# Preferred Argument Structure across time and space

## A comparative diachronic analysis of French and Spanish

William J. Ashby and Paola Bentivoglio
University of California, Santa Barbara /
Universidad Central de Venezuela

### Introduction

Du Bois' theory of Preferred Argument Structure relates observable discourse patterns and grammatical form, specifically the form of the "core" arguments of the verb: the subject and the direct object (Du Bois 1985, 1987a, 1987b). Du Bois first derived the theory from narratives in Sakapultek Maya, an ergative language. Because there is no unitary category of subject in Sakapultek, he follows the practice initiated by Dixon (1979) in denoting the subject of a one-argument verb as S, the subject of a two-argument verb as A, and the direct object as O.

Du Bois proposed that Preferred Argument Structure has both a grammatical dimension and a pragmatic dimension, expressed as "constraints" (i.e., measurable discourse preferences), as shown in Table 1. The grammatical dimension can be expressed by two constraints relating to the presence of full, lexical NP's in a clause. The One Lexical Argument Constraint reflects the paucity of clauses in which more than one of the core arguments is expressed as a lexical NP; any additional core arguments tend to be expressed as pronouns or as zero forms. The second grammatical constraint is the Non-lexical A Constraint. It reflects the tendency for the single lexical NP to occur either in the S or in the O role, and rarely in the A role.

Du Bois also expresses the pragmatic dimension of Preferred Argument Structure by a pair of constraints. The One New Argument Constraint reflects the tendency of clauses to contain no more than one piece of new information (see also Chafe 1994). Du Bois claims that this single piece of new information tends to appear in the O or S role, rarely in the A role; this patterning accounts for the Given A Constraint.

**Table 1.** Dimensions and constraints of Preferred Argument Structure

| | **Grammar** | **Pragmatics** |
|---|---|---|
| Quantity | One Lexical Argument Constraint | One New Argument Constraint |
| Role | Non-lexical A Constraint | Given A Constraint |

Source: Adapted from Du Bois (1987b: 829)

As is apparent in this volume, the patterns of Preferred Argument Structure first found by Du Bois in Sakapultek have been subsequently documented in a wide array of languages, both of the ergative-absolutive and of the nominative-accusative type.[1] Preferred Argument Structure has thus been rather extensively tested across space, but it has not as yet been extensively tested across time. The only studies of which we are aware that compare the manifestations of Preferred Argument Structure in the same language at different points in time are those of Bentivoglio (1994) for Spanish and of Josserand (1995) for hieroglyphic Maya. There is an obvious reason for this lacuna: namely, that the focus of Preferred Argument Structure has been on spoken language, primarily oral narratives and spontaneous conversational data. Oral data are obviously unavailable from the very distant past. Those wishing to study Preferred Argument Structure across time are thus obliged to work from written texts, preferably (for comparative purposes) with texts replete with dialogue and as reflective as possible of the spoken language of their time.

In this chapter we will test the Preferred Argument Structure hypothesis on an Old French text and an Old Spanish text, using the same methodology we have already brought to bear on the modern languages (Ashby 1995, Ashby & Bentivoglio 1993, Bentivoglio 1994, Ashby & Bentivoglio 1997).[2] We will then be able to compare the results for these different stages of French and Spanish in order to determine whether changes have occurred across both time and space: from medieval to modern France, and from Old Castile to present-day Latin America.

The medieval texts we have selected for our analysis are the *Chanson de Roland,* from late eleventh century France (Moignet 1969), and the *Cantar de Mio Çid*, from mid-twelfth century Castile (Menendez Pidal 1945). These texts are classics of the genre known as the medieval epic or *chanson de geste.* There has been considerable scholarly debate over whether the *Chanson de Roland*, and by extension other medieval epics, is the work of a single learned author or the written version of a traditional oral narrative.[3] Preferred Argument Structure theory may have something to contribute to this debate. That is, if the patterns of Preferred Argument Structure, which have been found manifested time and again in samples of spoken language, are also present in these medieval texts, this finding may provide confirmation of the essential oral quality of the epic and lend further support to the "traditionalist" viewpoint that sees this genre as oral literature. Thus, the comparison

of our medieval texts with our modern spoken data may not be as far fetched as it may at first appear, although the medieval epics are obviously of a different genre than the semi-formal conversational style captured in the corpora from which our modern data derive. In part, Prefered Argument Structure is based on the assumption that there are cognitive limits on planning strategies in spoken discourse. Even though they may be essentially oral in quality, the medieval epics are surely more planned than spontaneous conversation. In any case, the medieval epics are as close as we can get to spoken language of the time.

In the first section of this chapter, we will explain the methodology adopted for our study; we will then present the results of our analysis of Old French and Old Spanish, comparing them with the results of our earlier study of Modern French and Modern Spanish.

## Methodology

In order to have an equivalent amount of Old French and Old Spanish data, for each text we selected the first 800 main clauses, together with all embedded, combined and relative clauses occurring in the text from the beginning up to the eight-hundredth main clause.[4] The number of main and non-main clauses analyzed for each text is shown in Table 2.

We have followed the same methodology used in our earlier studies of the modern languages (Ashby 1995; Ashby & Bentivoglio 1993; Bentivoglio 1994; Ashby & Bentivoglio, 1997). For each corpus, every NP filling one of the core roles was coded for syntactic role, form, animacy, identifiability, and activation state, as shown in Table 3. The coded strings were then tabulated and analyzed with the help of Goldvarb 2.0 (Rand & Sankoff 1990). This program includes two applications of Sankoff's Variable Rule Program (Varbrul). One application generates probability weights for each variable, and the other selects the variables whose distributions are statistically significant.

