

- MacClure, Erica y Malcolm MacClure. 1988. Macro- and micro-sociolinguistic dimensions of code-switching in Vingard. Code-switching. Anthropological and Sociolinguistic Perspectives, ed. by Monica Heller, 34-47. Berlin: Mouton de Gruyter.
- Myers-Scotton, Carol. 1993. Social Motivation for Code-switching. Evidence from Africa. Oxford: Clarendon Press.
- Ninyoles, Rafael. 1969. Conflicte lingüístic valencià. València: Tres i Quatre.
- Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish y termino en español: towards a typology of code-switching. Linguistics 18 (7/8). 581-618.
- Pueyo, Miquel. 1994. Llenguas en contacte en la comunitat. València: Servei de Publicacions de la Universitat de València.
- Reyes, Rafael. 1976. Language Mixing in Chicano Bilingual Speech. Studies in Southwest Spanish, ed. by J.D. Bowen & J. Ornstein, 182-8. Rowley: Newbury House.
- Singh, R. .1983. We, They, and Us A note on Code-switching in Linguistic Contacts. Aspects of Bilingualism, de. By M. Paradis .ed. Columbia, SC, Hornbeam Press, 265-75.
- Tabouret-Keller, André. 1983. Switching from a Psychological Point of View. International Journal of the Sociology of Language, 39. 139-49.
- Tannen, Deborah. 1989. Talking voices. Repetition, dialogue, and imagery in conversational discourse. Cambridge: Cambridge University Press.
- Woolard, Katherin. 1988. Codeswitching and comedy in Catalonia. Codeswitching. Anthropological and Sociolinguistic Perspective, ed. by. Monica Heller, 53-74. Berlin: Mouton de Gruyter.

A CORPUS-BASED ANALYSIS OF SUBJECT RAISING IN MODERN SPANISH

Mark Davies

Illinois State University

This study of Spanish subject to subject raising (*María parece estar enferma* vs. *parece que María está enferma*) is based on a 7,700,000 word computer corpus of spoken and written Spanish, as well as acceptability judgments from nearly 150 native speakers. It provides data and motivations for a number of previously unexplained aspects of the Spanish raising construction. For example, the much higher degree of raising in written Spanish is due to universal strategies concerning the production of written and spoken language, as well as semantic-pragmatic factors such as topicality and the attributive/impersonal meanings of *parecer*. These factors are even more important in cases of non-third person subjects (*?parezco estar enfermo*) and 'partial raising' (*parecen que están enfermos*). Factors such as agentivity influence whether the indirect object raising construction will be used with certain embedded verbs (*me parece oír su voz / ?me parece cortar la leña*). Finally, the avoidance of main clause clitics (*?María me parece estar enferma, *ella me parece conocer*) are due to general functional principles such as the avoidance of ambiguity.

1. INTRODUCTION

Subject to subject raising, which is the movement of a lower clause subject into the main clause (1a to 1b)¹, has been the topic of many studies from within the generative framework as well as other syntactic models.

- (1a) ___ parece [que María está enferma]
NON-RAISED
it seems [that Mary is sick]
- (1b) María parece [___ estar enferma]
RAISED
Mary seems [___ to be sick]

However, while the English raising construction has received a great deal of attention, subject raising in Spanish has been studied much less and has been the main focus of only three previous studies. Bolinger (1972) is concerned in large part with the possible diachronic evolution of the *parecer* construction (such as its lexical development), and also points out in a preliminary way some interesting differences between the English and Spanish construction. Both García Pinto and Luján (1974) and Fernández Leboráns and Díaz Bautista (1992) are attempts to account for the 'standard' cases of raising (in (1a-b)) from within an autonomous, generative model of syntax. In addition to these three studies, a few other researchers have looked at subject raising in passing, but only as a small part of more general works on clausal structure in Spanish (Sauer 1972, Nuessel 1973, Ruíz Morales 1979, Lieó 1979, Seaver 1982, Subirats-Ruggeberg 1987).

There are still many questions regarding the construction, however, that have not been adequately answered in previous studies, and in many cases have not even been asked. In most cases this is due to the fact that there have been no data-based analyses, and nearly all of the studies have been written from within one particular syntactic framework. Some of the unexplored questions and issues are the following. In Section 2 we will see that the relative frequency of raised / non-raised sentences of the type *parece que María está enferma* vs. *María parece estar enferma* varies greatly between written and spoken Spanish, and we will suggest that this difference is due to both semantic and pragmatic factors as well as general differences in the way that written and spoken language are produced. In Section 3 we will consider why, unlike English, subject raising in Spanish is much less acceptable with non-third person subjects (*#yo parezco/tú pareces haber ofendido a Elena*), and we will investigate how the previously mentioned semantic and pragmatic factors relate to these marginal constructions. In Section 4 we will consider 'partial raising' sentences of the type *ellos parecen que la conocen*, and will see that general factors relating to the production of spoken language may explain why most native speakers consider these unacceptable, even though they are rela-

tively common in spoken corpora of Spanish. In Section 5 we will examine the 'indirect object' raising construction (*me parece oír su voz*), and see how general pragmatic factors explain the unacceptability of this construction with certain types of embedded verbs (*?*me parece haber matado a Elena por su dinero*). Finally in Section 6 we consider how functional principles may explain the unacceptability that results with main clause clitics in sentences of the type *?*María me parece estar enferma* and **María me parece conocer*.

In order to examine the different types of Spanish subject raising constructions just outlined, we will rely on a large number of tokens in a corpus of both spoken and written Spanish, as well as the acceptability judgments of a large number of native speakers. The more than 3600 examples are taken from one of the largest corpora of Modern Spanish presently available, which is composed of more than 7,700,000 words from both spoken and written Spanish, taken from 1350 novels, short stories, textbooks, and conversations. Table 1 gives a summary of the types of texts used, and a more complete listing can be found in Note 2.

TABLE 1
Composition of the corpus

	# texts / conversations	# words	# tokens
Spoken	955	3,469,000	1895
Habla - LatAm	385	2,193,000	1067
Habla - Spain	72	328,000	271
Oral - Spain	498	948,000	557
Written	395	4,294,000	1778
Novels	15	1,327,000	622
Short Stories	356	1,054,000	597
Argentina	22	1,913,400	559
Total	1350	7,763,000	3673

As Table 1 indicates, the computer corpus is based on materials from a number of sources. The 2,500,000+ word *Habla Culta* corpus represents transcripts of more than 450 conversations with native speakers from nine Latin American countries and from Spain, which were all re-

corded within the past 30 years. The *Oral-Spain* corpus is based on transcriptions of 498 conversations from just Spain (recorded within the past ten years), and represents nearly 1,000,000 additional words of text. The fifteen novels, which represent more than 1,300,000 words of text, are from both Latin America and Spain and were all written after 1950. The short stories are taken from collections of short stories from the same eleven countries as the *Habla Culta* and were all written by authors born after 1900. Finally, the nearly 2,000,000 word corpus from Argentina is composed mainly of non-fiction materials, but includes some other genres as well.

To extract the 3600+ cases of the *parecer* construction, we first needed to have all of the different corpora in electronic form. The *Habla Culta* materials and the corpora of short stories and novels were scanned into the computer, whereas the spoken texts from Spain and the Argentina materials were downloaded from the Internet. We then used *WordCruncher* to create an every-word index of the corpus, which permitted us to perform complex proximity and Boolean searches on the data. Proximity searching meant that we could find every case of *parecer* immediately followed by either an infinitive or *que*. Boolean searching meant that we could combine and limit searches, such as all cases of *parecer* followed by an infinitive where the infinitive was not *ser*, *estar*, or *haber*. Needless to say, without the computer corpus and *WordCruncher*, extracting the 3600+ tokens from the 7,700,000+ word corpus would have been virtually impossible.

