 words for <dirfem>, although judgements about what constitutes a sentence follow the edition of González Muela (1970), which is punctuated according to modern custom. A more significant statistic concerning sentence length,

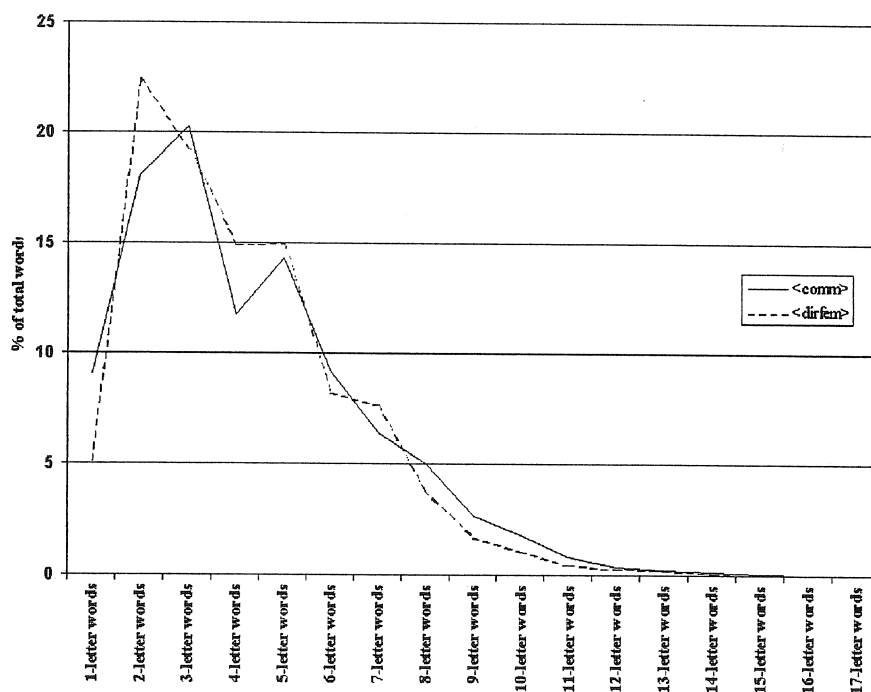


Figure 1. Word-length in *Corbacho*

but one which it is not possible to obtain automatically, might be the number of constituent clauses per 'sentence', as thus defined, and the length of the individual constituent clauses, a clause being defined by the presence of a verb. Figure 2 shows such statistics, obtained manually, for the first chapter of our corpus: it can be seen that, while the differential average sentence length is comparable in this sample, the two registers also have different profiles with regard to clause constituency, <dirfem> having a lower number of clauses per sentence and a lower average clause length.

One of the reasons for the difference in average clause length seems to be the higher incidence of enumeration in <comm>, a rhetorical device still associated especially with written rather than spoken register, and of comparative structures; both of these phenomena are illustrated in the following example:

- (1) Por quanto en aquello ponen su corazón e voluntad, mas non en el provecho de su casa, estado, e honra, synón en vanidades e locuras, e en cosa de poca pro. (*Corbacho*, p.128)

The difference in average sentence length is partly a function of clause length, although it is also related to the greater syntactic complexity often observable

clauses per 'sentence'	<comm>		<dirfem>	
	number of sentences	total clauses	number of sentences	total clauses
1	8	8	64	64
2	9	18	24	48
3	4	12	11	33
4	4	16	9	36
5	3	15	3	15
6	3	18	3	18
7	2	14	0	0
8	1	8	0	0
9	0	0	0	0
10	0	0	0	0
11	0	0	0	0
12	0	0	0	0
13	1	13	0	0
total	35	122	114	214
words		933		1166
av. sentence length (overall av. sentence length)		26.66 (29.81)		10.23 (10.16)
av. clauses per sentence		3.49		1.88
av. clause length		7.65		5.45

Figure 2. Sentences and clauses in <comm> and <dirfem> from Chapter 1 of *Corbacho*

in <comm>. Extract (2), for example, contains several subordinate clauses: a conditional clause introduced by *sy*, the subject complement of *es dubda* and the object complement of *espera*:

- (2) E por esta razón de avaricia muchas de las tales ynfinitos e diversos males cometen; que sy dineros, joyas preciosas e otros arreos yntervenga[n], o dados les sean, es dubda que a la más fuerte non derruequen, e toda maldad espera que cometrá la avariciosa muger con defrenado apetito de aver, asý grande como de estado pequeño. (*Corbacho*, p.121)