In the Appendix, Text 1 and Text 2 give extracts of the *Chanson de Roland* and of the *Cantar de Mio Çid,* respectively, showing our coding of core arguments for syntactic role and form. Consider, for example, line 8 of Text 1, reproduced here as (1).

(1) *Vers Engletere passat li la mer salse,*
'To England crossed he the salty sea'

In (1), the transitive subject, *li* 'he', is coded as A, because it expresses the most agentive argument of the two-argument verb *passat* 'crossed' and as P, because it is a pronoun. *La mer salse* 'the salty sea' is coded as O, because it is the transitive object of *passat*, and as N, because it is a full, lexical NP.

**Table 2.** Distribution of clause types in the Old French and Old Spanish corpora

| | Old French | Old Spanish |
|---|---|---|
| Main | 800 | 800 |
| Non-main | 289 | 453 |
| Total | 1089 | 1253 |

**Table 3.** Factor groups and factors coded for each core NP token

| Factor group | Factor | Definition |
|---|---|---|
| Syntactic role | A | Transitive Subject |
| | Se | Intransitive Subject, Copular |
| | Si | Intransitive Subject, Other |
| | O | Transitive Object |
| Form of NP | N | Lexical |
| | P | Pronominal |
| | C | Clausal |
| | Ø | Zero |
| Animacy | A | Animate |
| | I | Inanimate |
| Identifiability | I | Identifiable |
| | N | Non-identifable |
| Activation | N | New |
| | G | Non-new |

For one-argument verbs we distinguish between the single argument of the copular verb 'to be' (coded Se), as in *Charles* in (2), and the argument of all remaining one-argument verbs (coded Si), as in *los gallos* 'the roosters' in (3).

(2) *Merveilus hom est Charles*
'Charles is a marvelous man' (Text 1, 6)

(3) *Apiessa cantan los gallos*
'The roosters were singing repetitively' (Text 2, 1)

Unlike Modern French, Old French is a "pro-drop" language, as are Old and Modern Spanish. Thus, the form of the core arguments may not only be pronominal (P) and lexical (N), as illustrated in (1), (2), and (3), but may also be zero (Ø), as with the S argument of Fr. *asemblet s'est* '(he) has joined' in (4) and the A argument of Sp. *sopieron* '(they) heard' of (5):

(4) *Asemblet s'est as sarrazins messages*
'(He) has joined with the heathen messangers' (Text 1, 2)

(5) *i sopieron el mandado*
'and (they) heard the news' (Text 2,11)

We also included the clausal form (C), as in (6), where the infinitival clause, Spanish *crebar albores* 'to break day,' is coded as the O argument of the two-argument verb *quieren*, which is C in form.

(6) *e quieren crebar albores*
'and they wanted to break day' (Text 2, 2)

The arguments within this embedded infinitival clause were also coded. Thus, two arguments of *crebar* are coded: an A argument, which takes Ø form, and an O argument, *albores*, which is expressed as an N.[5]

We also considered the complement of verbs of saying to be of C form, as in (7), where the extended quoted speech is considered the O argument of the two-argument verb *respunt* 'replies'. The arguments of each verb within the quoted speech were also coded.

(7) *Guenes respunt: "Itels est sis curages.*
'Ganelon replys: "Such is his disposition.'
*Jamais n'ert hume ki encuntre lui vaille."*
'Never will there be a man who measures up to him."' (Text 1, 11–12)

## Results

### *The grammatical dimensions of Preferred Argument Structure*

Table 4a and Table 4b show the distribution of NP's of lexical, non-lexical, and clausal form in the various grammatical roles (A, Si, Se, O, C) in the Old French and Old Spanish corpora, respectively. Table 4a and Table 4b show that the non-lexical form (that is, pronoun and zero taken together) predominates for all but the O role. As will be demonstrated shortly, this is because the O role is preferred for the introduction of new information, and new information must perforce take lexical form. Table 4a and Table 4b support Du Bois' Non-Lexical A Constraint insofar as it is the A role that has the lowest ratio of lexical NP's of any of the roles (26% for Old French and 16% for Old Spanish). Note that most A's are zeros (56% for Old French and 74% for Old Spanish). This finding reflects the fact that referents coded in the A role tend to be animate and continuous[6].

Figure 1 shows the distribution of lexical NP's across core syntactic roles in the Old French and Old Spanish corpora, and project a comparison of these distributions

to those obtaining in our Modern French and Modern Spanish data (reported in Ashby & Bentivoglio 1993).[7] Contrary to what we would expect from the prediction of the Non-Lexical A Constraint, and contrary to what obtains in Modern French and Modern Spanish, in both Old French and Old Spanish we see a relatively high percentage of lexical NP's occuring in the A role (13% in Old French and 12% in Old Spanish, compared to 5% in Modern French and 6% in Modern Spanish).[8] We believe this relatively high ratio of lexical NP's in the A role role is specific to the epic genre and is in part the result of the frequent repetition of proper names. For example, lines 1 and 17 of Text 3, reproduced here as (8) and (9), are identical, except for the word order of the O relative to the verb, and both begin with a lexical A, *li reis Marsile* 'King Marsile'.

