

Third Person References

*Forms and functions
in two spoken genres of Spanish*

Jenny Dumont

John Benjamins Publishing Company

Third Person References

Studies in Functional and Structural Linguistics (SFSL)

ISSN 1385-7916

Taking the broadest and most general definitions of the terms functional and structural, this series aims to present linguistic and interdisciplinary research that relates language structure – at any level of analysis from phonology to discourse – to broader functional considerations, whether cognitive, communicative, pragmatic or sociocultural. Preference will be given to studies that focus on data from actual discourse, whether speech, writing or other nonvocal medium.

The series was formerly known as *Linguistic & Literary Studies in Eastern Europe (LLSEE)*.

For an overview of all books published in this series, please see

<http://benjamins.com/catalog/sfsl>

Founding Editor

John Odmark

Honorary Editors

Eva Hajičová

Charles University

Petr Sgall

Charles University

General Editors

Yishai Tobin

Ben-Gurion University of the Negev

Bob de Jonge

Groningen University

Editorial Board

Alexandra Y. Aikhenvald

La Trobe University

Joan L. Bybee

University of New Mexico

Ellen Contini-Morava

University of Virginia

Nicholas Evans

University of Melbourne

Victor A. Friedman

University of Chicago

Anatoly Liberman

University of Minnesota

James A. Matisoff

University of California, Berkeley

Jim Miller

Emeritus, University of Edinburgh

Marianne Mithun

University of California, at Santa Barbara

Lawrence J. Raphael

CUNY and Adelphi University

Olga Mišeska Tomić

Leiden University

Olga T. Yokoyama

UCLA

Volume 71

Third Person References. Forms and functions in two spoken genres of Spanish
by Jenny Dumont

Third Person References

Forms and functions
in two spoken genres of Spanish

Jenny Dumont
Gettysburg College

John Benjamins Publishing Company
Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/sfsl.71

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2015042239 (PRINT) / 2015046654 (E-BOOK)

ISBN 978 90 272 1581 9 (HB)
ISBN 978 90 272 6749 8 (E-BOOK)

© 2016 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Company · <https://benjamins.com>

Table of Contents

Acknowledgements	IX
CHAPTER 1	
Introduction	1
1.1 Discourse analysis	1
1.2 Genre	3
1.3 The present study	5
CHAPTER 2	
Research questions	13
2.1 Research questions	13
2.2 Grammar and competing motivations	13
2.3 Information flow	15
2.3.1 Activation status	15
2.3.2 Discourse referentiality	19
2.3.3 Identifiability	23
2.3.4 Specificity	27
2.4 Genre differences	28
2.5 Summary of research questions pertinent to information flow	30
2.6 Other factors influencing the linguistic coding of third person	31
2.7 Summary	33
CHAPTER 3	
Data	35
3.1 Data collection	35
3.2 Pear film narratives	36
3.3 Spontaneous conversations	37
3.4 Transcription method	38
CHAPTER 4	
Coding	41
4.1 Syntactic and semantic factors	45
4.1.1 NP realization	45
4.1.2 Modification	46

4.1.3	Syntactic role	47
4.1.4	Semantic class of verb	50
4.1.5	Summary	50
4.2	Factors related to information flow	51
4.2.1	Activation State	51
4.2.2	Specificity	53
4.2.3	Distance from previous mention	55
4.2.4	Realization of previous mention of the referent	56
4.2.5	Discourse referentiality	56
4.2.6	Summary of factors related to information flow	61

CHAPTER 5

Discourse referentiality 63

5.1	Factors related to discourse referentiality	63
5.1.1	Persistence	63
5.1.2	NP realization	66
5.1.3	Modification	71
5.1.4	Specificity	73
5.1.5	Grammatical role	74
5.2	Summary of grammatical manifestations of discourse referentiality	79

CHAPTER 6

Linguistic patterning of referents 83

6.1	Formal characteristics of discourse referents	83
6.1.1	Discourse referentiality and realization	84
6.1.1.1	New referents and realization	84
6.1.1.2	Given referents and realization	87
6.1.2	Discourse referential NPs and modification	101
6.1.2.1	New referents and modification	102
6.1.2.2	Given full NPs and modification	113
6.1.3	Activation status and syntactic role	115
6.1.3.1	New referents and syntactic role	115
6.1.3.2	Given referents and syntactic role	117
6.1.4	Summary of new referents	118
6.1.5	Summary of given referents	119
6.2	Additional considerations for referents	119
6.2.1	New referents and semantic class of verb	120
6.2.2	Other types of “given” referents	123
6.3	Summary	128

CHAPTER 7

Linguistic patterns of non-referential expressions 131

- 7.1 Discourse non-referential forms in core roles 132
- 7.2 Objects of prepositions 137
- 7.3 Predicate nominals 140
- 7.4 Adverbials 142
- 7.5 Vocatives 143
- 7.6 Summary of non-referential forms 144

CHAPTER 8

Genre differences 147

- 8.1 Distribution of referential NPs 147
- 8.2 Preferred argument structure 153
- 8.3 Transitivity across genres 157
- 8.4 Summary 158

CHAPTER 9

Conclusions 163

- 9.1 Summary of major findings 163
 - 9.1.1 Grammar and competing motivations 163
 - 9.1.2 Referentiality 164
 - 9.1.3 Genre 165
- 9.2 Avenues of future research 167

References 169**Appendix** 177**Index** 179

Acknowledgements

Looking back nearly ten years after the initial data collection for this project, I remain very grateful to my Ecuadorian research assistants and their friends and families for so generously sharing their conversations with me. My friend and colleague José Luis Santana and his wife Paola played a crucial role in connecting me to the speech community, and in particular, to Pilar Bastidas, who graciously hosted me for six weeks during the summer of 2006. I've had many friends cheer me on from the sidelines throughout the many stages of the writing process; they include Mary Copple, Stephanie Kerschbaum, Damián Wilson, Eleanor Hogan, my colleagues in the Spanish Department at Gettysburg College, and my mentors Catherine Travis and Rena Torres Cacoullos. This work never could have been accomplished without the support of my family, especially my father, Richard Dumont, who made it possible for me to retreat into writing mode while keeping my household afloat. My son Alex provided me with much joy, laughter, love and affection, which keeps me centered. Lastly, two anonymous reviewers and the editorial board provided very helpful feedback and helped shape this book into the present form.