In addition to the computer corpus, we also used comments and acceptability judgments from 146 native speakers of Spanish. An introductory survey was sent to eighteen different newsgroups on the Internet that deal with Hispanic culture (e.g. *soc.culture.peru.soc.culture.spain*). 146 speakers from many different countries responded, including a majority who still lived in the Spanish-speaking country of their birth. In this survey speakers were asked to rate with a score of 1-5 eleven different *parecer* sentences, and to add any relevant comments concerning the sentences. In a second survey, to which 82 of the speakers responded, the same speakers were asked to simi-

larly rate another thirteen sentences that followed up on some of the more problematical aspects of the raising construction. In a third and final survey, to which there were 59 responses, the speakers were asked to provide very brief comments on four sentences of the type *ellos parecen que la conocen*. The speaker's judgments as well as their comments provided important insight into a number of the more problematic issues, especially those dealing with relevant pragmatic factors that affected acceptability.

In addition to providing insight into a number of subject raising phenomena described above, this study will also hopefully be of value from a methodological standpoint. While there have been some studies of Spanish syntax that are based on large computer-based corpora, the corpus employed here is both much larger as well as more diverse, in terms of the range of both written and spoken texts. The expectation is that corpora such as this one can begin to accurately model syntactic phenomena in a number of different registers. In addition to the corpora, this study also suggests that the Internet can be used to obtain valuable data. As shown, we were able to obtain in a very short period of time the acceptability judgments and native speaker intuitions from nearly 150 Spanish speakers on the Internet, and the study is one of the first to utilize Internet resources to this extent. In summary, there are now available a number of computer-based tools that make it possible for researchers to gain access to much more extensive data than previously possible, and the hope is that this data can be used to explain a wide range of syntactic phenomena in Spanish.

2. RAISED/NON-RAISED CONTRAST: PARECE QUE MARÍA ESTÁ ENFERMA / MARÍA PARECE ESTAR ENFERMA

The first phenomena of Spanish subject raising that we will consider concerns the frequency of raised (2a) vs. non-raised clauses (2b) in both spoken and written Spanish. They show quite clearly that spoken Spanish very much favors non-raised constructions (3a), whereas written Spanish favors the raised construction (3b):

- (2a) **parece que** María está enferma NON-RAISED
 (2b) María **parece estar** enferma RAISED
 (3a) A mí **me parece que** es una gran fuga la de él (Buenos Aires:M21:B:26)
 (3b) **Me parecía ser** el paciente aniquilado por la enfermedad (Venezuela 1:Alizo:46)

Table 2 shows the number and percent of both types of constructions in both the spoken and written registers:

TABLE 2
Raising / non-raising in spoken / written Spanish

	NON-RAISED parece que María está enferma	RAISED María parece estar enferma	% RAISED	RAISED without ser/estar/ haber	% RAISED
Spoken	1705	190	.10	42	.02
Habla-LatAm	1009	58	.05	27	.02
Habla - Spain	254	17	.06	0	.00
Oral - Spain	442	115	.21	15	.02
Written	601	1177	.66	1020	.57
Novels	260	362	.58	337	.54
Short Stories	190	407	.68	366	.61
Argentina	151	408	.73	307	.55

The first two columns show the total number of non-raised and raised sentences with *parecer* in the different corpora, and the third column gives the corresponding percentage of sentences with raising (*María parece estar enferma*). As one can readily see, the difference between spoken and written Spanish is quite marked. In only one of the blocks of spoken texts (a nearly 1,000,000 word corpus from Spain) is the percentage of raised clauses even near one third that of the written texts. In the *Habla Culta* texts, on the other hand, raised clauses are only about one sixth as common as in the written texts.

The last two columns refer to a certain subset of raised clauses, those in which the embedded verb is not *ser*, *estar*, or *haber*. Some linguists, such as García Pinto and Luján (1974), have claimed that raising is only possible when the embedded verb is stative, such as with the three verbs just mentioned. Ruíz Morales (1979) and Seaver (1982:173) relax this restriction somewhat to include a few other verbs,

but still mention that raising is best with stative verbs. Subirats-Ruggeberg (1987:123), on the other hand, argues that raising can occur with any embedded verb. Let us consider what the data from both the written and spoken corpora indicate about the effect of the semantics of the embedded verb. Looking first at written Spanish, Table 2 shows that in this register the semantics of the embedded verb do not appear to be important. More than 86% of the cases of raising in written Spanish (1020/1177 cases) are with verbs other than *ser*, *estar*, or *haber*. In spoken Spanish, on the other hand, the semantics of the embedded verb are quite important. As Table 2 shows, nearly two thirds of the cases of raising in the *Habla Culta* corpus are with *ser*, *estar*, or *haber* (48/75), and in the *Corpus Oral* from Spain 100/115 cases occur with one of these three verbs.

Nevertheless Subirats-Ruggeberg is right in arguing that there are still a number of verbs with which raising can occur even in spoken Spanish. In the spoken corpus we find raising with embedded verbs like *aislar*, *aliviar*, *brindar*, *buscar*, *funcionar*, *irritar*, *participar*, *rendirse*, *ubicar*, and *ver*:

(4a) Ahora, cuando estaba en el banquillo, mientras tu comentabas esto, **parecía buscar** algún tipo de golpe en la parte derecha de su cuerpo (Deporte EDEP007A:49)

(4b) En esos momentos, la vida **parecía brindar** al poeta todas sus oportunidades (Havana:M37:677)

Yet these few cases of raising with verbs besides *ser*, *estar*, and *haber* should not obscure the fact that raising is quite uncommon in spoken Spanish, in which only 10% of cases with *parecer* take raising, and in which nearly 80% of these cases are with one of the three verbs *ser*, *estar*, or *haber*.

Now that we have provided data showing that there is a clear difference in raising between spoken and written Spanish, let us suggest some possible motivations for this difference. One possible motivation has to do with the discourse status of the raised noun phrase. Givón (1990:771-75) argues that raising is a function of the topicality of the noun phrase; the more topical the noun phrase,

the more likely it is that it will undergo raising. Suppose we have the following two sentences:

- (5a) — parece [que Juan estar distraído]
low topicality
(5b) Juan parece [— estar distraído]
high topicality

In a context where we have been talking about Juan and certain strange behavior that he exhibited at a recent party, we might add a sentence with raising '*y parecía estar bastante distraído también*', since he is the topic of the sentence. If, on the other hand, we have been talking about María and certain things that have her worried and then shift to talking about Juan, it would then be more natural to not have raising, since Juan is low in topicality: '*y bueno, como parecía que Juan estaba distraído también ...*'.

In addition to topicality, there is also a semantic factor concerning *parecer* that might be related to raising. As Ramsey (1956:394) and Fernández Leborans and Díaz Bautista (1990:353-54) mention, there seem to be two different meanings with *parecer*, one in which it is 'impersonal' and the other in which it is 'attributive'. According to these researchers, the 'attributive' *parecer* refers to appearance and can be paraphrased in English by cases like 'Mary looks like she's sick (see how flushed she is)'. The 'impersonal' *parecer*, on the other hand, refers to belief and opinion: 'it looks like Mary is sick again (judging by what Tom said)'. According to these researchers, the attributive reading lends itself to raising, whereas the impersonal one does not.

This is a hypothesis that had not previously been tested empirically. We attempted to do so and presented the 82 speakers in the follow-up survey with the following two contexts. The first example is attributive because it refers to physical perception, while the second is impersonal since it involves indirect evidence or hearsay:

- (6a) Hace varios días que no nos vemos, pero la última vez que la vi, _____.
(6b) Ya sabíamos que Pedro y Juan estaban enfermos, y como _____ también (según las noticias que nos trajo Alejandro), decidimos cancelar el programa.