**Table 4a.** Distribution of NP's according to grammatical role and form in the Old French corpus

| Role | Lexical | | Pronoun | | Zero | | Clause | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N | (%) | N | (%) | N | (%) | N | (%) | N | (%) |
| A | 175 | (26) | 126 | (18) | 384 | (56) | 0 | (0) | 685 | (37) |
| Se | 48 | (44) | 28 | (26) | 33 | (30) | 0 | (0) | 109 | (5) |
| Si | 139 | (47) | 61 | (21) | 95 | (32) | 0 | (0) | 480 | (24) |
| O | 403 | (53) | 173 | (23) | 5 | (1) | 178 | (23) | 730 | (37) |
| Total | 765 | (41) | 388 | (21) | 517 | (28) | 178 | (10) | 2004 | |

**Table 4b.** Distribution of NP's according to grammatical role and form in the Old Spanish corpus

| Role | Lexical | | Pronoun | | Zero | | Clause | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N | (%) | N | (%) | N | (%) | N | (%) | N | (%) |
| A | 115 | (16) | 76 | (11) | 506 | (73) | 0 | (0) | 697 | (35) |
| Se | 27 | (29) | 26 | (28) | 40 | (43) | 0 | (0) | 93 | (5) |
| Si | 158 | (33) | 78 | (16) | 241 | (50) | 3 | (1) | 480 | (24) |
| O | 351 | (48) | 249 | (34) | 0 | (0) | 130 | (18) | 730 | (37) |
| Total | 651 | (33) | 42 | (21) | 787 | (39) | 133 | (7) | 2000 | |

(8) *Li reis Marsile out sun cunseill finet.*
'King M. had held his counsel.' (Text 3,1)

(9) *Li reis Marsile out finet sun cunseill.*
'King M. had held his counsel.' (Text 3,17)

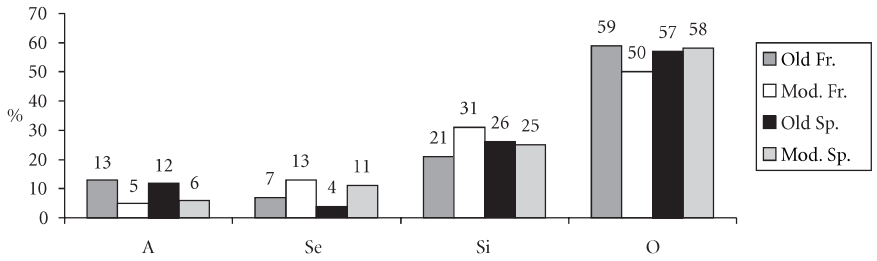


**Figure 1.** Comparison of lexical NP's in Old French, Modern French, Old Spanish and Modern Spanish, by syntactic role.

The second of these verses appears neither to convey new information nor to advance the story line, but it does have a certain rhetorical effect and perhaps had mnemonic value for the medieval *jongleur.* Similar repetitions occur in the Old Spanish data, as in lines 1 and 4 of Text 4, reproduced here as (10) and (11).

(10) *Mio Çid don Rodrigo a la puerta adelinyava;*
'My Çid don Rodrigo was advancing toward the door;'

(11) *Mio Çid Ruy Días por las puertas entrava…*
'Mio Çid Ruy Díaz entered through the doors…'

Here, too, it seems that the use of the second lexical S, *Mio Çid Ruy Días,* is not so much required by discourse constraints as by rhetorical or aesthetic considerations (see Martin, this volume, on the use of repetition in Mocho narrative).[9]

According to Du Bois' One Lexical Argument Constraint, it is unlikely that both the A and the O roles in a given clause will be lexical in form. This constraint is strongly supported by the Old French and Old Spanish data. Table 5a and Table 5b give the distributions of lexical and non-lexical tokens in the A and O roles in the Old French and Old Spanish corpora, respectively. These tables demonstrate that there are only 37 clauses in the Old French corpus and 27 clauses in the Old Spanish corpus that contain both a lexical A and a lexical O.

Figure 2 compares the distribution of A and O in the same clause in Old French, Modern French, Old Spanish, and Modern Spanish. The distributions shown are very similar, not only within the same language, but also between French and Spanish. In all four data sets, there is a low percentage of lexical A and lexical O in the same clause: only 7% in Old French, 5% in Modern French, 5% in Old Spanish, and 2% in Modern Spanish. We can thus affirm that both Old French and Old Spanish follow the One Lexical Argument Constraint, just as do the modern languages.

Summarizing the results of the analysis thus far, we have shown that both Old French and Old Spanish conform to the grammatical constraints of Preferred Argument Structure. Both medieval texts show that the preferred form for the A role is

**Table 5a.** Distribution of A and O in the same clause, by form (N, P) in Old French

| | A-N | | A-P* | | Total | |
|---|---|---|---|---|---|---|
| | N | (%) | N | (%) | N | (%) |
| O-N | 37 | (11) | 286 | (89) | 323 | (48) |
| O-P | 46 | (26) | 130 | (74) | 176 | (26) |
| O-C | 88 | (50) | 89 | (50) | 177 | (26) |
| Total | 171 | (25) | 505 | (75) | 676 | |

*Combines P and Ø

**Table 5b.** Distribution of A and O in the same clause, by form (N, P) in Old Spanish

| | A-N | | A-P* | | Total | |
|---|---|---|---|---|---|---|
| | N | (%) | N | (%) | N | (%) |
| O-N | 27 | (9) | 286 | (91) | 313 | (46) |
| O-P | 43 | (17) | 205 | (83) | 248 | (36) |
| O-C | 36 | (29) | 88 | (71) | 124 | (18) |
| Total | 106 | (15) | 579 | (85) | 676 | |

*Combines P and Ø

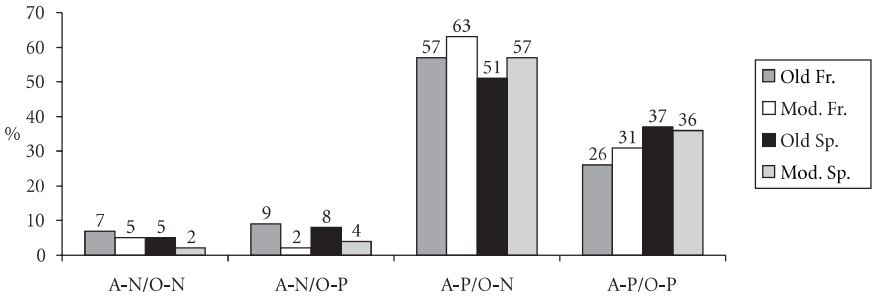


**Figure 2.** Distribution of lexical (N) and non-lexical (P) forms in A and O roles in same clause, in Old French, Modern French, Old Spanish, and Modern Spanish.

either pronoun or zero, thus supporting the Non-Lexical A Constraint. Both medieval texts show a low incidence of lexical NP's in both the A and O roles, thus confirming the One Lexical Argument Constraint. Despite some quantitative differences between the medieval and modern corpora — differences that seem to be genre dependent — we can affirm that both Old and Modern French, and Old and Modern Spanish exemplify the grammatical constraints of Preferred Argument Structure.