Introduction

1.1 Discourse analysis

The study of language, or the discipline of linguistics and related fields, has been revolutionized over the course of the last five decades with the advent of voice recorders and other technology that have allowed researchers to capture spontaneous language at the time of production and preserve it in a way in that it can be analyzed at a later date. As such, the field of Discourse Analysis has grown exponentially as researchers have access to phenomena that were once fleeting. Consequently, our understanding of language structure and grammar has been radically altered.

At first glance, to the untrained eye, a transcript of conversational language appears messy and disorganized. Language in its raw and unedited form appears chaotic – full of unfinished sentences, the haphazard stringing together of clauses or fragments of speech with no punctuation, and the frequent occurrence of nuisances or interruptions such as pauses, hesitations, truncated words, laughter, and other speakers. It bears little resemblance to the more polished written genres. However, when our expectations are adjusted and we dismiss the notion that speech is but a poor representation of an underlying grammar, patterns and order become visible. Organization is viewed in the highly sophisticated turn-taking system, which allows for speakers to practically seamlessly transition from one speaker to another with remarkably few problems (cf. Sacks, Schegloff & Jefferson, 1974). Structure is revealed in the architecture of the clause, in which the same patterns are repeated over and over, building conversations and narratives (cf. Chafe, 1994; Du Bois 1987). Punctuation is found in the intonation patterns that regulate the flow of information and contribute to the turn-taking system (cf. Chafe, 1994; Ford & Thompson, 1996).

What, then, are the factors that contribute to the organization and structure of spoken language? A great deal of work has concentrated on the cognitive factors that shape the grammar of spoken language. The work of scholars such as Wallace Chafe (1980, 1987, 1994, *inter alia*), Talmy Givón (1981, 1983, 1995, *inter alia*) and

John Du Bois (1980, 1987, 2003a and 2003b) have shown how cognitive concerns contribute to patterns of language use. Through these and similar studies, we have learned a great deal about how speakers manage the translation of thoughts and memories into coherent speech. In the process of verbalization, speakers are faced with limitations regarding how much information can be in the focus of consciousness at one time (cf. Chafe, 1994), as well as how to keep track of multiple referents in a conversation at a time without creating confusion. Grammar is said to reflect speakers' need to distinguish between what information is presumed to be in the consciousness and what is not.

Discourse analysis has also shed light on categorization. Hopper and Thompson (1984) show that while speakers universally orient toward *things*, which are represented by nouns, and *actions*, represented by verbs, it is only within actual discourse that categories are imposed on the forms. In a similar vein, Hopper and Thompson (1980) and Thompson and Hopper (2001) show that intransitive, transitive and ditransitive are not discrete grammatical categories, but that transitivity is scalar and depends upon several markers of transitivity within the clause. The valency of a verb is determined by its use in discourse, rather than the other way around. The significance of these studies is that language is a dynamic system which is grounded in use. Without discourse, there are no nouns or verbs, nor transitive or intransitive verbs, and referentiality is not inherent to linguistic form. Without discourse, our assumptions about linguistic structure are often misguided. Where the traditional view of language or grammar may be somewhat unidimensional, the analysis of discourse reveals that there are a multitude of interrelated dimensions (cognitive, pragmatic and interactional, among others) which bear relevance on linguistic form.

This study of language in spoken form allows us a privileged glimpse into how social and interactional practices, which are absent from many other genres, influence grammar. In addition to the cognitive pressures associated with creating language on-line with no time to edit, speakers are also simultaneously faced with additional interactional concerns, ranging from taking turns to ways of expressing stance and attitude. These interactional pressures have grammatical consequences that are manifested in a number of ways.

Numerous studies have examined the turn-taking system (Ford & Thompson, 1996; Ford, Fox & Thompson, 2002; Goodwin, 1981; Lerner, 1991; Sacks et al., 1974; Schegloff, 1996; *inter alia*). This research indicates that turn-taking is a highly organized and predictable system within conversation. Participants recognize the appropriate times in a conversation in which to take a new turn at talk, have developed strategies for dealing with problems that arise when turn-taking does not follow as anticipated, and can even collaboratively construct sentences across two speaker turns.

One of the fundamental principles of interactional linguistics is the idea that grammar emerges as a set of patterns that arise in response to repeated actions (cf. Bybee, 2007, 2010). For example, Thompson's (2002) study of complement clauses challenges the traditional view of these clauses as subordinate to complement-taking predicates (e.g., *think*, *know*, *realize*, *wonder*, etc.) by showing that complement-taking predicates are more accurately described as epistemic/evidential/evaluative fragments denoting a speakers' stance toward the clause. Here, our understanding of these structures as main clause + subordinate clause must be abandoned as we see the rich interactional functions that they perform in conversation.