Speakers were asked to fill in the blanks with either the raised *María parecía estar enferma* or the non-raised *parecía que María estaba enferma*. The clear majority of the respondents indicated that the raised construction was preferred in the first sentence (6a), which involves direct perception of María's sickness (*la última vez que la vi*). The second sentence, on the other hand, is impersonal since the news of her sickness comes second-hand (*según las noticias que nos trajo Alejandro*), and speakers generally concurred in their preference for the non-raised construction.

As we have seen, certain pragmatic and semantic factors are related to raising. Yet neither of these two factors alone could be sufficient to explain the spoken/written contrast, unless we assumed an unlikely scenario in which noun phrases are more topical or more 'attributive' in written than in spoken Spanish, which would produce a higher degree of raising in the written register. There must be some additional factor(s) that can explain the difference between the two registers. Our suggestion is that the missing piece to this puzzle has to do with general factors relating to the production of spoken and written language. These factors have been the focus of corpus-based research in a number of other languages. Miller (1994:4301), for example, reminds us that one obvious difference between the two registers is that spoken language is produced in real time, whereas written language has more careful editing, which permits the writer to develop more complex syntax. Chafe (1982:37) adds that 'in writing, we have time to mold a succession of ideas into a more complex, coherent, integrated whole, making use of devices we seldom use in speaking'. In more specific terms, Biber (1988:47) notes that infinitives are used to achieve a higher degree of clausal integration (see also Beaman 1984 and Chafe 1985). Thus we see a cross-linguistic preference of written texts for a higher degree of clausal integration, such as with

the infinitival clauses found in the raising sentence *María parece estar enferma*.

Shorter, more defined clauses are characteristic of spoken language. These clauses are generally introduced with conjunctions that separate the subordinate from the main clause. In studies carried out on extensive corpora of spoken English, Biber (1988:159, 195) notes that 'the primary use of *that* complements . . . seems to be for informational elaboration under real time production constraints' and 'in discourse that cannot be carefully planned and integrated'. This fits in with a more general 'one-clause-at-a-time' constraint that has been hypothesized for spoken language (Sacks, Schegloff, and Jefferson 1974, Pawley and Snyder 1983), and corresponds in turn to the non-raising (*parece que María está enferma* type sentences. Some comments from Spanish grammars also make oblique reference to the tendency of spoken Spanish to (over)employ short clauses introduced by *que* (Gili Gaya 1961:288-89, Ramsey 1956:581-82), while noting that overly artificial written Spanish often omits *que*: *creía haber dicho la verdad*, or *les rogamos nos comuniquen sus deseos* (cf. Butt and Benjamin 1988:336-7).

Some of the comments made by the native speakers from the survey regarding the use of the non-raised *parece que* sentences of spoken Spanish can also be related to these more general studies. Several speakers from the survey noted that when they were in school, their teachers encouraged them to avoid overuse of *que* as a *muletilla* -- a syntactic 'crutch'. One speaker mentioned that *que* was part of 'el lenguaje informal, por falta de tiempo o de recursos lingüísticos', while another suggested that it was a 'muletilla común para crear una pausa que facilite la respiración'. While these explanations might seem overly informal or anecdotal, they do find support in more formal studies. Thus there is strong cross-linguistic evidence to support the claim that tightly integrated clauses (with infinitives) are characteristic of careful written language, while looser, 'one clause at a time' sentences joined together with conjunctions are more characteristic of spoken language.

One remaining problem, however, is integrating this explanation with the fact that the raised / non-raised distinction is also in part a function of the topicality of the noun phrase and whether or not the sense of *parecer* is attributive or impersonal. We have suggested that no one factor alone can be responsible for the spoken / written distinction because the factor would not be significantly different in either spoken or written Spanish. As Kemmer (p.c.) has suggested, however, these factors might yet be integrated into general constraints on spoken and written language. As she comments, 'written registers tend to do more in the way of making the form-/meaning relations less transparent than spoken registers do. Highly planned discourse gives language users much greater opportunity to do more complex things, more marked things, for particular purposes'. In other words, raising is a function of factors such as topicality and information structure, but it is in carefully planned written language that speakers are best able to make these distinctions. Hence, we would argue that subject raising in Spanish is both a function of general cross-linguistic factors such as topicality and constraints on spoken and written language, as well as perhaps more language specific factors such as the oft-commented distinction between the attributive and impersonal meanings of *parecer*.

3. NON-3SG USES OF PARECER: PARECIAS ESTAR TRISTE

In the preceding section we saw that subject raising in Spanish is more restricted than in English, and is especially limited in spoken Spanish. In this section we will consider a subset of the raising sentences that is even more limited than those described above, but which provide interesting insight into how semantic and pragmatic factors influence raising. As Bolinger observes (1972:73-74), the less prototypical cases of raising in Spanish involve non-third person singular subjects (*tú pareces, ellos parecen*), which differ in their distribution from the more common third person singular subjects. As shown above, subject

raising in English can occur freely with first, second, and third person subjects:

- (7) John / they / you / I seem(s) to have offended Mary

In Spanish, on the other hand, it is most acceptable with third person singular, somewhat less with third person plural, even less with second person, and is quite awkward with first person subjects:

- (8a) Juan parece haber ofendido a Marta
 (8b) ellos parecen haber ofendido a Marta
 (8c) ? tú pareces haber ofendido a Marta
 (8c) ?? yo parezco haber ofendido a Marta

Bolinger suggests the restrictions on Spanish may be due to the fact that historically, true raising constructions (9a) developed from the 'partial raising' constructions (9b):

- (9a) Juan parece haber ofendido a Marta
 (9b) Juan, parece que ha ofendido a Marta

In cases of non-third person subjects, however, there is no agreement between the fronted noun phrase and the following third person singular *parece*, and for this reason there are fewer cases of raising with these subjects.

- (10a) ellos, *parece* que han ofendido a Marta
 (10b) tú, *parece* que has ofendido a Marta

Both Bolinger (1972:74) and Fernández Leborans and Díaz Bautista (1990:367) suggest that there may be a semantic explanation as well. The typical reading with *parecer* is that something seems a certain way to *me*, in which I am the observer. It is awkward, then, to have a first person subject, which would mean 'it seems to me that I have done a certain thing', in which I am both the doer and the observer:

- (11) ?? yo parezco haber ofendido a Marta

What is lacking to this point, however, is concrete data on the extent of raising with non-3SG subjects in Spanish. Bolinger points out that his data seen in (8a-d) is based on informal, anecdotal evidence gained from speakers of Spanish in his classes, and there are no other studies that have looked at the issue of non-3SG *parecer*. Yet the phenomenon is of interest for a number of reasons. First, we might want to see whether the extension beyond the 3SG uses of *parecer* has taken place in any register of Spanish, as it has with *seem* in English, and if so where, and to what extent. Second, it might be interesting to consider how the non-3SG cases of raising relate to factors such as topicality and to the attributive / impersonal nature of *parecer*, which are two factors that affect raising. By studying the extent and nature of non-3SG raising, which represent the most innovative cases of raising in Modern Spanish, we may well gain insight into the future evolution of subject raising in Spanish.

Let us first consider the data from our corpus of Modern Spanish. Table 3 shows the number of non-3SG cases in the different registers of the corpus. The first column shows the number of cases with 3PL (*ellos parecen haber ofendido a Marta*) and the last column shows all remaining cases with first and second person subjects (*yo, nosotros, tú, vos, vosotros*).