### *The pragmatic dimensions of Preferred Argument Structure*

We now turn to the pragmatic dimensions of Preferred Argument Structure, focussing on how new and non-new information is linguistically encoded in the Old French and Old Spanish corpora. In Ashby & Bentivoglio (1993), we demonstrated that, when they encode new participants to the discourse, speakers of Modern French and Modern Spanish do not randomly choose one of the core grammatical roles, but instead give preference to the O and Si roles, avoiding not only the A role, but also the Se role. To test this hypothesis on the Old French and Old Spanish data, we set activation state (new vs. non-new) as the dependent variable for the Goldvarb analysis. The results for Old French are displayed in Table 6a and for Old Spanish in 6b. These tables also give the results for Modern French and Modern Spanish, as reported in Ashby & Bentivoglio (1993). Figure 3 presents these same distributions in graphical form.

Table 6a and Table 6b, together with Figure 3, provide dramatic confirmation of the Given A Constraint not only in the Modern French and Modern Spanish data, but in the Old French and Old Spanish data as well. In all four data sets, it is the A role that has the lowest Goldvarb probability weight for encoding referents that are new to the discourse. The weight for A in Old French is only .198; in the Modern French data reported in Ashby & Bentivoglio (1993), there were no new A's at all (hence, no probability weight is given).

Likewise in Modern Spanish, the probability of new A's was very low, only .197. In the Old Spanish data, it was higher, .415, although still the lowest of the four weights given for Old Spanish in Table 6b. It must be pointed out, however, that Goldvarb did not select role as a statistically significant factor in the analysis of the Old Spanish data; hence the probability weights given for Old Spanish in Table 6b are in brackets. (The fact that role was not selected by Goldvarb does not necessarily indicate that it is an unimportant determinant of the distribution of new vs. non-new tokens; it may simply indicate that there were simply not enough data to meet Goldvarb's test of significance.)

We also see from the data given in Table 6a, Table 6b, and Figure 3 that in all four data sets it is the O role that has by far the highest probability of encoding

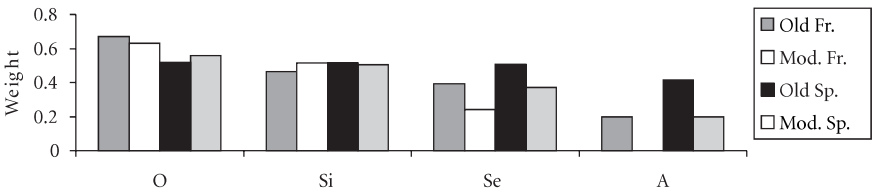


**Figure 3.** Distribution of new NP's in Old French, Modern French, Old Spanish, and Modern Spanish, by role.

**Table 6a.** Distribution of new NP's in Old French and Modern French, by role

| | Old French | | | Modern French | | |
|---|---|---|---|---|---|---|
| **Role** | **Count** | **% New** | **Weight** | **Count** | **% New** | **Weight** |
| O | 283/403 | (70) | .670 | 143/324 | (44) | .629 |
| Si | 52/139 | (37) | .465 | 64/203 | (32) | .517 |
| $S^e$ | 15/ 48 | (31) | .393 | 11/87 | (13) | .241 |
| A | 19/175 | (11) | .198 | 0/32 | – | – |
| | N = 765 | Input = .460 | | N = 645 | Input = .314 | |

**Table 6b.** Distribution of new NP's in Old Spanish and Modern Spanish, by role

| | Old Spanish | | | Modern Spanish | | |
|---|---|---|---|---|---|---|
| **Role** | **Count** | **% New** | **Weight** | **Count** | **% New** | **Weight** |
| O | 121/351 | (35) | [.520] | 142/341 | (42) | .558 |
| Si | 29/158 | (18) | [.517] | 44/150 | (29) | .506 |
| Se | 5/ 27 | (5) | [.509] | 12/ 65 | (18) | .372 |
| A | 11/115 | (10) | [.415] | 2/ 35 | (6) | .197 |
| | N = 651 | Input = .229 | | N = 591 | Input = .311 | |

referents that are new. Thus, we are confident in affirming that O favors new information and A disfavors new information, both in the Old and the Modern stages of French and Spanish.