1.2 Genre

It must be noted that discourse analysis is not limited to just the study of conversation, but includes other spoken and written genres of language. It should be emphasized that one genre does not take precedence over another in the field of discourse analysis, rather the focus is on naturally-occurring data and the study of language beyond isolated sentences. This study, building upon previous studies of the cognitive and interactional dimensions of language use, examines two genres of spoken language – spontaneous conversations and monologic narratives of the Pear Film – and finds that genre is an important notion in interpreting differences in frequency and distribution of form between the two datasets.

The shift toward usage based study of language has prompted a closer look at the divergent ways that language is used and has called for a refinement in the way scholars talk about language variation as it relates to different situational or communicative demands. Terms such as genre, register, style and diaphasic variation have all been used in an attempt to capture these differences. The canonical works of Douglas Biber (cf. Biber, 1988, 1995; Biber & Finegan, 1994; Biber, Conrad & Reppen, 1998; Biber & Conrad, 2009) have made great gains toward a deeper theoretical understanding of nuances of this type of variation, yet it can be argued that many linguistic subfields lack an awareness of the importance of considering contextual variation and the necessary methodological tools for doing so.

Variation in distribution of linguistic forms between genres and styles was noted even in the earliest usage based studies. Early sociolinguistic studies emphasized the style as a linguistic variable (cf. Labov, 1972; Trudgill, 1974), but as later studies have alluded to (cf. Biber & Finegan, 1994), when the variation is only understood in a limited range of styles, the bigger picture of how different types of situational variation relate to linguistic form fails to develop. As corpus linguistics has evolved, the notion of contextual or situational variation has been explored

in a number of other types of studies. A few of these are discussed in more detail here, and important connections to the present study are highlighted, but a fuller discussion of the complexities of register, genre, style and the like are better found in the aforementioned works of Biber.

Biber (1988) dedicates an entire volume to genre effects on a wide range of linguistic phenomena, from tense and aspect to negation, and this fine grained focus on the form-genre connection is further explored in the present study. As corpus linguistics has expanded, these fundamental concepts have been subjected to empirical testing in an increasing number of languages, different theoretical frameworks and different practical models are emerging. One influential example of this extension to non-English languages and the incorporation of a more sophisticated form of modeling is found in Biber et al. (2006), which reports the findings of a Multi-Dimensional analysis of register variation in Spanish that describes six different dimensions of variation (that correspond to different registers) that can be identified by the co-occurrence of linguistic features (e.g., the subjunctive mood, progressive aspect, present tense, etc.). In a sociolinguistic account of variation, Travis (2007) explores genre effects and subject expression in two varieties of spoken Spanish, finding genre to be a significant factor in the rate of first person subject expression in spoken Spanish, as well as in the duration of the priming effect. A greater understanding of how this dimension of variability is related to linguistic form is essential; in particular a greater understanding of how to differentiate disparate rates of occurrence of linguistic forms between genres and true differences in linguistic conditioning between genres is an important development in discourse analysis (and related fields of study), and has significant implications for the advancement of linguistic study. As more and more researchers conduct quantitative research of language and comparisons between studies are drawn, it is essential that we understand how the external or situational circumstances may shape the linguistic patterns uncovered in one study.

Having highlighted the importance of recognizing register differences or genre effects and the need to produce a coherent and consistent understanding of this dimension of variation, this study focuses on the functional basis for differences in linguistic patterning that emerge in two different genres of spoken Ecuadorian Spanish. The word genre is used throughout this work in the following manner, following Bauman (1999, p.84): a genre is a "speech style oriented to the production and reception of a particular kind of text". Note that genres can be defined broadly, as in conversation versus narrative, or on a more specific level, as in the comparison of spontaneous service encounters (as a type of conversation) and oral narratives of personal experience (one type of narrative), or even more specifically within those levels. The genres that provide the data for the analyses presented in this work are as follows: (1) spontaneous conversations between friends and

family members, and (2) Pear Film narratives (cf. Chafe 1980). The notion of genre is invoked in order to provide functionally based interpretations of the data within the specific situational circumstances in which they were produced; and the analyses within this work should be understood with respect to these particular datasets and not understood as entirely representative of the larger genres to which they belong. That is, these conversations are not intended to be representative of all conversations between Spanish-speakers or even all Ecuadorians, nor are these narratives representative of all oral narratives (and indeed this particular narrative genre is unique in that it was designed by linguists for the explicit purpose of obtaining data for linguistic analysis). The analysis centers on the communicative actions common to the genre, rather than the genre itself. It should be noted that most of the communicative actions are not exclusive to one genre or the other (for example, introducing referents and narrating events occur in both genres), although a few communicative actions (such as turn-taking) are exclusive to the conversations. More specific information about the speakers, the genres and how these data were collected is given in Chapter 3.

1.3 The present study

This tome is a case study of the linguistic forms used to designate third person as they occur in two genres of spoken Ecuadorian Spanish. Why was third person chosen? The sheer abundance of third person references in any given corpus immediately makes it an attractive candidate for a quantitative study. It is through the quantitative study of language that we achieve an understanding the routinized linguistic structures that are the grammatical realization of recurrent social and communicative needs. In addition, the study of third person expressions provides a veritable gold mine of information about the underlying cognitive processes involved in language production – it is through the study of the third person that we trace the information status of referents, study anaphora, and measure the ways in which information flow pressures shape discourse. The frequency of third person expressions also guarantees that they occur in a wide variety of interactional contexts, allowing for an understanding of how interactional factors help shape grammar.