However, complex subordinating syntax is not necessarily revealed by sentence length, since sentences can also be extended by coordination, as is evident in (2) in the case of the clause beginning *que sy dineros...*

The difference between the two categories is also made plain by what we might see as pragmatically-determined features. Exclamations, characterised by the absence of a main verb as in

- (3) ¡Ay, triste, desconsolada! ¡Quántos meneos por non nada! ¡O qué linda galana! (*Corbacho*, pp.125, 140 and 142)

are common in the modern spoken language, and it is no surprise to find that, in this text, both <dirfem> and <dirmale> contain an overwhelmingly higher number than <comm> and <narr>, as shown in Figure 3. Ellipsis (in the specific sense of the absence of a main verb from a sentence), which is another well-known feature of spoken register, is also well represented in <dirfem> (13 examples) as opposed to just one in <comm>.

These preliminary figures should give us confidence in pursuing these data further in search of linguistic indicators of register. I have established what seems to be significant variation with regard to the features shown in Figure 4.

Several of these features would be entirely expected, again for pragmatic reasons. They are also comparable to features of different registers in the modern language, which can be observed more empirically. A higher frequency of logical connectors and greater 'exactness' in their use is likely to be demanded pragmatically by expository prose, where a coherent argument is being developed, so it is not surprising to find that such items are most frequent in <comm>, and in fact discriminate <comm> from <narr>, as well as from <dirfem> and <dirmale>. Conversely, the use of the 'vague' logical connector *que* is more frequent in speech, as it is today. *Pues* shows a frequency in <dirfem> comparable to that of the modern spoken language (in the Esgueva and Cantarero corpus it has an incidence of 8.92 per 1,000 words), but it is also characteristic of <comm>, and distinguishes <comm> sharply from <narr>. A number of the grammatical features identified (absolute constructions, the 'accusative and infinitive' construction, qualificative adjectives anteposed to the noun and verb-final word order) are readily associable with 'learned' Latin syntax and hence more likely to be a feature of the cultured written language.

So far, this would all seem to be unexceptional and, at the same time, would appear to represent a convenient demonstration of the recoverability of fifteenth-century register variables; but anyone who is familiar with *Corbacho* will be feeling, rather uneasily, that things cannot be so straightforward; after all, Martínez de Toledo is renowned amongst literary critics for what is usually described as his 'mixing' of styles. González Muela, in the introduction to his edition, writes:

"Es un estilo de compromiso: tiene algo de "arte" refinado, algo del decir y sentir del pueblo, y algo (quizá mucho) de genial creación personal." (González Muela 1970: 20)

<dirfem>	<dirmale>	<comm>	<narr>
70 (8.42)	12 (25.00)	4 (1.10)	0 (0)

Figure 3. Exclamations in *Corbacho* (total occurrences followed by percentage of total number of sentences per register)

Feature	<dirfem>	<dirmale>	<comm>	<narr>
Logical connectors				
<i>así que</i>	0 (0)	0 (0)	7 (0.64)	0 (0)
<i>mas</i>	2 (0.41)	0 (0)	21 (1.92)	3 (2.2)
<i>pero</i>	3 (0.62)	2 (3.47)	35 (3.19)	2 (1.47)
<i>por ende</i>	0 (0)	0 (0)	33 (3.01)	1 (0.73)
<i>por quanto</i>	0 (0)	0 (0)	18 (1.64)	2 (1.47)
<i>pues</i>	51 (10.55)	1 (1.73)	33 (3.01)	1 (0.73)
<i>porque</i>	1 (0.21)	2 (3.47)	13 (1.19)	1 (0.73)
<i>yten</i>	0 (0)	1 (1.73)	5 (0.46)	0 (0)
Grammatical features				
'accusative and infinitive'	0 (0)	0 (0)	7 (0.64)	0 (0)
infinitive with overt subject	0 (0)	0 (0)	16 (1.46)	1 (0.73)
Word order				
Separation or inversion of auxiliary and infinitive	0 (0)	0 (0)	21 (1.92)	0 (0)
verb-final order	6 (1.24)	1 (1.73)	69 (6.30)	0 (0)
Other				
<i>fazer fazer</i>	0 (0)	0 (0)	3 (0.27)	1 (0.70)
<i>ello</i>	1 (0.21)	2 (3.47)	35 (3.19)	1 (0.70)
<i>por adj que + subjunctive</i>	0 (0)	0 (0)	3 (0.43)	0 (0)
use of <i>es / ay</i> as an existential	0 (0) /	0 (0) /	18 (0.73) /	0 (0) /
	3 (0.62)	0 (0)	23 (2.10)	6 (4.18)
qualificative adjective preceding the noun	0 (0)	1 (1.73)	57 (5.20)	4 (2.79)
conjunctive <i>que</i>	56 (11.58)	6 (10.4)	34 (3.10)	0 (0)
adverbs in <i>-mente</i>	2 (0.41)	0 (0)	36 (3.29)	4 (2.79)