The role of S is less clear, however. In Ashby & Bentivoglio (1993) we reported that S also appeared to be a site that favored new information, just as expected from Preferred Argument Structure theory. Thus, Si shows a weight of .517 for Modern French (Table 6a) and of .506 for Modern Spanish (Table 6b). These weights are not strong, however, being only slightly above the neutral point of .500. In Ashby & Bentivoglio (1997), we found that many of the tokens we had coded as S in Ashby & Bentivoglio (1993) were in fact arguments of the single-argument presentative structures, *il y a* (French) and *hay* (Spanish) 'there is/there are'. Once these frequently occurring presentatives were split off from the rest of the S tokens, the probability of new S dropped to .314 in Modern Spanish and to only .138 in Modern French. Although the presentative structure proved to be very frequent in the modern languages, they were relatively infrequent in our medieval corpora: 31 tokens in the Old French data and only 2 tokens in the Old Spanish data.[10] Since there were so few such tokens in our medieval data, we decided to revert to our original practice (Ashby & Bentivoglio 1993) and to group NP's occurring after these presentatives with the other Si tokens. Even so, Table 6a shows that the

probability weight associated with encoding new information in the Si role in Old French is only .465; Table 6b shows that it is .517 for Old Spanish, just above the neutral point of .500. Thus, S does not appear to be a role that clearly and strongly favors the introduction of new information in any of our data sets. In all cases, however, the probability weights for Si are higher than for A, as we would expect. (We have not examined the possible role of "information pressure" [Du Bois 1987b: 836] on the variable distribution of new S in our four data sets.)

We now turn to the probability weights for Se. It was shown by Ashby & Bentivoglio (1993) that in Modern French and Modern Spanish the Se role strongly disfavors new information.[11] As reported in Table 6a, the probability weight for new Se in the Modern French corpus is only .241; Table 6b shows that it is .372 in Modern Spanish. At .393, the weight for Old French is likewise low. In Old Spanish the probability weight for new Se is higher (.509); but it is unclear whether this is meaningful, since there were so few tokens in Se in Old Spanish. (Recall that for the Old Spanish data, the distributions were not selected by Goldvarb as being statistically significant.)

We can thus arrive at the following tentative conclusions from the information given in Tables 6a and 6b:

1. In all of the data sets, A is strongly disfavored for encoding new referents; this finding provides confirmation of the Given A Constraint.
2. In all of the data sets, O is strongly preferred as the site for encoding new referents.
3. It is not obvious that S is a role that favors new referents, contra the predictions of Preferred Argument Structure. This is further evidence of the "ambivalent status of the S role" (Durie, this volume).
4. The Se role strongly disfavors new referents in all but Old Spanish. (Again, we note that Goldvarb did not select role for the Old Spanish data; moreover, there are relatively few Se tokens in Old Spanish.)

Let us now consider the pragmatic status of the arguments in the Old French and Old Spanish tokens which contain both a lexical A and a lexical O, in order to evaluate the One New Argument Constraint. Of the 37 Old French tokens, only three have *new* NP's in *both* the A and O roles. They are represented in (12), (13), and (14).

(12) O A
*Mult grand eschech en unt si chevaler*
much great booty of-it had his knights
'His knights got great booty from it' (VIII, 99)

(13) O A
*L'estreu li tint sun uncle Guinemer.*
the-stirrup DAT-3RD-SG held his uncle Guinemer
'His uncle Guinemer held the stirrup for him.'

A

(14) *Reis Almaris del regne de Belferne*
king Almaris of-the kingdom of Belferne
'King Almaris from the kingdom of Belferne'

O

*Une bataille lur livrat le jur pesme.*
a battle DAT-3PL delivered that day worst
'gave them a terrible battle that day.' (LXV, 813)

In none of these examples are *both* NP's "brand new" (Prince 1992), that is, both new and non-identifiable. In (12) the booty (*mult grand eschech*) can be considered both new and non-identifiable, but the knights (*sis chevalers*) are made identifiable by the possessive adjective and by schematic association with the emperor. In verses (13) and (14), one of the new arguments is a proper name, which can also be considered an indentifiable referent (see Chafe 1994:93–107 on identifiability). Similarly, the single Old Spanish example containing new NP's in both the A and O roles, shown as (15), has referents that are identifiable, hence not brand-new.

A O

(15) *el abbat don Sancho… rezaba los matines …*
'the abbot don Sancho was reciting the morning prayers…' (I,14,237)

The small number of tokens with transitive verbs in which both arguments are new, and the total absence of tokens in which the two arguments are both new and non-identifiable constitute strong evidence in support of the One New Argument Constraint.[12]

## Conclusion

Our analysis of a sample of Old French and Old Spanish must be considered a preliminary approach to the study of Preferred Argument Structure across time. It will be necessary to extend this study to other time periods and to other genres. The results obtained nevertheless validate the general hypothesis as it relates both to the grammatical and pragmatic dimensions of Preferred Argument Structure.

Our medieval and modern data sets are clearly of different genres. The modern data come from recorded conversations, where the contribution of the French and Spanish-speaking interlocutors represents discourse of an unplanned nature; consequently, it is not surprising to find that Preferred Argument Structure is followed. The medieval data, on the other hand, consist of samples of epic poetry, a genre that comes to us in written form, a genre that is surely more planned than is conversation. Nevertheless, our counts have demonstrated that Preferred

Argument Structure is manifested even in the medieval epics. This finding may support the hypothesis that the medieval epics are essentially oral in quality. In any case, we have demonstrated the robustness of Preferred Argument Structure in two very different genres. Despite this difference in genre, and despite a gap of seven centuries, we have shown that the medieval and modern forms of French and Spanish are remarkably similar in their manifestations of Preferred Argument Structure.

## Notes

1. Three Romance languages have been shown to exhibit the characteristics of PAS. They are Brazilian Portuguese (Dutra 1987), French (Ashby 1995, Ashby & Bentivoglio 1993, Ashby & Bentivoglio 1997, Lambrecht 1987, Lambrecht 1988) and Spanish (Ashby & Bentivoglio 1993, Bentivoglio 1992, Bentivoglio & Ashby 1993, Ocampo 1993).

2. The Modern French data were recorded by W. J. A. in the French city of Tours in 1976. The Modern Spanish data were recorded in 1987 in Caracas, Venezuela, by university students under the direction of P. B. See Ashby & Bentivoglio (1993) for details.