Only those linguistic expressions used to designate third person are examined here. One reason for this is purely practical – 1st and 2nd person references are nearly absent from one of the genres under consideration (narratives of the Pear Film). It is also because third person has been less widely studied than first and second person in Spanish (e.g., Cameron, 1994; Flores-Ferrán, 2002; Torres Cacoullos & Travis, 2015; Travis, 2005, 2007 *inter alia*; but see Bentivoglio, 1993

and Dumont, 2006) and because there are different factors that affect the linguistic coding of first and second person references that do not apply to third person. Lastly, there is simply more variation in both the information flow parameters (discourse referentiality, information status, specificity, identifiability) for third person than first or second person, and a wider range of linguistic expressions used to designate third person. It is precisely this variation which is central to the investigation of the cognitive processes involved in language production examined here.

How do we investigate the role of third person references in the cognitive and interactional processes of language production? Or conversely, how do we investigate the role of cognitive and interactional pressures in shaping the grammar of third person references in spoken language? The best approach toward beginning to answer these questions is to first sort out the cognitive factors from the interactional pressures. For that reason, two genres of spoken language – monologic narratives of the Pear Film and dialogic conversations – were chosen for this study. These genres differ from each other in two important ways – the information flow pressures and the level of interaction. In terms of cognitive processes, these narratives have higher information flow pressures (cf. Du Bois, 1987), which allows us to understand how speakers maximize the available grammatical resources of a language to keep track of a relatively high number of referents. The low information pressures of the conversations allow for a comparison of the form-function links as evidenced in references to the third person under different conditions. As far as interactional concerns, the conversational data teems with interaction as speakers take turns, compete for the floor, finish either other's sentences, question what others have said, and jointly build narratives. Interaction is not absent from the narratives (cf. Schegloff, 1982 and Goodwin, 2007), but we can expect interactional concerns to be less pervasive in narratives than in the conversations and there may also be fewer types of interactional concerns (i.e., turn-taking is less important in a largely monologic narrative). Compare examples (1) and (2), for example. In the first example, two speakers are jointly constructing a narrative, taking turns with each other, competing for the floor (overlapping speech is seen in square brackets, see the Appendix for a full list of transcription conventions), agreeing and disagreeing with each other. In the second example, the speaker is recalling the story on her own – she has the floor to herself, there is no one to help her remember details or to challenge her memory, nor to agree with her.

(1) Jointly constructed narrative in conversational data

- A: ... (H) *vos te fuiste una vez trotando con nosotros?*
 R: .. *hasta el --*
 claro,

- pues,
hasta= --
y después [Ø cogimos bus].
- A: [que fue] --
no,
y esa camioneta,
que nos llevó hasta la Mitad del Mundo,
- R: ah,
claro,
para llegar hasta la Mitad del Mundo.
- A: .. (H) y que luego nos Ø querían llevar <@ más allá @>.
.. y Ø estaban borrachos,
Ø creo.
- R: claro,
se iban para Calacalí,
creo.
- A: .. ah,
a la costa,
a dónde también Ø se irían?
- R: .. pero Ø estaban --
bien borrachos,
nosotros golpeábamos el [vidrio],
- A: [hm],
- R: .. para que Ø nos pare,
- A: y Ø no nos pa- --
Ø nos paró más allá del redondel?
- R: ... <@ claro @>.
- A: una media [cuadra].
- R: [y de ahí] Ø queríamos subir trotando de nuevo.
- A: ... no,
Ø cogimos bus hasta.. <@ Pomasqui @>.
- R: Ø cogimos bus y Ø nos bajamos en --
claro,
[en Pomasqui].
- A: [Pomasqui].
y de ahí Ø seguimos trotando.
- A: \... (H) did you go jogging with us one time?
- R: .. to the --
of course,
well,
to= --
and then [we caught a bus].

A: [that went] --
no,
and that pickup truck,
that took us to La Mitad del Mundo,
R: ah,
of course,
to get to La Mitad del Mundo.
A: .. (H) and then later (they) wanted to take us
<@ farther @>.
.. and (they) were drunk,
(I) think.
R: of course,
they were going to Calacalí,
I think.
A: .. ah,
to the coast,
where would (they) go?
R: .. but (they) were --
really drunk,
we hit the [window],
A: [hm],
R: .. so that (they) would stop,
A: and (they) didn't st- --
(he) let us out past the roundabout?
R: ... <@ of course @>.
A: a half a [block].
R: [and from there] we all wanted to go
jogging again.
A: ... no,
(we) took a bus to.. <@ Pomasqui @>.
R: (we) took a bus and (we) got off in --
right,
[in Pomasqui].
A: [Pomasqui].
and from there (we) kept jogging.' (Fumar:155-194)

(2) Excerpt from monologic narrative

Ø está yendo en la bicicleta,
eh,
.. en el que misma --
en el mismo camino,
pero en el sentido contrario,
pasa una niña,
en otra bicicleta.
.. (H) a lo que Ø están pasando juntos,

eh,
 .. el sombrero del niño vuela,
 (H) y el niño,
 por regresar a ver el sombrero,
 no ve una piedra grande y se choca.
 ... (H) eh,
 Ø se choca y se caen las manzanas,
 se cae el canasto,
 se riegan todas las peras,
 (H) la niña sigue su camino,
 (H) y el niño a lo que= --
 .. a lo que Ø esté en el piso,
 se levanta su pantalón,
 se baja sus medias,
 se X su pierna lastimada,
 y se da cuenta que a lado de él están parados tres niños.
 '(he) is going on the bicycle,
 eh,
 .. in the same --
 on the same road,
 but on the other side,
 comes a girl,
 on another bicycle.
 .. (H) when (they) pass by each other,
 eh,
 .. the boy's hat flies off,
 (H) and the boy,
 upon looking back at the hat,
 doesn't see a big rock and crashes.
 ... (H) eh,
 (he) crashes and the apples fall,
 the basket falls,
 the pears all spill out,
 (H) the girl goes on her way,
 (H) and the boy when --
 .. when (he) is on the ground,
 raises his pants,
 lowers his socks,
 X his hurt leg,
 and realizes that standing next to him are three boys.'
 (PS 100:131-154)