Figure 4. Features sharply differentiating register in *Corbacho* (total occurrences followed by frequency per thousand words in brackets)

By contrast, what I have been suggesting so far is that the elements in this compromise are fairly clearly differentiated.

However, quantitative comparison can also reveal when linguistic features are not distinguished in the way we might have expected on the basis of other evidence.

Feature	<dirfem>	<dirmale>	<comm>	<narr>
negative-clitic / clitic-negative	103 (21.30) / 1 (0.21)	10 (17.33) / 0 (0)	46 (4.20) / 7 (0.64)	1 (0.70) / 1 (0.70)
<i>ge</i> / <i>se</i> + other clitic	3 (0.62) / 10 (2.07)	1 (1.73) / 0 (0)	7 (0.64) / 10 (0.91)	1 (0.70) / 0 (0)
<i>ser</i> -passive	17 (3.52)	2 (3.47)	55 (5.02)	7 (4.89)
'adverbial' gerund (i.e., excluding periphrastic verbal constructions)	17 (3.52)	0 (0)	59 (5.39)	18 (12.56)
absolute constructions	3 (0.62)	0 (0)	6 (0.55)	7 (4.89)
Relatives				
<i>que</i>	40 (8.27)	6 (10.4)	141 (12.87)	13 (9.-7)
<i>quien</i> , etc	4 (0.83)	1 (1.73)	13 (1.19)	0 (0)
<i>el qual</i> , etc	0 (0)	0 (0)	5 (0.46)	0 (0)
<i>cuyo</i> , etc	0 (0)	0 (0)	1 (0.09)	0 (0)
<i>el que</i> , etc	18 (3.72)	2 (3.47)	67 (6.12)	4 (2.79)
Total	62 (12.82)	9 (15.6)	227 (20.72)	17 (11.85)

Figure 5. Features which are undifferentiated in *Corbacho* (total occurrences followed by frequency per thousand words in brackets)

There is clearly some variation at this time in the order of clitics and the negative. The insertion of a negative between clitics and the verb seems to have been a feature of some northern Old Castilian texts (Pountain 2001: 87–8). It is possible, therefore, that it covaried as an indicator of register. In all the discourse types in which it appears, it is heavily outnumbered by examples of the negative preceding the clitic sequence, which would suggest that it is in fact archaic, a change that is nearly completed — just the kind of feature to which a strong subjective attitude might be taken by speakers. However, although it is more frequent in <comm> and <narr> than in <dirfem> and <dirmale>, it is attested in <dirfem>, and not in a context which suggests that it has any special stylistic marking:

- (4) ¡...que quantas cerraduras le puso e tanto me vedó que la non abriese!
(*Corbacho*, p.153)

The converse possibility, that it is a feature of colloquial usage and that Martínez de Toledo is using it to introduce a colloquial tone in more formal register, would not be inconsistent with some of the instances in <comm>; in the following example, the feature appears in a clause representing the female subject's behaviour (*ella faze que ge lo non prescia*), which could be interpreted as being a kind of indirect speech ("ge lo non prescia") mimicking her:

- (5) Dízele todo el secreto; ella faze que ge lo non prescia, nin le plaze oyrlo
(*Corbacho*, p.150)

To explore this matter further, I employed another technique, that of using a contemporaneous source to act as a control. I took a sample consisting of the first 5,775 words of Pérez de Guzmán's *Generaciones y semblanzas*, which may be taken as a more consistently formal text (cf. González Muela, pp.25–6). In this text, there are 12 examples of negative-clitic order as against two of clitic-negative order, figures broadly comparable with those obtained for *Corbacho* (Figure 6). The inescapable conclusion is therefore that this feature is not in fact a register indicator. A similar account could be given of the variation that is observable between *ge* and *se* in clitic sequences. The next three features in Figure 5 (the *ser*-passive, the 'adverbial' gerund and the absolute construction) are all features which might be thought of as attributable to learned Latin influence and thus more likely to be typical of formal written language (see Blatt, 1957, and Alvar and Mariner, 1967). They are all similarly attested in *Generaciones y semblanzas* (see Figure 6).