3. Enders (ms.) notes "three major schools of thought" in the debate over the nature of the epic, as represented by the *Chanson de Roland*. She characterizes these schools as "the individualists, the traditionalists, and a position somewhat midway between these two". She explains the debate in the following terms:

> The individualists see the *Chanson de Roland* as the creation of one cultivated author, and argue that the epic is too well written, too near perfection to be the product of an oral literature tradition. On the other hand, the traditionalists see the *CR* as a traditional oral poem, passed down from performer to performer. The midway position attempts to reconcile these two views by claiming that a man of genius took a poem of the oral tradition and transformed it from a rude song of battle into a highly idealistic work.

As for Moignet, he is obviously of the traditionalist school (1969: 5):

> Bien que le texte des chansons que nous avons nous soit parvenu par la tradition écrite…, le genre appartient, de soi, à la littérature orale: il est créé et transmis par des *jongleurs*, à la fois auteurs, remanieurs et interprètes, qui pratiquent la récitation publique. Rien n'est plus éloigné de l'art des jongleurs que l'idée, toute moderne, d'un texte fixe, propriété littéraire d'un auteur: la matière équipe est un bien commun que chacun peut traiter à sa guise.

See also Lord (1960).

4. All clause types were included (declarative, interrogative, exclamative, exhortative). Clauses containing impersonal verbs (e.g., Sp. *mucho es huebos* 'it is very necessary') and time-indicating verbs (e.g., Sp. *Antes que anochesca* 'Before it gets dark') were not included, however.

5. The tokens of C form were not coded for animacy, identifiability, or activation state, but the NP's within these embedded clauses were coded for these factors.

6. Interestingly, the overall percentage of the null form is higher in Old Spanish than in Old French (39% for Old Spanish, only 28% for Old French). This suggests that even though Old French is considered a pro-drop language, it was already using a higher ratio of pronouns than

Old Spanish. This trend apparently continues, with the pronoun having become obligatory in Modern French, but not in Modern Spanish.

7. The tokens representing the C role, shown in Table 3a and Table 3b are not included in Figure 1, because clausal arguments were not included in our previous analysis of the modern languages.

8. Du Bois (1987a) found that 5% of A's were lexical in Sakapultek; Dutra (1987) found 8% lexical A's in Brazilian Portuguese.

9. On the other hand, proper nouns are relatively low on Ariel's "accessibility scale" (Ariel 1990), suggesting that the repetition of the proper noun may be necessary for cognitive reasons.

10. The verb on which this presentative is based is 'to have'. French *il y a* derives from Vulgar Latin *ibi habet,* literally 'there has' (*ibi* > *y* and *habet* > *a*, with the impersonal pronoun *il* added beginning in Old French). Spanish *hay* comes from *habet ibi.* Traditionally, the NP following this verb is considered the object, because in Latin it was marked for accusative case. Indeed, this NP exhibits some features of the direct object in the grammar of Modern French and Spanish (in pronominalization and in interrogation, for example). Nevertheless, in Ashby & Bentivoglio (1993), we considered it to be an S, rather than an O, because it is the sole argument of the verb. In colloquial spoken French, *il* is optional, and no other pronominal or nominal argument is possible. In Spanish, no additional argument can be used with existential *hay* (e.g., *Hay muchas muñecas* 'There are a lot of dolls', but not **Ello hay muchas muñecas*).

11. In Ashby & Bentivoglio (1993), the Se role was called X.

12. We also note in Table 6a and Table 6b that all of the referents coded as A are [+animate], and that all of the referents coded as O are [−animate].

## Appendix

**Text 1.** Extract from the "Chanson de Roland", coded for syntactic role and form of core NP's

1 Si-N ***Guenes*** *chevalchet suz une olive halte,*
'Ganelon rides under a tall olive tree,' (XXVIII, 366)

2 Si-Ø (**Ø**) *Asemblet s'est as sarrazins messages;*
'(He) has joined with the heathen messangers;' (367)

3 Si-P *Mais Blancandrins* ***ki*** *envers lu s'atarguns a l'altre.*
'But (here is) Blancandrin who is lingering near.' (368)

4 Si-P *Par grant saveir parolet* ***li uns*** *a l'altre.*
'The one talks to the other with great skill.' (369)

5 A-N *Dist* ***Blancandrins:***
'B. says,' (370)
O-C ***"Merveilus hom …marche."***
'A marvelous man…province' (370–374)

6 Se-N *Merveilus hom est* ***Charles,***
'Charles is a marvelous man' (370)

7 A-P ***ki*** *cunquist*
'who conquered'

O-N ***Puille e trestute Calabre!***
'Apulia and Calabria!' (371)

8 A-P *Vers Engletere passat* ***li***
'To England he crossed'
O-N ***la mer salse,***
'the salty sea' (372)

9 A-Ø *Ad oes seint Perre en* ***(Ø)*** *cunquist*
'he won (the tribute) for Saint Peter'
O-N ***le chevage:***
'the tribute' (373)

10 O-P ***Que***
'What'
A-0 *nus****(0)*** *requert ça en la nostre marche?*
'does he ask of us here in our province?' (374)

11 A-N ***Guenes*** *respunt:*
'G. replies:'
O-C ***"Itels est sis curages…lui vaille."***
'Such is his disposition…measures up to him.' (375–376)

12 Se-N ***Itels*** *est sis curages.*
'Such is his disposition.' (375)

13 Se-N *Jamais n'ert* ***hume***
'Never will there be a man'

14 Si-P ***ki*** *encuntre lui vaille.*
'who measures up to him' (376)

15 A-N *Dist* ***Blancandrins:***
'Says B.'
O-C ***"Francs….confundent"***
'The Franks…overwhelm (him).' (XXIX, 377–379)

16 Se-N ***Francs*** *sunt mult gentilz home!*
'The Franks are very noble men!' (377)

17 O-N ***Mult grant mal***
'Great harm'
A-N *funt e* ***cil duc***
'do both these dukes'
A-N *e* ***cil cunte***
'and these counts' (378)