TABLE 3
Non-3SG raising (*parecías estar triste*)

	3PL <i>parecen saber la respuesta</i>	1-2 SG-PL <i>parezco/pareces/ parecemos saber la respuesta</i>
Spoken	12	0
Habla - LatAm	11	0
Habla - Spain	0	0
Oral - Spain	1	0
Written	262	6
Novels	77	3
Short Stories	88	2
Argentina	97	1

These figures clearly demonstrate that raising with non-3SG subjects is much more common in written than in spoken Spanish, which is not surprising considering the much higher incidence of raising in written Spanish generally. There are only twelve examples of non-3SG subjects in the entire spoken corpus (12a-b), while there are 262 in the written corpus (12c-d):

- (12a) *y parecían aislarlo de las fuerzas francesas, ¿no?* (La Paz: M5:63)
 (12b) *A mí me decepcionó que parecían entender las nuevas ideas* (Buenos Aires: M11:A:175)
 (12c) *Siempre, a esta hora, parecían desayunar los niños* (Muerte 150:3)
 (12d) *Ahora, al avance hacia el Nuevo Mundo, parecen irse borrando poco a poco Colombia 1: Gómez:335)*

If we look at non-3SG subjects as a percentage of all cases of raising, we see that it is quite a marked construction in spoken Spanish. Only about six percent of all cases of raising in spoken Spanish are with non-3SG subjects (12 non-3SG / 190 3SG). In written Spanish, on the other hand, nearly 20% of all cases involve non-3SG subjects (262 non-3SG / 1177 3SG).

As might be expected, the most uncommon cases of raising are with first and second person subjects. As Table 3 shows, there are no cases at all of these constructions in the spoken corpus, and only six total cases in the written corpus, three with *tú/vos*, two with *nosotros*, and just one with *yo*:

- (13a) *Tú parecías andar con paso firme, más yo no tenía bien claro lo que quería hacer* (Colombia 2: Bahamón: 18)
 (13b) *Unicamente ellos y yo parecíamos haber quedado en la casa.* (México 2: Tarió: 128)
 (13c) *Parecíamos temer a los recursos pero disfrutábamos de los tantos locos sueltos que pululaban en esos días* (Argentina: Novios: Chapter3)
 (13d) *A vos parece darte lo mismo que ... No parecés darte cuenta de que esas irrupciones ...* Rayuela 299:9)
 (13e) *mi hermanita enferma, valerosa, terriblemente valerosa, con su vientre de soltera maltratado raspado hasta el vivo; toda la sangre que salió por a herida ... Pareces preguntar si aún hay mundo afuera* (Barro 195:2)

(13f) *No sé qué posición tomar aquí, debajo de la línea del Ecuador; encogiendo las piernas, pareczo defecar en el piso; estimarlas, a más de resultar impropio de toda etiqueta, me haría tropezar con las macetas grabadas a cuchillo en la caja del escritorio* (Barro 90:2)

Since these cases of raising with first and second person represent the most innovative type of raising in Spanish, we might consider for a moment what contributing factors are present in these cases. The four cases in (13a-d) all involve contrast, which is suggestive of a topicalized element ('*tú parecías andar ... pero ...*', '*a vos parece darte ... no yo*', '*parecíamos ... pero ...*', '*a vos parece darte ... no parecés darte*'), while (13e-f) are cases of the 'attributive' use of *parecer*, in which it is used to provide physical description. (13f) is a particularly interesting case because although raising with first person subjects is uncommon, owing to the semantic incongruity of a person being both the doer and the observer, in this case it makes sense because the narrator is concerned with how he must have looked to others. In summary, all of the handful of cases with non-3 raising in the corpus can be motivated by one of the semantic-pragmatic factors discussed previously.

Turning to the results from the survey, we find a nice match between the grammaticality judgments of the native speakers and the data from the corpus. In the survey, the 82 speakers in the follow-up survey were given three different sentences, with 3SG (14a), 3PL (14b), and 1PL (14c) subjects:

- (14a) *María parece conocerlas*
 (14b) *ellos parecían haber ofendido a Marta*
 (14c) *parecíamos haberlo visto antes*

As expected, the speakers who responded gave the highest score to the 3SG example (4.39), a slightly lower score to the 3PL sentence (3.91), and a significantly lower score to the 1PL example (2.45). This fits in well with Bolinger's hunch that there was a slight difference between 3SG and 3PL subjects, but a larger acceptability gap between those and 1-2 person subjects. In addition to simply providing scores for the different sentences, several speakers also

added notes indicating that although the non-3(SG) sentences are 'correct', they were in fact 'not said', and that they sounded 'too complicated and proper', which fits in with the argument made above that these highly marginal cases of subject raising in Spanish are limited primarily to written, formal registers.

Native speakers provide valuable insight into what type of context might make the marginal non-3(SG) sentences acceptable. Several speakers in the surveys noted that in a context where the emphasis was on the appearance or attitude of the subject, or especially where the subject was only pretending to have carried out the action, then the sentences became more acceptable ('*se ve la actitud culpable de los ofensores*', '*se mostraban como si hubieran ofendido a Marta*', '*tienen la apariencia de haberla ofendido*'). In fact several speakers suggested that in order to get this reading, they would have preferred to use the verb *aparentar* 'to look, feign, give the impression', which expresses the idea more clearly. Thus there appears to be a semantic-pragmatic continuum in which the most marginal uses of *parecer* merge into other lexical items, which again suggests that raising is in part a function of semantic-pragmatic factors and that it is not a strictly syntactic phenomenon.

4. 'PARTIAL' RAISING: ELLOS PARECEN QUE ESTÁN ENFERMOS

Another less prototypical case of subject raising in Spanish is the construction in which the subject raises up to the main clause subject position and produces verbal agreement but is not 'deleted' in the embedded clause:

- (15) parecen que están enfermos

This construction has received scant attention from previous researchers. Subirats-Ruggeberg (1987:130) suggests that these sentences are cases of topicalization, which is also a motivation for true subject raising, but says no more about the construction. Bolinger (1972:73-74), basing his comments on a few informal grammaticality judgments

by native speakers, suggests that the construction is most acceptable with third person subjects (*parecen que han perdido el dinero*) and is much less acceptable with first or second person subjects (*yo parezco que estoy enfermo*).

The data from the corpus show this to be quite an interesting construction, both because of its marginal status and also because of its distribution. There are only nine examples of this construction, but eight of them come from spoken Spanish (16a-d), while only one is from written Spanish (16e):

(16a) este otro grupo de ciencias más especulativas, son las que me **parecen que dan** esta ... esta capacidad (Madrid M2:37)

(16b) No, Irene, eso era en la película de Indiana Jones hija, cuando dice: 'Yo, yo **parezco que he pisado** pan tostado'. Y encienden una bombilla y **son** todos los bichos. (Conversacional CC0032A:142)

(16c) Estoy hablando - a nivel burocrático me **pareces que no puedes** cambiarlo durante un año (Conversacional PC00007B:35)

(16d) ¿qué problemas te **parecen que son** los más importantes? (Sevilla-P M7:176)

(16e) Todas, casi todas ya, **parecen que quieren** levantarse de entre los cardales y esas manchas, pero se han quedado atrapadas (Hombres 244:1)

Since constructions of this type do appear in the corpus (albeit sparingly) and since the majority of them are in spoken Spanish, it was expected that a fair amount of the 82 speakers in the follow-up survey would accept such sentences, but this was not the case. In the first survey in which they were asked to judge this type of sentence, they gave an average score of just 1.71 for (17a) and 2.05 for (17b).

- (17a) ellos parecía que habían ofendido a Marta
(17b) nosotros parecía que lo habíamos visto antes

Since the data from the corpus showed these type of constructions to be more acceptable than did the results of the survey, we gave speakers a final survey in which they were presented with (16a-b) above, as well as:

(18a) parecen que reconocen a María

(18b) parecemos que hemos ganado las elecciones

Even when they were given actual examples from the corpus (16a-b), nearly all speakers rejected examples of the construction. Those who did not reject all of the sentences outright did mention that third person subjects (16a and 18a) were preferable to non-third person subjects (16b and 18b). A few speakers also mentioned that given the right context, some of the grammaticality judgments might be more positive. For example, one speaker suggested that if (16b) meant that 'mi aspecto ha cambiado como si hubiera pisado pan tostado', then the sentence would be more acceptable. A handful of other speakers agreed and mentioned that the sentences were best in a context where the emphasis was on the physical appearance or the (feigning) attitude of the subject, i.e. they are just pretending that they know María (18a), we are putting on airs as though we had one the election (18b). As we saw previously with non-third person uses of *parecer* (*pareces estar enferma*), these cases of 'partial raising', which are marginal to begin with, become more acceptable if a highly marked context can be provided.