The most interesting of these features is the *ser*-passive, since we know that this is, or has become, extremely infrequent in spoken register today. There are just 15 occurrences in a 134,500 word corpus of spontaneous Madrid speech of the 1970s (Esgueva and Cantarero 1981). In tracing the history of the construction in the Romance languages, it would be very interesting to know if it has always had this register value or whether in the Middle Ages it was a more common feature of speech. At first sight, the evidence of *Corbacho* would suggest that it was not an indicator of register in the same way as it is today. But,

negative-clitic / clitic-negative	12 (2.08) / 2 (0.34)
<i>ge</i> / <i>se</i> + other clitic	1 (0.17) / 2 (0.34)
<i>ser</i> -passive	36 (6.23)
'adverbial' gerund	40 (6.93)
absolute constructions	3 (0.52)
Relatives	
<i>cuyo</i> , etc.	5 (0.87)
<i>el qual</i> , etc.	26 (4.50)
<i>que</i>	99 (1.71)
<i>quien</i> , etc.	5 (0.87)
<i>el que</i> , etc.	21 (0.36)
Total relatives	156 (27.01)

Figure 6. Some features of *Generaciones y semblanzas* (total occurrences followed by frequency per thousand words in brackets)

as I have pointed out elsewhere (Pountain, forthcoming), closer inspection of these examples reveals that 8 of the 17 examples in <dirfem> and <dirmale> are blessings or curses following the model of *¡maldito/bendito sea!* and may be considered formulaic expressions (all but one show inversion of the auxiliary and past participle), and five others are the passives of stative verbs in which the past participle might be considered to have a purely adjectival value (e.g. *soy desfavorecida; perdido soy*). It is thus possible to sustain the hypothesis that the *ser*-passive is likely to have been more frequent in written register at this time, even though its occurrence in the spoken language was not as significantly constrained as today.

We cannot make such claims for the ‘adverbial’ gerund, which seems to be well established in <dirfem>. As far as absolute constructions are concerned, the figures are not statistically significant beyond revealing that they do appear to be used in <dirfem>, apparently quite naturally.

It is a pity that the statistics concerning relatives do not allow firmer conclusions to be drawn, since *cuyo* and *el cual*, etc., have a very low frequency in modern speech and can today certainly be considered indicators of formal written register. *Corbacho* has a very low incidence of *cuyo* and *el qual* overall by comparison with many other Old Castilian texts (cf. the comparisons with *Generaciones y semblanzas* given in Figure 7), and this may be an indication of the presence in <comm> of some ‘popular’ features.

(*Cuyo* is absent from some Old Castilian texts, most notably the *Poema de Mio Cid* and the *Santa María Egipcíaca*, though it is attested in Berceo and becomes more frequent in 15th-century legal texts. A reason for the infrequency of *cuyo* in <comm> may be the frequency of what may be described as the use

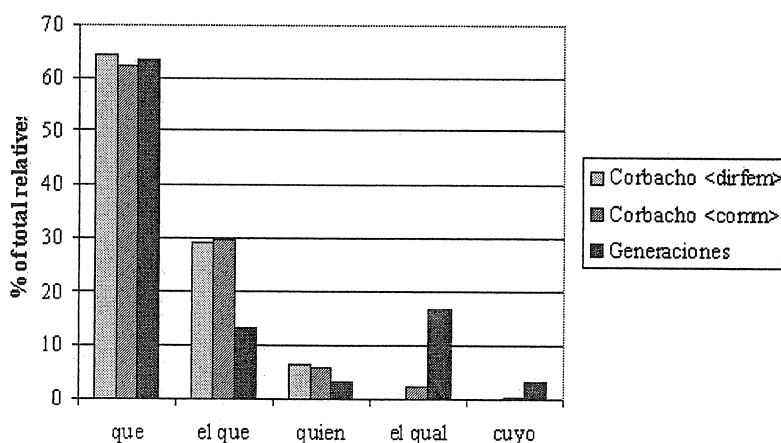


Figure 7. Relatives in *Corbacho* and *Generaciones y semblanzas*

of a 'resumptive' possessive in the relative clause; in (6), the relative clause is detached from its antecedent *cosas* (with double underlining) and the use of the possessive *sus* (with single underlining) makes the reference to *cosas* clearer. In Modern Spanish, this would be considered unacceptable by purists, though it continues to be found in spontaneous spoken language.