The corpus used for this study is described in detail in Chapter 3. The data are from a larger, three genre corpus of Ecuadorian Spanish. Fifteen narratives of the

Pear Film were selected, totaling 6430 words, from which all linguistic expressions used to designate 3rd person were extracted (exclusions are discussed in Chapter 4). An equal number of speakers were chosen for the conversational data. A sample of third person references from these speakers was extracted from transcripts of ten conversations, totaling 74,673 words.¹

The findings presented in this work shed light on a number of important issues. The comparison of two genres allows for an empirically based understanding of what recurrent grammatical forms are linked to the different actions of narrative and conversational data, and to better interpret the subsequent patterns by observing links between form and function. For example, the higher information flow pressures (communicative action or goal) in the narrative data are reflected in the forms used in this genre, notably a higher rate of transitive constructions, proportion of referential to non-referential mentions of third person, and a stricter adherence to Preferred Argument Structure (PAS, cf. Du Bois, 1987). In other words, the collective patterning of these forms together in the narratives but not the conversation is interpreted as intrinsically linked to the unique communicative goals of the Pear Film narratives. Other grammatical forms are seen at different rates in the conversations than in the narratives and are linked to different communicative actions, such as the use of the definite article to introduce of new referents with definite markers, which is interpreted to reflect genre-specific discourse actions. In this case, different coding reflects the more frequent action of introducing referents that are assumed to be shared between speakers in the conversations than in these Pear Story narratives.

Categoriality is explored, especially with respect to the distinction between referential and non-referential forms, and the typology of the different groups of non-referential forms. The terms referential and non-referential are used here in the sense of Du Bois (1980) and Hopper and Thompson's 1984 "discourse manipulable" sense. Referential expressions are those that speakers use to track a referent

1. Although at first blush it looks as though there is a large discrepancy in the proportion of the corpus that belongs to the narratives and the proportion that belongs to the conversations, it was necessary to take a sample from a larger set of conversational data for several reasons. The narratives contain almost exclusively 3rd person, whereas 3rd person in the conversations is interspersed with first and second person, meaning that one needs to have a larger corpus to get the same number of tokens. Secondly, given the wider range of communicative actions in the conversations, it was decided to extract more tokens from this genre to ensure a better representation of the discourse functions of third person expressions. Lastly, only a sample of NPs was taken from the 74,673 words in the conversations, so the actual word count of the proportion used is much smaller (but very difficult to count precisely, as often only one speaker's NPs were used, but the speech of others intervenes between turns of the target speaker).

whereas the different kinds of non-referential expressions perform different discourse functions, including predicating, classifying and characterizing. Compare the noun *sol* ('sun') in (3) and (4). It is the same lexical form, yet the discourse function, and consequently the grammar, is very different between the two examples. In the first example, *sol* is referential, that is, it exhibits the prototypical functions of a noun – it refers to an entity that has continuity of identity within the discourse. In addition, it shows nominal morphology (i.e., the definite article), whereas non-referential expressions tend to exhibit less morphology characteristic of NPs (e.g., bare NPs, as in (4)). The difference between the linguistic expressions referentiality and non-referentiality will be discussed in depth in Chapter 5, where the operationalization of the coding for referential and non-referential expressions is explained, but perhaps one of the easiest ways to understand the difference between the two is the often used analogy of computer files (cf. Du Bois 1980, p. 220–23, 1987 p. 817). When a speaker mentions a referent for the first time in a conversation, the listener creates a new file or accesses an existing cognitive file for that referent. The identity of this referent is stable, information about this particular referent is stored in the mind, and new information can be added to the file. Linguistic expressions of non-referentiality, on the other hand, such as the noun *sol* in (4), have no associated cognitive file. The noun *sun* in this example is not used to talk about the entity that is yellow, is a star, and is located approximately 149.6 million kilometers from the Earth. The discourse function of *sol* in this second example is part of a verbal predicate and refers to the weather conditions. As we will see throughout this volume, there is ample evidence in the discourse that the grammar of linguistic forms used to designate third person reflects the differences between referentiality and non-referentiality and that speakers use them for quite distinct discourse functions.

(3) Referential use of the noun *sol*

F: *amarillo es el sol,*

F: 'the sun is yellow,'

(Birthday: 286)

(4) Non-referential use of the noun *sol*

A: *pero está haciendo basta=nte so=l.*

A: 'but it's really sunny' (lit. '(it) is making a lot of sun')

(Classes: 7)

The previously understudied roles of free NPs (see *gastroenteritis* and *hepatitis* in (5)) are also explored in this work. We see that speakers use these NPs for a variety of functions related to both information flow and interactional concerns.