18 A-P *A lur seignur,* ***ki***
'to their lord, (they) who'
O-N ***tel conseill*** *li dunent:*
'give him such counsel:' (379)
O-P e ***altrui***
'and others'
A-Ø ***(Ø)*** *travaillent*
'(They) torment'

A-Ø *e (**0**) confundent."*
'and (they) overwhelm' (380)

20 Si-N ***Un faldestoet** out suz l'umbre d'un pin;*
'There was a throne in the shade of a pine' (XXXI, 407)

21 Se-Ø *Envolupet (**Ø**) fut d'un palie alexandrin:*
'(It) was covered with a cloth of silk from Alexandria.' (408)

22 Si-N *La fut **li reis***
'There was the king'

23 A-P *ki*
'who'

O-N ***tute Espaigne** tint;*
'held all of Spain.' (409)

24 *Tut entur lui vint milie Sarrazins*
'Around him, twenty thousand Sarrasins.' (410)

25 Si-P N'i a *celoi*
'There are none'

26 A-P ***ki***
'who'

O-N ***mot** sunt*
'sound a word'

27 A-Ø *ne (**Ø**)*
'or (who)'

O-N *mot* tint;
'chime a word' (411)

**Text 2.** Extract from the "Cantar de mio Çid", coded for syntactic role and form of core NP's

1 Si-N *Apriessa cantan **los gallos***
'The roosters were singing repetitively'

2 A-Ø *e (**Ø**) quieren*
'And (they) wanted'

O-C ***crebar albores,***
'to break the day' (14, 235)

3 A-Ø *(**0**) crebar*
'to break'

O-N *albores*
'day'

4 Si-N *quando llegó a San Pero **el buen Campeador;***
'when the good *Campeador* arrived at San Pero' (236)

5 A-N *el abbat don Sancho, cristiano del Criador, rezaba*
'the abbot, Don Sancho, a Creator's Christian, was reciting' (237–238)

O-N ***los matines** abuelta de los albores.*
'the morning prayers at dawn.' (238)

6 Si-N *Y estava **doña Ximena** con çinco duenyas de pro,*
'And Doña Ximena with five good ladies was there' (239)

7 A-Ø *(**Ø**) rogando a San Pero e al Criador:*
'imploring Saint Peter and the Creator' (240)
O-C ***"Tú que a todos guías, val a mio Çid el Campeador.**"*
'You who guide everyone, protect my Cid the *Campeador*' (241)
8 A-P ***Tú** que… guías*
O-P ***a todos***
9 A-Ø *(**Ø**) val*
O-N ***a mio Çid el Campeador***
10 Si-Ø *(**Ø**)Llamavan a la puerta,*
'They were knocking at the door'
11 A-Ø *i (**Ø**) sopieron*
'and heard'
O-N ***el mandado;***
'the news' (15, 242)
12 Se-N *Dios, qué alegre fo **el abbat don Sancho!***
'God, how happy the abbot Don Sancho was!" 243)
13 A-Ø *Con lumbres e con candelas al corral (**Ø**) dieron* [*salto*],
'With lights and candles (they) went out from (*lit.* gave a jump from) the courtyard' (244)
O-N *salto,*
'(a) jump'
14 A-Ø *con tan grant gozo (**Ø**) reçiben*
'with such great joy (they) welcome'
O-P ***al*** [*que en buena hora nasco*]
'the one [who was born happily]'
15 Si-P ***que** en buena hora nasco. (245)*
16 O-C ***"Gradéscolo a Dios, mio Çid…ospedalo."***
'I am grateful to God (*lit.* I accept it with pleasure to God), my Cid…hospitality.' (246–247)
A-N *dixo **el abat don Sancho;***
'said the abbot Don Sancho'
17 A-Ø *(**Ø**) gradésco*[*lo*]
'(I) accept with pleasure'
O-P ***lo***
'it'
18 A-Ø *"pues que aquí vos(**Ø**) veo,*
'because (I) see you here'
O-P ***vos***
'you'
19 A-Ø *(**Ø**) prendet de mí*
'take from me'
O-N ***ospedado.***
'hospitality'

**Text 3.** Extract from the "Chanson de Roland"

1 *Li reis Marsile out sun cunseill fenet,*
'King Marsile had held his counsel' (V, 62)
2 *Sin apelat Clarin de Blaguet,*
'He called Clarin de Balaguer,' (63)
3 *Estamarin e Eudropin, sun per,*
'Estramin and Eudropin, his peer,' (64)
4 *E Priamun e Guarlan le barbet*
'And P. and G. the bearded one' (65)
5 *E Machiner e sun uncle, Maheu,*
'And M. and his uncle, M.' (66)
6 *E Joüner e Malbien d'ultremer*
'And J. and M. from across the sea,' (67)
7 *E Blancandrins, por la raisun cunter.*
'and Blancandrin, in order to lay out his plan. (68)
8 *Des plus feluns dis en ad apelez:*
'Among the most treacherous, he called ten.' (69)
9 *"Seignurs baruns, a Carlemagnes irez.*
'Lord Barons, you will go to Charlemagne' (70)
10 *"Il est al siege a Cordres la citet.*
'He is at the siege of the city of Cordres. (71)
11 *"Branches d'olives en voz mains porterez*
'You will carry olive branches in your hands' (72)
12 *"Ço senefiet pais et humilitet.*
'This signifies peace and humility.' (73)
13 *"Par voz saveirs sem puez acorder,*
'If, by your skill, you can make an agreement between him and me,' (74)
14 *"Je vos durrai or e argent asez,*
'I will give much gold and silver,' (75)
15 *"Teres e fiez tant cum vos en vuldrez."*
'Lands and fiefdoms, as many as you wish.' (76)
16 *Dient paien: "De ço avun nus asez!"*
'The pagans reply, "With that we will be satisfied.' (77)
17 *Li reis Marsile out finet sun cunseill.*
'King M. had held his counsel.' (VI, 78)