Before discussing a possible motivation for the partial raising facts, let us consider briefly an additional set of partial raising examples from the corpus. Recall that eight of the nine examples represented in (16a-e) come from spoken Spanish, but these are just the cases of non-3SG *parecer* + *que* (*pareces que, parecen que*). It would have been quite difficult to sort through the many cases of 3SG '*parece que*' and eliminate the 'impersonal' cases (*parece que está enfermo*) to extract just those cases of NP_i *parece que NP_i V* (*Juan_i parece que está, enfermo*). Yet it is possible to find selected case of 3SG, in which *parecer* is preceded by *él* or *ella*, or *usted*. There are an additional nine cases of these, all of which are found in spoken Spanish, as in the following:

(19a) *Sí, ella parece que quiere o por lo menos está haciendo el esfuerzo por ver lo que ellos están haciendo, ¿no?* (Havana M7:206)

(19b) *Él parece que le dijo, después de casado ya, le dijo que no le arreglaba la ropa.* (Buenos Aires M27:B:331)

(19c) como *usted parece que conoce a las dos partes* (Jurídico AJUR016A:170)

With these additional nine examples, we have provided a total of eighteen cases of partial raising from the corpus, seventeen of them from spoken Spanish. We now want to consider why it is that the cases of partial raising are found almost exclusively in the spoken register, but are then rejected by native speakers as ungrammatical. The answer may simply be that as speakers produce these sentences, they start into raised construction (*ella parece [querer] ...* in (19a), *me pareces [no poder] ...* (16c), and then 'backtrack' and revert to a non-raised construction (*ella parece que quiere, me pareces que no puedes*). In some cases this might be due to what would be awkward strings of infinitives (as in (16c), (16e), and (19a), but there are other cases like (16a), (16b), (16d), and (19c), where a normal string would result (*parecen ser, parece conocer*). The explanation for all cases, however, is probably found in the research presented above that shows that cross-linguistically, spoken language tends to avoid infinitival subordination in favor of shorter *que/that* clauses, which would explain why speakers start into an infinitival construction but then switch to a finite *que* clause. It also explains quite nicely why there are almost no examples of this construction in written Spanish, where the writer is able and more likely to take the time to produce a complex infinitival construction. Finally, this scenario also explains why the speakers in the survey were so reluctant to accept any of these sentences. Since the strings result primarily from speakers attempting to 'escape' from an awkward construction, the speakers in the survey do not recognize them as part of the 'normal' grammar they use in production. If this scenario is correct, we can see that the 'partial' raising constructions provide us with valuable insight into the nature and functioning of the internal grammar of native Spanish speakers, and the way in which they produce this particular subset of sentences.

5. THE INDIRECT OBJECT CONSTRUCTION: ME PARECÍA OÍR SU VOZ

Another subject raising construction that has received very little attention in previous studies is one in which there is an indirect object that refers to both the both the subject of the embedded verb as well as the 'observer' or 'experiencer' of *parecer*:

(20) me parecía oír su voz

Bolinger (1972:75) suggests that this construction is analogous to the English personal construction (*I seemed to hear her voice*) and is an alternative to the marginal *yo parecía oír su voz* discussed above, and Ramsey (1956: 466) suggests that the indirect object construction is somewhat less colloquial than the equivalent finite clause *me parecía que (yo) oía su voz*. Yet there are a number of questions concerning the construction that have not been addressed in previous research. For example, we might ask whether the construction is limited primarily to written Spanish, as are the other raising constructions, or whether it is commonly found in spoken Spanish also. Another issue might be the restrictions placed on this construction in terms of the semantics of the embedded verb.

Concerning the first question, it appears that as with the other raising constructions, the indirect object construction is primarily found in formal, written Spanish. The corpus shows that while there are 54 cases in written Spanish (21a-b), there are only three cases in spoken Spanish (22a-b).

- (21a) **Le pareció caer** hacia abajo y que las rocas de lava se despeñaban junto con ella (Cuba 2: Montaner:131)
 (21b) A eso de la media noche **me pareció escuchar** un mugir dolorido (Puerto Rico 2: Díaz:57)
 (22a) la recuerdo mucho. **Me parece verla**, a ella, delgada (Lima: M20:268)
 (22b) **me parece ver** la viejita con su vestido rojo (Bogotá: M44:617)

The 82 speakers in the follow-up survey concur with the data from the corpus and with Ramsey's assertion that the

construction is somewhat less common than the non-raised *parece que* construction. For example, while speakers gave the non-raised construction '*parecía que habían ofendido a Marta*' an average score of 4.3, they gave the related indirect object construction '*les parecía haber ofendido a Marta*' a score of only 3.7.

Perhaps a more interesting issue than the frequency of the construction in different registers is the relationship between the indirect object construction and the semantics of the embedded verb, which is something that has not been considered in previous research. The corpus shows quite clearly that nearly all of the embedded verbs in the indirect object construction are what are known as 'experiencer' or 'non-agentive' verbs, in which the subject is not a willful agent who carries out an action on another person or thing, but rather the 'experiencer' or 'undergoer' of a certain phenomena or condition. For example, of the 57 cases in the corpus, more than half are with verbs of perception (*ver* 17 cases, *escuchar* 6, *oír* 3, and *percibir* 3):

(23a) **le parecerá oír** la risa de Boris y su voz quejumbrosa (México 1:Pit01:159)

(23b) **Les pareció ver**, a lo lejos, una blanca sombra que aparecía y desaparecía entre los troncos (Colombia 1: Zalamea:242)

The others are with verbs that express cognition or belief (*apreciar*, *reconocer*, *conocer*, *distinguir*, *discernir*, *entrevener*, *descubrir*, *comprobar*, *admitir*, *advertir*, *notar*) or verbs in which the subject is very low on the agentivity scale (*caer*, *vivir*, *girar*, *obtener*, and *llegar*):

(24a) **Le pareció reconocer** cada matorral, cada repliegue del terreno ... (Colombia 1: Zalamea:240)

(24b) **le parecía girar** suavemente en el borde de un gran círculo (Chile 2: Guzmán:183)

To test the hypothesis that the indirect object construction was in fact limited (or most acceptable) with non-agentive verbs, the 82 native speakers were asked to compare the following three sentences:

- (25a) nos parecía haberlo visto antes
 (25b) les parecía haber ofendido a Marta
 (25c) les parecía haberla matado por su dinero

These three sentences range from the 'experiencer' type with *ver* in (25a) to *ofender* in (25b), which is potentially but not obligatorily agentive (*queríamos ofenderla para que no nos molestara más vs. la ofendimos sin querer*), to *matar* (25c), which is highly agentive. The speakers indicated a high degree of acceptance with *ver* (4.6), a lower degree with *ofender* (3.7), and the lowest degree of acceptance with *matar* (2.5).