- (6) Así que muchas cosas tyenen buenos comienços que sus fines son diversos (*Corbacho* <comm>, p.142)
(cf. MSp. Así que muchas cosas cuyos fines son diversos tienen buenos comienços)

The register value of this 'resumptive' construction is difficult to judge in *Corbacho*, although, since it does not appear in <dirfem> or <dirmale>, we might conclude that Martínez de Toledo did not associate it particularly with speech. Its use in <comm> certainly facilitates the expression of complex comparative observations in which this register is rich; the following example shows the use of a resumptive possessive (*sus*) and a resumptive pronoun (*las*), both of which relate to the antecedent *aquellas*:

- (7) Desta regla las monjas son maestras, e dezir dellas en particular non conviene, pues mugeres son, e so la regla dellas se comprenden: las buenas como buenas e honestas religiosas loando, e la malas, sy las ay, como aquellas que sus fechos las fazen malas, reprovando. (*Corbacho* <comm>, p.145)

4. Conclusions

What I hope to have suggested in this paper is that it is feasible to investigate register variation in the history of Spanish, and that this task is an urgent one, given the good progress that has already been made in English linguistics. We increasingly have at our disposal the electronic tools that permit the collection and analysis of the large amounts of data that such an enterprise necessarily involves. The experience of looking at *Corbacho* shows, however, that such data will need critical philological markup and interpretation before it can appropriately serve for register analysis. It reveals the danger of considering that an individual text belongs to one particular genre which can be equated with register, and hence the questionability of the notion of a single-genre corpus. Admittedly, not all texts are likely to be as complex as this one; but, as I hope to have shown, it may be precisely such texts, judiciously interpreted, which can offer us an insight into the more elusive registers of the history of the language.

Although I do not have clear answers to all the particular questions I have raised, I hope to have shown that Martínez de Toledo's prose does provide evidence of contemporary register variation and that this variation can be successfully quantified. There will of course always be lacunae in historical data, but this should not impede tentative conclusions, duly acknowledged as such, being drawn. I have also suggested how data concerning register variation in the modern language and the use of 'control' texts may be used to assist in the enterprise.

References

- Alvar, Manuel & Sebastián Moliner. 1967. "Latinismos". *Enciclopedia lingüística hispánica* ed. by M. Alvar, A. Badía, R. de Balvín and F. Lindley Cintra, II, 3–49. Madrid: Consejo Superior de Investigaciones Científicas.
- Beaugrande, Robert de. 1993. "'Register' in discourse studies: A concept in search of a theory". *Register analysis: Theory and practice* ed. by Mohsen Ghadessy, 7–25. London and New York: Pinter.
- Bia, Alejandro & Andrés Pedreño. 2001. "The Miguel de Cervantes Digital Library: The Hispanic voice on the web". *Literary and linguistic computing* 16(1). 161–77.
- Biber, Douglas & Edward Finegan. Eds. 1994. *Sociolinguistic perspectives on register*. Oxford/New York: Oxford University Press.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Blatt, Franz. 1957. "Latin influence on European syntax". *Travaux du Cercle Linguistique de Copenhague* 5. 33–69.
- Cano Aguilar, Rafael. 1997. *El español a través de los tiempos*, 3rd ed. Madrid: Arco.
- Cano Aguilar, Rafael. 2001. "La construcción del discurso en el siglo XIII: Diálogo y narración en Berceo y el *Alexandre*". *Lengua medieval y tradiciones discursivas en la Península Ibérica. Descripción gramatical-pragmática histórica-metodología* ed. por Daniel Jacob and Johannes Kabatek, 135–51. Madrid: Vervuert.
- Dorian, Nancy C. 1994. "Variation in a language restricted to private-sphere use". *Sociolinguistic perspectives on register* ed. by Douglas Biber and Edward Finegan, 217–32. Oxford/New York: Oxford University.
- Ferguson, Charles A. 1994. "Dialect, register, and genre: Working assumptions about conventionalization". *Sociolinguistic perspectives on register* ed. by Douglas Biber and Edward Finegan, 15–30. Oxford/New York: Oxford University Press.
- Halliday, M.A.K. 1978. *Language as social semiotic. The social interpretation of language and meaning*. London: Arnold.
- Jacob, Daniel & Johannes Kabatek. Eds. 2001. *Lengua medieval y tradiciones discursivas en la Península Ibérica. Descripción gramatical-pragmática histórica-metodología*. Madrid: Vervuert.