(5) Free NP

E: *qué problema es el que da,
cuando está --*

.. cuando Ø hacen cosas sucias?
...(2.0) cómo se llama?
L: .. **gastroenteritis**.
E: .. **gastroenteritis**.
o la otra,
que te pones amarillo?
L: .. **hepatitis**.
E: .. **hepatitis**.
E: 'what problem is it,
when it --
.. when (they) do dirty things?
...(2.0) what's it called?
L: .. **gastroenteritis**.
E: .. **gastroenteritis**.
or the other,
when you turn yellow?
L: .. **hepatitis**.
E: .. **hepatitis**.' (Food, 6:13–622)

As we trace the form and distribution of references to third person throughout discourse, we see how the varying forms and roles reflect local cognitive and interactional demands. At the local level, we can situate a reference within the immediately surrounding discourse and understand why it is that a speaker has chosen a particular way of encoding the reference. On a more global scale, clusters of similar form-function links within genres and the comparison of these clusters across genres reveal patterns that are in turn interpreted as evidence of the unique of communicative actions and goals of each genre, and the value of a discourse analytic approach to the study of language. The findings here attest to the view that spoken language is in fact orderly and structured and that speakers are highly capable of managing several dimensions of external circumstances (i.e., turn-taking, information flow) in the on-line production of language.

Research questions

2.1 Research questions

A great deal of research has focused on how information flow correlates with linguistic form, and there has recently been an interest in how interactional factors may also relate to third person form and distribution. What is lacking is a methodologically refined study of information flow, particularly with respect to forms in Spanish, and a holistic approach to the form and distribution of linguistic expressions of third person which takes into account factors beyond information flow. The current study enhances our understanding of these linguistic forms and distribution by (1) addressing methodological concerns, (2) contributing to our understanding of the effects of both cognitive and interactional factors on NP form, and (3) comparing data from two genres in order to better understand genre effects.

2.2 Grammar and competing motivations

When engaging in conversation or story-telling, speakers need to use language in a way that effectively, accurately and efficiently reflects the message(s) they are trying to convey. From a functional perspective, linguistic forms are a resource that speakers employ to negotiate the different needs presented in the discourse context. Grammar, or a description of the way groups of speakers use these forms, arises from repeated patterns of use (cf. Hopper, 1987; Bybee & Hopper, 2001) and is said to “code best what speakers do most” (Du Bois, 1985 p. 363).

To the non-linguist and to many linguists, grammar is often understood to be a rigid set of rules to which speakers conform. To the functional or socio-linguist, grammar is adaptable, flexible, and speakers are presented with choices in constructing grammatical units, i.e., a clause or a phrase. However, these choices are not entirely up to the whim of the speaker, rather they are constrained by a number of different factors. The types of factors of interest here are cognitive factors

that influence the flow of information in discourse and the social and interactional factors that shape conversation.

Du Bois (1985) points out that discourse is shaped by competing motivations. Unmarked grammar, he says “is distributed on the basis of “unmarked” (statistically most frequent) discourse” (Du Bois 1985, p. 362). With this in mind, we expect to see broad patterns of use emerging, repeated across speakers and genres. This is the case with information flow – speakers are constantly managing the flow of information, and as a result, the link between forms used to regulate information flow and rate of occurrence of these forms in different communicative contexts lead to patterns related to this discourse goal. For example, Preferred Argument Structure (cf. Du Bois, 1987) has been observed in a number of typologically unrelated languages. The two key findings are that there tends to be only one new argument per clause and that new arguments are typically found as subjects of intransitive verbs or objects. Clauses that are consistent with the pragmatic and morphosyntactic tendencies of PAS can be considered unmarked. Marked grammar, on the other hand, may be evidenced in situations in which another factor, for example a particular interactional need such as disagreement, outweighs the cognitive motivation behind the expected pattern, for example, a deviation from PAS. There may be instances in which there is no clash between the competing motivations, that is, the grammar of the language allows for one grammatical coding that accommodates both factors. On the other hand, there may be times when the grammar of the language does not allow for both factors to be accommodated, therefore, speakers must “mark” the grammar according to the strength of each external motivation.

Regarding the grammatical correlates of the third person in Spanish, speakers are faced with a number of choices. For third person referents, they choose among three types of realization (full NP, pronoun and unexpressed), and the choice is commonly thought to be governed by information flow, pragmatic and discourse-sensitive considerations (Blackwell, 2003; Silva-Corvalán, 1994).¹ In addition, a choice must be made about how to fit the referent into the rest of the discourse (i.e., syntactic role), and when a lexical form (NP) is used, they must choose how to modify the noun (i.e., choice of determiner). Information flow is one of the most influential forces in helping speakers to make these choices, and the effects of information flow have been widely observed in a variety of languages, with speakers roughly adhering to universal tendencies such as presenting only one new piece of information at a time. Yet speakers are also often dealing with other types

1. One of the basic assumptions here is that discourse shapes grammar. For alternate (formal) approaches to subject expression in Spanish, see Toribio 1993, 1996 and 2000.

of pressures, sometimes of a more interactional or social nature, or genre specific needs, at the same time, which can also influence the linguistic forms they use for third person references. The purpose of this study is to examine the linguistic expressions used to designate third person in Spanish in different communicative contexts and to look for evidence of patterns that reflect the different types of pressures that contribute to the choice of a particular third person expression in spoken discourse.

2.3 Information flow

Information flow involves the relationship between the activation status, identifiability, specificity and referentiality and the bearing that these factors have upon the linguistic coding. When participating in a conversation or other speech event, speakers and interlocutors are faced with the task of keeping track of referents that are being talked about and distinguishing those from forms that serve other discourse functions (e.g., orienting, predicating, vocatives, etc.) and require a different type of cognitive effort. When coding third person, speakers use a combination of linguistic devices to convey to the hearer important information such as whether or not the expression is a referent to be tracked, along with clues that enable them to identify the referent or create a new cognitive file. Among the devices that speakers use are a combination of morphological markers (e.g., definite versus indefinite articles), realization (e.g., full NP or pronoun), and syntactic role.