The fact that nearly all verbs are non-agentive and of the experiencer verb class is quite interesting from a cross-linguistic point of view. Across the languages of the world, there is a hierarchy of 'subjecthood' (cf. Comrie 1989), in which noun phrases with certain verbs are seen to be strongly agentive because they willfully carry out an action that affects a patient (*Juan ha pintado el coche, Susana ha escrito un libro*). With these verbs nearly all languages code the noun phrase as subject. Yet with certain verbs, especially verbs of perception, cognition, and emotion, the noun phrase is seen as being more of an 'experiencer', because it experiences an action or state, but does not willfully cause an action that affects a patient (*Juan vio a María, Pedro entiende la lección*). Since these noun phrases are lower on the subjecthood scale, some languages code them as subject (Modern English *I like the book*), whereas in other languages their non-agentive status is marked by coding them as indirect objects, the next grammatical role on the hierarchy of subjecthood (Spanish *me gusta el libro*, Old English *me likes the book*). Hence it is not surprising that there is a link between subjects that are experiencers or weak agents and the indirect object construction, since indirect objects are often used to encode the subjects of such verbs.

There is also another explanation, one that is perhaps more intuitive and does not make reference to concepts such as 'experiencer verbs' and 'subject hierarchies'. The fact is that there is incongruity between stating that it 'appears' to someone that an action has been carried out,

which indicates uncertainty or doubt, when they themselves are the willful agent of that action, which means they are fully aware that the action has been carried out (*les parece haberla matado por su dinero*). Some speakers mentioned that a sentence such as this is acceptable only in a context in which there is still some uncertainty about her murder, such as whether she really died, or if they couldn't remember whether they had killed her for her money or for some other reason. In the case of *les parece haber visto a María*, on the other hand, the act of 'catching sight of someone' does not involve willful intent, and so the uncertainty involved with *parecer* makes sense. In summary, then, we see an interesting relationship between the use of the indirect object construction and the semantics of the embedded verb, which ties into more general cross-linguistic notions such as experiencers and subjecthood hierarchies.

6. OTHER CLITICS WITH RAISING: MARÍA ME PARECE ESTAR ENFERMA/ MARÍA ME PARECE CONOCER

Let us consider finally two additional aspects of the Spanish subject raising construction, both of which deal with the use of object clitics. In the previous section we examined the indirect object construction, in which there is coreference between the observer and the subject of the embedded verb (*a María le parece oír voz*). But what about cases where there is no coreference with the subject of the embedded infinitive:

- (26a) me parece existir una duda acerca de ese relato
 (26b) María me parece estar enferma

Ramsey (1956:466) is the most lenient and simply suggests that these type of sentences are best when there is coreference. Bolinger (1972:75) is less accepting of such sentences, and argues that examples like (26a) are doubtful at best. Fernández Leborans and Díaz Bautista (1990:419-20), in a passing reference, claim that there is obligatory coreference with the embedded subject, since there is a 'control' struc-

ture in which the indirect object controls the PRO subject of the embedded verb. No previous study, however, has provided hard evidence for or against obligatory coreference.

The data from the corpus show that there are occasional cases of non-coreference. In total, there are four cases in which the embedded verb is *ser*, three of them in the written corpus and just one in the spoken corpus:

- (27) distinguió un pequeño barco que le parecía ser de los que acostumbradamente navegan por esas aguas (Perú: Alegría:144)
 (28) en tercer lugar quisiera indicarle lo que me parece ser la ... problemática más decisiva (San Juan: M12:249)

Aside from these cases with *ser*, there are three other cases of non-coreference, all of them in the written corpus:

- (29a) Nunca lugar alguno del mapa le parecía tener sentido permanente (Chile 1: Echeverría:87)
 (29b) en el aspecto espacial no le parecía dejar nada por desear (Tiempo 84:7)

Thus the data show that while coreference is not absolutely obligatory, the exceptions are rare and are limited primarily to the simple verb *ser* and to written Spanish.

To follow up on these facts, the 150+ native speakers were asked to rate the following two sentences, in which there is no coreference:

- (30a) me parece ser una solución fácil
 (30b) María me parece estar enferma

Speakers were somewhat accepting of impersonal cases with *ser* as in (30a), where there is no such subject, and assigned it an average score of 2.5. However they nearly uniformly rejected cases like (30b) in which there is a specified subject of the embedded verb, and gave it an average score of just 1.5. Thus it does appear that there are some cases of non-coreference between the indirect object and the embedded clause subject, although they are quite infrequent and do seem quite awkward to native speakers.

The other question regarding clitics is whether in subject raising constructions an embedded clause object can move into the main clause (31a) as it can with object raising (31b) or with EQUI/control constructions (31c):

- (31a) María me parece conocer < parece [María conocer me]
 (31b) ellos me lo hicieron limpiar < ellos hacer [yo limpiar lo]
 (31c) María lo quería comprar < María querer [María comprar lo]

A number of researchers have argued that sentences such as (31a) are at least partially acceptable (Subirats-Ruggeberg 1987:129-30, Butt and Benjamin 1988: 250, Contreras 1979:174), while others clearly rule it out (Fernández Leboráns and Díaz Bautista 1990:372).

In our search of the computer corpus, we found no examples of cases like (31a), in which an embedded clause object raises along with the subject to the main clause. In terms of data from the survey, the 146 native speakers in the initial survey were given the following two sentences:

- (32a) María las parece conocer
 (32b) María se parece despertar

The speakers gave a low average score of 1.75 to (32a) and an even worse 1.29 to (32b). In fact (32b) was so unacceptable that several speakers facetiously asked if they could give it a score of 0 or -1 on a scale of 1-5. Although part of the problem with (32b) may be due to the use of the present tense (as opposed to the progressive: *María (se) parece estar despertando(se)*), it may also be due to the fact that non-reflexive clitics in general climb more easily out of the embedded clause than do reflexive clitics (see Davies 1995:376). At any rate, both the data from the corpus as well as the judgments of native speakers show that clitic climbing is at best very marginal with subject raising in Spanish.

Finally, we might ask why it is that clitic climbing is unacceptable with subject raising when it is acceptable with other constructions like object raising (31b) and EQUI/control structures (31c). The solution may have to do with the indirect object construction discussed above.

There are examples with the indirect object construction where it is at first sight unclear whether the clitic refers to the observer of the action or whether it is the object of the embedded verb, or at least cases like (33c-d) in which temporary 'garden-path' confusion could result. Thus the strategy may be to simply disallow embedded object clitics in the main clause, to avoid this type of ambiguity.

- (33a) *le parecía advertir* hasta qué extremo estaban unidos y separados (México 1: García:138)
 (33b) y en las manos y en la nariz *me parecía comprobar* que además de pipí se estaba haciendo algo más feo (Julia 274:1)
 (33c) *me pareció ver* por un instante, sólo por un instante, sombras de cuerpos humanos (Cuba 1: Leante:397)
 (33d) y *le parecía escuchar*, en el silencio de la noche, el lento latir del mangó (Puerto Rico 3: Ferré: 113)

7. CONCLUSION

One goal of this study was to demonstrate the value of a strong data-based approach to syntax. The more than 3600 tokens from the 7,700,000 word corpus of both spoken and written Spanish, as well as the acceptability judgments and intuitions from the nearly 150 native Spanish speakers on the Internet, were invaluable in providing evidence for the different phenomena that we have tried to explain. In some cases our data simply provided more concrete evidence for claims made by others, such as the idea that raising was more common with 3SG subjects and that there was in fact a 'partial' raising construction. In other cases it resolved certain disagreements, such as whether main clause clitics were acceptable in cases other than the indirect object construction, and whether the standard raising construction occurred with a full range of verbs or just the statives *ser*, *estar*, and *haber*. In many other cases the data provided evidence of certain phenomena for the first time. Examples of this are the idea that the basic raising construction as well as the indirect object construction were much more common in written than in spoken Spanish, that the partial raising construction was mainly limited to spoken Spanish, and that the indirect object construction

occurred almost exclusively with a certain class of embedded verbs. Our hope is that studies such as this demonstrate the methodological value of using large computer-based sources of data.