- Jacob, Daniel. 2001. "¿Representatividad lingüística o autonomía pragmática del texto antiguo? El ejemplo del pasado compuesto". *Sociolinguistic perspectives on register* ed. by Douglas Biber and Edward Finegan, 153–176. Oxford/New York: Oxford University.
- Keniston, Hayward. 1937. *The syntax of Castilian prose. The sixteenth century*. Chicago: University of Chicago Press.
- Kloss, Heinz. 1967. "Abstand-Languages and Ausbau-Languages". *Anthropological Linguistics* 9(7). 29–41.
- Meurman-Solin, Anneli. 2001. "Structured text corpora in the study of language variation and change". *Literary and linguistic computing* 16(1). 5–27.
- Milroy, James. 1992. *Linguistic variation and change. On the historical sociolinguistics of English*. Oxford: Blackwell.
- Nevalainen, Terttu & Helena Raumolin-Brunberg. 2003. *Historical sociolinguistics*. London: Longman.
- Oesterreicher, Wulf. 1998. "Pragmática del discurso oral". *Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza*, vol. 4 ed. por Giovanni Rufino, 203–12. Tübingen: Niemeyer.
- Penny, Ralph. 2000. *Variation and change in Spanish*. Cambridge: Cambridge University Press.
- Pountain, Christopher J. 1998a. "Nuevo enfoque de la posición del adjetivo atributivo". *Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza*, vol. 2 ed. por Giovanni Rufino, 697–708. Tübingen: Niemeyer.
- Pountain, Christopher J. 1998b. "Learnèd syntax and the Romance languages: The "accusative and infinitive" construction with declarative verbs in Castilian". *Transactions of the Philological Society*, 96(2). 159–201.
- Pountain, Christopher J. 2000. "Capitalization". *Historical linguistics 1995, volume 1: General issues and non-Germanic languages* ed. by John Charles Smith and Delia Bentley, 295–309. Amsterdam: John Benjamins.
- Pountain, Christopher J. 2001. *A history of the Spanish language through texts*. London: Routledge.
- Pountain, Christopher J. forthcoming. "Syntactic borrowing as a function of register". *Languages in Contact* ed. by Arturo Tosi and Laura Lepschy. Oxford: Legenda.
- Reid, T.B.W. 1956. "Linguistics, structuralism and philology". *Archivum Linguisticum* 8. 28–37.
- Romaine, Suzanne. 1982. *Socio-historical linguistics: Its status and methodology*. Cambridge: Cambridge University Press.
- Ruffino, Giovanni. Ed. 1998. *Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza*. Tübingen: Niemeyer.
- Sperberg-McQueen, C.M. & Lou Burnard. Eds. 2004. *TEI P4: Guidelines for electronic text encoding and interchange*. Text Encoding Initiative Consortium. XML Version: <http://www.tei-c.org/P4X/>.

Corpora

- Biblioteca Virtual Miguel de Cervantes*. University of Alicante: www.cervantesvirtual.com.
- CORDE = *Corpus Diacrónico del Español*. Real Academia Española www.rae.es/cordenet/html.
- Corpus Oral de Referencia del Español Contemporáneo, Laboratorio de Lingüística Informática*. Departamento de Lingüística, Lenguas Modernas, Lógica y Filosofía de la Ciencia, Universidad Autónoma de Madrid.
- Davies, Mark. *Corpus del Español*. www.corpusdelespanol.org.
- Samper Padilla, José Antonio Clara Eugenia Hernández Cabrera & Magnolia Troya Déniz. *Macrocorpus de la norma lingüística culta de las principales ciudades del mundo hispánico (MC-NLCH)*. 1998. Las Palmas de Gran Canaria: Servicio de Publicaciones de la Universidad de Las Palmas de Gran Canaria.

Texts

- González Muela, J. 1970. *Arcipreste de Talavera o Corbacho*. Edición. Madrid: Castalia.
- Esgueva, M. & M. Cantarero. 1981. *El habla de la ciudad de Madrid*. Madrid: Consejo Superior de Investigaciones Científicas.
- Rico, Francisco. 1987. *Lazarillo de Tormes*, Edición. 2nd ed.. Madrid: Cátedra.
- Tate, R.B., 1965. *Fernán Pérez de Guzmán, Generaciones y semblanzas*. Edición crítica. London: Tamesis.

Author's address

Christopher John Pountain
School of Modern Languages
Queen Mary, University of London
Mile End Road
London E1 4NS
United Kingdom
c.j.pountain@qmul.ac.uk