**Text 4.** Extract from the "Cantar de Mio Çid"

1 *Mio Çid don Rodrigo a la puerta adelinyava;*
'Mio Çid don Rodrigo was advancing toward the door;' (23, 467)
2 *los que la tienen, quando vidieron la rebata,*
'those who hold it [the door] when they saw the sudden assult,' (468)
3 *ovieron meido e fo desenparada*
'they had fear and it [the door] was abandoned' (469)
4 *Mio Çid Ruy Días por las puertas entrava….*
'Mio Çid Ruy Díaz entered through the doors…' (470)

## References


Ariel, Mira. 1990. *Accessing Noun-Phrase Antecedents.* London and New York.

Ashby, William J. 1995. "French presentational structures". In Jon Amastae et al. (eds), *Contemporary Research in Romance Linguistics.* Amsterdam & Philadelphia: Benjamins, 91–104.

Ashby, William J. and Paola Bentivoglio. 1993. "Preferred Argument Structure in spoken French and Spanish". *Language Variation and Change* 5: 61–76.

Ashby, William J. 1997. "Strategies for introducing new referents into discourse: a comparative analysis of French and Spanish presentational structures". In Robert M. Hammond and Marguerite G. MacDonald (eds), *Linguistic Studies in Honor of Bohdan Saciuk.* West Lafayette, IN: Learning Systems, Incorporated, 9–26.

Bentivoglio, Paola. 1994. "Spanish Preferred Argument Structure across time and space". *D.E.L.T.A. (Revista de Documentaçao de Estudos em Lingüística Teórica e Aplicada)* 10: 277–293.

Chafe, Wallace L. 1980. "The deployment of consciousness". In W. Chafe (ed.), *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production.* Norwood, NJ: Ablex, 7–40.

Chafe, Wallace L. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing.* Chicago: University of Chicago Press.

Chafe, Wallace L. 1987. "Cognitive constraints on information flow". In Russell Tomlin (ed.), *Coherence and Grounding in Discourse.* Amsterdam & Philadelphia: Benjamins, 21–51.

Dixon, Robert M. W. 1979. "Ergativity". *Language* 55: 59–138.

Du Bois, John W. 1985. "Competing motivations". In John Haiman (ed.), *Iconicity in Syntax.* Amsterdam & Philadelphia: Benjamins, 343–65.

Du Bois, John W. 1987. "The discourse basis of ergativity". *Language* 63: 805–55.

Du Bois, John W. and Sandra A. Thompson. 1991. "Dimensions of a theory of information flow". Unpublished ms., University of California, Santa Barbara.

Duggan, Joseph J. 1973. *The Song of Roland.* Berkeley: The University of California Press.

Dutra, Rosalia. 1987. "The hybrid S category in Brazilian Portuguese: Some implications for word order". *Studies in Language* 11: 163–180.

Enders, Jody. Review of Joseph J. Duggan, *The Song of Roland.* Unpublished ms, University of California, Santa Barbara.

England, Nora C. 1986. "Mamean voice: syntactic and narrative considerations". Unpublished ms. University of Iowa.

Josserand, J. Kathryn. 1995. "Participant tracking in Maya hieroglyphic texts: Who was that masked man?" *Journal of Linguistic Anthropology* 5: 65–89.

Lambrecht, Knud. 1987. "On the status of SVO sentences in French discourse". In Russell Tomlin (ed.), *Coherence and Grounding in Discourse.* Amsterdam & Philadelphia: Benjamins, 217–61.

Lambrecht, Knud. 1988. "Presentational cleft constructions in spoken French". In John Haiman and Sandra A. Thompson (eds), *Clause Combining in Grammar and Discourse.* Amsterdam & Philadelphia: Benjamins, 135–79.

Lord, Albert B. 1960. *The Singer of Tales.* Cambridge, MA: Harvard University Press.

Martin, Laura. 2002. "Narrator virtuosity and the strategic exploitation of Preferred Argument Structure in Mocho narrative: repetition and constructed speech in Mocho narrative. In John W. Du Bois et al., *Preferred Argument Structure: Grammar as Architecture for Function.* Amsterdam & Philadelphia: Benjamins.

Menéndez Pidal, Ramón (ed.). 1945. *Cantar de mio Cid. Texto, Grammática y Vocabulario.* Madrid: Espasa-Calpe.

Moignet, Gérard (ed.) 1969. *La Chanson de Roland.* Paris: Bordas.

Ocampo, Francisco. 1993. "The introduction of new referents in French and Spanish discourse: One constraint, two strategies". In William J. Ashby et al. (eds), *Linguistic Perspectives on the Romance Languages.* Amsterdam & Philadelphia: Benjamins, 351–362.

Prince, Ellen. 1992. "Subjects, definites, and information-status". In William C. Mann and Sandra A. Thompson (eds), *Discourse Description: Diverse Linguistic Analyses of a Fund-raising Text.* Amsterdam & Philadelphia: Benjamins, 295–325.

Rand, David and David Sankoff. 1990. GoldVarb 2.0 (computer program obtained from authors). Montreal: University of Montreal.

Sankoff, David. 1990. "Variable rules". In David Rand and David Sankoff, *GoldVarb Version 2: A Variable Rule Application for the Macintosh.* Montreal: University of Montreal, 1–26.