The second goal of this study was to show how a number of pragmatic, semantic, psycholinguistic, and other functional factors can explain quite nicely a number of phenomena related to Spanish subject raising. These are the type of explanations that have typically been overlooked in previous studies on the topic. We have shown that pragmatic and semantic factors such as topicality and the 'attributive / impersonal' distinction are relevant, especially in the case of the less prototypical cases like non-3SG subjects (*parecíamos estar enfermos*) and partial raising (*ellos parecen que la conocen*). The more the construction deviates from the prototypical 3SG full raising construction, the more important the pragmatic and semantic factors are in creating a felicitous context. These pragmatic and semantic factors are also at play with the indirect object construction, which is used mainly with embedded clause verbs with non-agentive subjects (*me parecía oír su voz / ??me parecía cortar la leña*). Functional factors such as the avoidance of ambiguity with the indirect object construction may be the motivation for avoiding other main clause clitics (*me parecía ver unas sombras / ??Juan me parecía ver unas sombras / ??Juan me parecía ver*). Finally, psycholinguistic factors dealing with the production of spoken and written language provide the most likely explanation for the overall difference between the predominance of raising in written Spanish as opposed to spoken Spanish. This factor is probably also related to speaker's rejection of partial raising (*ellos parecen que la conocen*), even though these types of constructions are not uncommon in the spoken corpus. In summary, we hope that the strong data-based approach and the reliance on a wide range of linguistic motivations will have provided a much greater degree of insight into several different aspects of subject raising in Spanish.

NOTES

¹ In this paper we often use the term 'subject raising' in place of the more cumbersome 'subject to subject raising'. Yet one should keep in mind that in classical transformational grammar, there were assumed to be other types of raising to and from subject position, such as subject to object raising (*lo vi trabajar < vi [él trabajar]*) and object to subject raising (*el libro es difícil de leer < es difícil [leer el libro]*). For theory-internal reasons, these two types of raising are not permitted in the Government and Binding model. In the GB model there is also no longer a specific rule of subject to subject raising per se, but rather it is subsumed within generalized 'noun phrase movement' or the even more general rule of 'Move α' '. In this paper, however, we will continue to use the traditional terminology of 'subject raising'.

² Composition of the corpus. *Note:* While the *Corpus Oral* (2) and *Argentina* (5) materials are in the public domain, the electronic versions of other materials are not (*Habla Culta*, *Novels*, *Short Stories*).

1. **Habla (LatAm and Spain).** Materials from the Habla Culta project. Listed by city, with title, date of publication in book form, and number of words in computer corpus. BOGOTÁ: (240,000) *El habla de la ciudad de Bogotá. Materiales para su estudio*, 1986; BUENOS AIRES: (253,000) *El habla culta de la ciudad de Buenos Aires. Materiales para su estudio*, 1987; CARACAS: (280,000) *El habla culta de Caracas. Materiales para su estudio*, 1979; HABANA: (140,000) From thirty interviews taken from an unpublished manuscript in possession of Prof. Joseph Matluck, Univ. Texas-Austin; LA PAZ: (178,000) *El habla de la ciudad de La Paz. Materiales para su estudio*, 1992; LIMA: (119,000) *El español de Lima. Materiales para el estudio del habla culta*, 1989; MADRID: (143,000) *El habla de la ciudad de Madrid. Materiales para su estudio*, 1981; MÉXICO CULTO: (178,000) *El habla de la ciudad de México. Materiales para su estudio*, 1971; MÉXICO POPULAR: (174,000) *El habla popular de la ciudad de México*, 1976; SAN JOSÉ (COSTA RICA): (99,000) *El habla culta de San José, Costa Rica. Materiales para su estudio*, Ms. inédito, c1994; SAN JUAN: (162,000) *El habla culta de San Juan. Materiales para su estudio*, 1990; SANTIAGO: (396,000) *El habla culta de Santiago de Chile. Materiales para su estudio*, 1979-1990; SEVILLA CULTO: (75,000) *Sociolingüística andaluza 2. Material de encuestas para el estudio del habla urbana culta de Sevilla*, 1983; SEVILLA POPULAR: (77,000) *Sociolingüística andaluza 4. Encuestas del habla urbana de Sevilla - nivel popular*, 1987. I would like to thank George DeMello for his considerable assistance in acquiring these materials. [Citations refer to Ciudad:Muestra:Página]

2. **Oral - Spain.** Corpus oral de referencia de la lengua española contemporánea. (From the Facultad de Lingüística General en la Universidad Autónoma de Madrid. Francisco Marcos-Marín, Director. Downloaded from the Internet at <ftp://lola.llf.uam.es/pub/corpus/oral/>). The file *corpus.lee* at the FTP site gives a complete listing of all of the transcrip-

tions used in the corpus. Contains the following types of conversations (with the number of words in each genre): Administrativo (6,000), Científico (31,000), Conversacional (189,000), Debate (73,000), Deporte (43,000), Documentario (24,000), Educativo (54,000), Entrevista (132,000), Humanístico (48,000), Instruccional (5,000), Jurídico (33,000), Lúdico (46,000), Noticiario (60,000), Político (50,000), Publicitario (23,000), Religioso (99,000), Técnico (33,000). [Citations refer to Genre:Block Number (follows computer corpus)]

3. **Novels.** Fifteen novels from Latin America and Spain from 1950-1975. Obtained from the Humanities Research Center at Brigham Young University in Provo, Utah. Full publication information not available. Titles, authors, and number of words: Los pies de BARRO, Salvador Garmendia (99,000); EL CAMINO, Miguel Delibes (50,000); CIEN años de soledad, Gabriela García Márquez (135,000); Casa de CAMPO, José Donoso (128,000); De DÓNDE son los cantantes, Severo Sarduy (35,000); GAZAPO, Gustavo Sainz (43,000); Los HOMBRES de a caballo, David Viñas (138,000); La tía JULIA y el escribidor, Mario Vargas Llosa (128,000); La guaracha del MACHO camacho, Luis Rafael Sánchez (42,000); MALADRÓN: epopeya de los Andes Verdes, Miguel Asturias (67,000); LA MUERTE de Artemio Cruz, Carlos Fuentes (97,000); RAYUELA, Julio Cortázar (172,000); LA REINDIVICACIÓN del conde Don Julián, Juan Goytisolo (55,000); SOBRE héroes y tumbas, Ernesto Sábato (49,000); TIEMPO de silencio, Martín Santos (89,000). [Citations refer to Title:Page Number]

4. **Short stories.** For each country there are between 100,000 and 110,000 words of text from the following collections of short stories. ARGENTINA: *Antología de cuentistas argentinos*, 1979; BOLIVIA 1: *Los mejores cuentos bolivianos del siglo XX*, 1989; 2: *Antología de cuentos: II Concurso nacional*, 1968; CHILE 1: *Antología del nuevo cuento chileno*, 1954; 2: *Cuentos de cabecera*, 1967; COLOMBIA 1: *El cuento colombiano*, 1980; 2: *Premios que cuentan*, 1988; CUBA 1: *Cuentos cubanos del siglo XX*, 1975; 2: *Narradores cubanos de hoy*, 1975; ESPAÑA 1: *22 narradores españoles de hoy*, 1990; 2: *Cuento español de posguerra*, 1986; MÉXICO 1: *Los mejores cuentos mexicanos*, 1982; 2: *Antología de cuentos mexicanos*, 1977; PERÚ: *El cuento peruano*, 1968-1974, 1984; PUERTO RICO 1: *Reunión de espejos*, 1983; 2: *Cuentos puertorriqueños de hoy*, 1968; 3: *Apalabramiento: cuentos puertorriqueños de hoy*, 1983; VENEZUELA 1: *Narrativa venezolana contemporánea*, 1971; 2: *Dos siglos de prosa venezolana*, Edime, 1965. [Citations refer to Collection:Author:Page]