The different information flow properties of third person interact with different areas of third person grammar – specifically the corresponding form (realization), morphosyntactic markers (such as articles or relative clauses) and syntactic role. In the sections that follow I provide explanations of the different dimensions of information flow, show how they are interrelated, and how they have been found to correlate third person forms. In addition, I draw attention to research questions that have not yet been addressed.

2.3.1 Activation status

Many researchers have highlighted the role of activation status in the choice of form used to designate third person referents and distribution of these forms (cf. Ariel, 1990; Du Bois, 1987; Givón, 1983; Gundel, Hedberg & Zacharski, 1993; Lambrecht, 1994; *inter alia*), and as this research has emerged, advances have been made toward defining activation state and operationalizing methods of data analysis according to activation state. As Chafe (1976) explains it, activation status has

to do with where in the hearer's consciousness a speaker assumes a referent to be located. A concept can be considered active (given) if it is "knowledge which the speaker assumes to be in the consciousness of the addressee at the time of the utterance" (Chafe, 1976, p. 30). Conversely, it is inactive if it is not in the consciousness of the addressee (or, more specifically, not assumed so by the speaker). For a concept to be inactive, it does not mean that the addressee has no prior information about the referent, but that the speaker assumes that the referent is not in the addressee's focus of consciousness at that particular moment in time. However, Chafe also points out that this characterization is based upon introspection, and it is irresponsible for the linguist to assume that he or she has access to what the speaker is thinking.

While a speaker's use of particular forms does not in and of itself point directly to that speaker's beliefs about hearers' consciousness, linguists have nevertheless developed a number of hypotheses regarding how various cues from the discourse, such as other mentions of the referent and the forms which the speaker uses to code the referent, can shed light on whether or not speakers invoke notions of activation state in the linguistic coding of a referent. Many researchers (see, for example, Ariel, 1990; Chafe, 1987; Givón, 1983; Prince, 1981, 1992) agree that if a referent has been mentioned in the recent discourse, then it is probable that the speaker and hearer will regard it as active or given, but is debatable how long a referent stays active after it has been activated. Chafe (1987) suggests that active concepts are found in short-term memory, and given that short-term memory can only hold seven plus or minus two items at a time (Miller, 1956) the number of active concepts at a time must be very limited. Givón (1983) first proposed a measure of twenty clauses for operationalizing referential distance, and later modified the number to ten (1995). As Fox (1987) observes, however, this linear view of discourse and anaphora does not always represent what speakers are *doing* in the discourse, and she shows in her research how the notion of conversational sequences and other discourse structures are relevant to studies of anaphoric coding of referents. Tomlin (1987) also finds that discourse structure is more relevant than elapsed time (as measured in clauses in Givón's work); specifically he reports that episode boundaries are crucial in the realization of referents in narrative data.

While the precise amount of time that a referent remains active in the mind and the role of other factors is unknown, most would agree that activated referents can and probably do change status after a relatively short period of inactivity. This leads to the question of what happens to referents after they have been activated. Prince's 1981 model and Chafe's 1987 model with three states are two of the most often cited models. Prince's model is more detailed than Chafe's model, with seven different categories of assumed familiarity to his three activation statuses, but both

models share several core traits. In both models there are “new” and “given” referents, with (an) intermediate state(s).

Chafe (1987) argues that active referents are moved to an intermediate activation state, semi-active (“accessible”). Semi-active referents are not limited to those which have been previously active, and Chafe (1987) considers those referents that are part of a greater schema to also be semi-active, i.e., “they are more accessible to recall than they would have been as inactive concepts” (Chafe, 1987, p. 29). For example, if the schema “tree” is invoked, then it may follow that concepts such as branch, leaf, and roots are more accessible or easily identified due to their association with “tree”. Note that Givón’s (1983) model does not include a semi-active or accessible state.

The third activation status distinguished by Chafe is inactive, or new. This status applies to referents that have not been active or semi-active via previous mention, nor can they be considered semi-active due to association with a schema. According to Chafe (1987, p. 31), the change in state from inactive to active carries the heaviest cognitive burden; the cost of activating these referents is much higher than other changes in activation status (i.e., activating a semi-active concept, the change from active to semi-active). As a result, it is expected that this higher cost will be reflected linguistically.

Prince’s “brand-new” (both unanchored and anchored) corresponds most closely to Chafe’s “inactive” or “new”. New-unused, noncontaining inferrable and containing inferrable are all similar to Chafe’s “semi-active” or “accessible”, and her “textually and situationally evoked” referents could correspond to Chafe’s semi-active or active. Prince’s later (1992) work reflects the important distinction between *hearer* new/old and *discourse* new/old.²

Example (6) illustrates several changes in activation state and the anaphoric coding predicted by models of information flow. In line 2, the boy is first introduced, representing a change from inactive to active. The cognitive burden is represented in its expression as a full NP, *un niño*. The referent then continues as an active concept, seen in lines 5, 7 and 8, which show lesser coding (pronoun in line 5, unexpressed in lines 7 and 8, and again pronominally in 10, 11 and 12). With the introduction of the girl in line 15 the boy becomes semi-active, and when he becomes active again he is referred to pronominally (line 16), which is more coding than the previous unexpressed mentions in lines 7 and 8, but less than the first mention and activation (line 2).

2. Other researchers may have different interpretations of how Prince’s categories align with Chafe’s. O’Dowd (1990), for example, codes Prince’s Unused referents as New, and only inferrable and containing inferrable as Accessible.