5. **Argentina.** Corpus lingüístico de referencia de la lengua española - Argentina. (From the Facultad de Lingüística General en la Universidad Autónoma de Madrid. Francisco Marcos-Marín, Director. Downloaded from the Internet at <ftp://lola.llf.uam.es/pub/corpus/argentina/>). The file *coarginl.txt* at the FTP site gives a complete listing of all of the texts in the corpus. We include here just the number of texts and number of words in each major division of the corpus. Humanístico: 5 texts, 379,000 words; Escolar: 2 texts, 187,000 words; Jurídico: 3 texts, 168,000 words; Científico: 4 texts, 343,000 words; Literario: 2 texts, 123,000 words; Técnico: 2 texts, 130,000 words; Periodístico: 529,000 words from 660 arti-

cles from three different newspapers; Commercial: 313 letters. [Citations refer to ARG:Title:(Chapter/Article)]

REFERENCES

- Beaman, Karen. 1984. Coordination and subordination revisited: syntactic complexity in spoken and written narrative discourse. *Coherence in spoken and written discourse*, ed. by Deborah Tannen, 45-80. Norwood, NJ: Ablex.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge Univ. Press.
- Bolinger, Dwight. 1972. The syntax of parecer. *Papers in linguistics and phonetics to the memory of Pierre Delattre*, ed. by Albert Valdman, 65-76. The Hague: Mouton.
- Butt, John and Carmen Benjamin. 1988. *A new reference grammar of Modern Spanish*. London: Edward Arnold.
- Chafe, Wallace. 1982. Integration and involvement in speaking, writing, and oral literature. *Spoken and written language: explaining orality and literacy*, ed. by Deborah Tannen, 35-54. Norwood, NJ: Ablex.
- Chafe, Wallace. 1985. Linguistic differences produced by differences between speaking and writing. *Literature, language, and learning: the nature and consequences of reading and writing*, ed. by D.R. Olson, et al., 105-23. Cambridge: Cambridge Univ. Press.
- Comrie, Bernard. 1989. *Language universals and linguistic typology: syntax and morphology*. 2nd ed. Oxford: Blackwell.
- Contreras, Heles. Clause reduction, the saturation constraint, and clitic promotion in Spanish. *Linguistic Analysis* 5.161-82.
- Davies, Mark. 1995. *Analyzing variation with computer-based corpora: the case of Modern Spanish clitic climbing*. *Hispania* 78.370-80.
- Espinosa, Aurelio M. Jr. and John P. Wonder. 1976. *Gramática analítica*. Lexington, MA: D.C. Heath and Company.
- Fernández Leboráns, María Jesús and María del Díaz Bautista. 1990. *Sobre la sintaxis del verbo español 'parecer'*. *Boletín de la Real Academia Española (BRAE)* 70.353-420.
- Gili Gaya, Samuel. 1961. *Curso superior de sintaxis española*. Madrid: Vox.
- Givón, T. 1990. *Syntax: a functional-typological introduction*. Vol 2. Amsterdam: John Benjamins.
- García Pínto, Magdalena, and Marta Luján. 1974. *On subject raising in Spanish*. *Colloquium on Spanish and Portuguese linguistics*, ed. by William T. Milan, et al. Washington, DC: Georgetown Univ. Press.
- Nuessel, Frank H. 1973. *Complement structures in Spanish*. Ph. D. dissertation, Univ. of Illinois.
- Lleó, Conxita. 1979. Some optional rules in Spanish complementation: Towards a study of the speaker's intent. Tübingen: Niemeyer.
- Miller, J. E. 1994. *Speech and writing*. *Encyclopedia of language and linguistics*, ed. by R.E. Asher, 4301-06. Oxford: Pergamon Press.
- Pawley, A. and F. H. Syder. 1983. *Natural selection in syntax: notes on adaptive variation and change in vernacular and literary grammar*. *Journal of Pragmatics* 7.551-579.
- Ramsey, M. 1956. *A textbook of Modern Spanish*. [1894] Rev. by Robert K. Spaulding. New York: Holt, Rinehart, and Winston.
- Ruiz Morales, Hildebrando. 1979. *Infinitives and patterns of sentential complements in Spanish*. Ph. D. dissertation, Univ. of Indiana.
- Sacks, Harvey, Emmanuel A. Schegloff, and Gail Jefferson. 1974. *A simplest systematics for the organization of turn-taking for conversation*. *Language* 50.696-735.
- Sauer, Keith E. 1972. *Sentential complementation in Spanish*. Ph. D. dissertation, Univ. of Washington.
- Seaver, Patricia. 1982. *The syntax and semantics of grammatical relations in reduced complement clauses in Spanish*. Ph. D. dissertation, SUNY-Buffalo.
- Subirats-Ruggeberg. 1987. *Sentential complementation in Spanish: a lexico-grammatical study of three classes of verbs*. Amsterdam: John Benjamins.

FIGURE 2

Sound wave of the word *vista* [βi'ʃ.ta] (duration in ms: [βi'=79, ʃ.ta=162]) 'view', in the context (...) *desde un punto de vista narratológico*, 'from a narratological point of view'. Speaker and procedures are the same as in Figure 1, and the pause that follows *vista* is indicated by a slash (/). The word *narratológico* is not in the figure, but it appears after the pause.

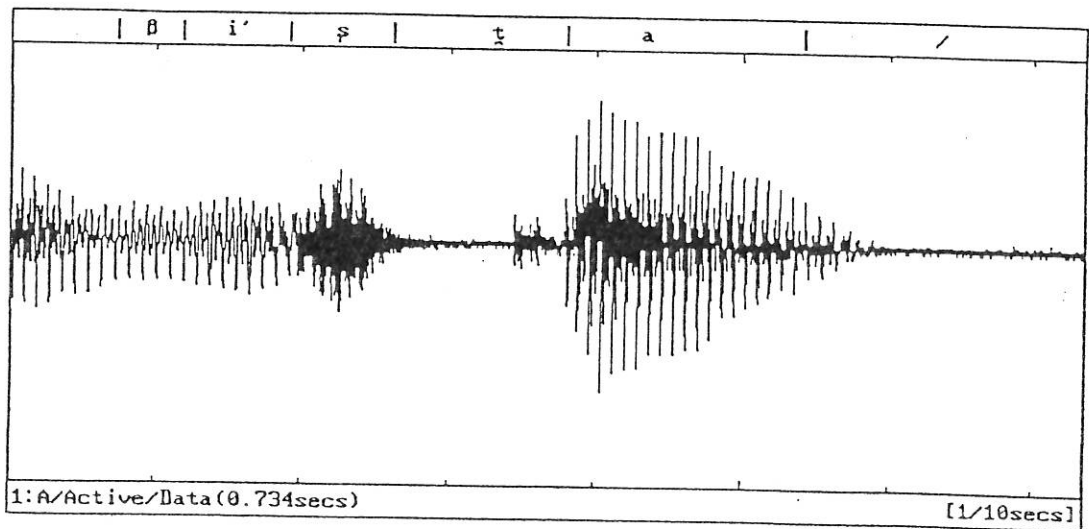


TABLE OF CONTENTS

ARTICLES

JOSÉ LUIS BLAS ARROYO <i>Dimensiones sociolingüísticas del cambio de código. A propósito de un corpus radiofónico.</i>	1
MARK DAVIES <i>A corpus based analysis of subject raising in modern Spanish.</i>	33
ANNA MARÍA ESCOBAR <i>From time to modality in Spanish in contact with Quechua</i>	64
MECEDES NIÑO-MURCIA <i>Ideología lingüística hispanoamericana en el Siglo 19: Chile (1840-1880)</i>	100
RAFAEL A. NÚÑEZ CEDENO <i>Liquid gliding in Cibaëño and feature geometry theories</i>	143
MARÍA E. PLACENCIA <i>Address forms in Ecuadorian Spanish</i>	165