(6) Changes in activation state

- 1 *el momento que él estaba distraído recogiendo,*
 2 *pasó **un niño**.*
 3 ... (H) con --
 4 *en una bicicleta.*
 5 .. **é=1** se acer- --
 6 *se acercó,*
 7 (H) y $\emptyset$ quiso coger --
 8 .. $\emptyset$ quiso coger una,
 9 *pero=,*
 10 *la tentación **le** --*
 11 *.. **le** --*
 12 *.. (H) **le** venció y $\emptyset$ cogió todo un canasto en su*
 13 *bicicleta.³*
 14 (H) *al seguir por el camino,*
 15 *eh,*
 16 *pasó una niña,*
 17 *y $\emptyset$ **le** botó el sombrero.*
 18 (H) *al momento de botarle el sombrero **él** se --*
 19 ***é=1**.. se descuidó y cayó.*
- 1 'the moment that he was distracted collecting (pears),
 2 along came **a boy**.
 3 ... (H) with --
 4 on a bicycle.
 5 .. **he** appr- --
 6 approached,
 7 (H) and (**he**) wanted to take --
 8 .. (**he**) wanted to take one,
 9 bu=t,
 10 the temptation to **him** --
 11 .. to **him** --
 12 .. (H) conquered **him** and (**he**) took a whole basket on his
 13 bike..
 14 (H) upon following the path,
 15 eh,
 16 along came a girl,
 17 and knocked off **his** hat (lit. knocked off the hat to **him**).
 18 (H) with the hat being knocked off **he** --
 19 **he=**.. got distracted and fell.'

(PS 29:27-44)

3. Although the boy is not the referent of this NP, he is referred to by the use of the possessive adjective (i.e., the possessor) and this also contributes to the maintenance of the activation state.

It should be noted that Chafe (1987, p. 25) applies different information states to *concepts*, which includes “the ideas of objects, events, and properties”, which are expressed as verbs, adjectives, etc. In practice, however, most researchers find it practical to only code for the information status of referents. As Du Bois (1987, pp. 816–17) points out, “it is easier to decide whether a nominal reference to an entity is new or not, or whether there are one or two entities within a single clause, than it is to decide whether a particular sequence of verb + adverb + complement constitutes three two, or just one (complex) piece(s) of new information”. Furthermore, he states that “it may be that speakers attend to new nominal information on a separate cognitive track, as it were, from new verbal information” (1987, p. 817).

While there is a consensus that given-new distinction is not simply binary, but rather scalar or multi-dimensional, and that different factors (previous mention, shared knowledge, participation in a schema, episodic structure, etc.) can influence the information status of a referent, in practice most researchers find it sufficient to distinguish between new and non-new (given), although they arrive at these distinctions in different ways. Lambrecht (1994, p. 100) argues that “from a strictly grammatical (phonological or morphological) point of view, only a binary distinction is justified”. Although it is possible to examine a text or transcript and find clear examples of each of Chafe and Prince’s categories, when the activation status of each and every third person referent in a dataset must be accounted for, researchers are faced with the task of operationalizing these notions. A variety of different practices for operationalizing this variable have been used, with both Chafe and Prince’s models often cited. The only criterion that can be neatly operationalized for 3rd person referents is previous mention (see Ewing, 2005, pp. 122–124 for further discussion), and it is this measure that will also be used in this study. However, alongside this I consider other dimensions, including distance from previous mention, and at times, intervening referents.

This study seeks to add to the body of literature that explores the relationship between how speakers code third person referents and information status. Specifically, this study will examine the preferred form and distribution of both given and new referents with special attention to preferred presentational structures across genres. It also considers other dimensions of information flow in conjunction with information status and fills a gap in previous research by showing how different dimensions of information flow are intertwined with each other and have bearing on the linguistic coding of third person.

2.3.2 Discourse referentiality

It is clear that not all linguistic expressions of third person have the same discourse function and research in this domain has traditionally considered the NP to be the principal data point for information about third person. Prince (1981, p. 235)

states this variation in function very succinctly: “all discourse entities in a discourse-model are represented by NPs in a text, though not all NPs in a text represent discourse entities”. Hopper and Thompson (1984, p. 710) explain that the prototypical function of nouns is to “introduce participants and “props” and to deploy them”. NPs performing these functions are referential, in the sense that they “have continuity of identity” (Du Bois 1980, p. 208), and “are important in the subsequent discourse” (Hopper & Thompson, 1984, p. 711). In (7), the referent Janine’s cousin (and subsequent mentions) is clearly referential. In the first mention, a participant (*la prima de la Janine*) is introduced, and is later seen to have both continuity of identity and to be important in the discourse.

(7) NP performing a referential function

- A: .. (H) *y te acuerdas que una vez **la prima de la Janine**,*
<@ se pegó una matada @>?
 .. (H) *jugando a las escondidas?*
... en la casa mismo de la Janine,
afuera que [había tierra]?
- R: [cuál prima]?
- A: **una prima de la Janine,**
que tenía.
- R: **la Mónica?**
- A: *ah=.*
.. ∅ iba corriendo y shu=,
∅ se resbala,
y ∅ se pega una matada,
y ∅ se raspó la rodilla,
te acuerdas?
- A: ‘.. (H) and do you remember one time that **Janine’s cousin**,
<@ took a fall @>?
 .. (H) playing hide-and-go-seek?
 ... at Janine’s house,
 outside where there [was dirt]?
- R: [which cousin]?
- A: **one of Janine’s cousins,**
that (she) had.
- R: **Mónica?**
- A: *yeah=.*
.. (she) was running and shu=,
(she) slips,
and (she) takes a fall,
and (she) scrapes her knee,
remember?’

(Fumar:218–232)

Du Bois (1980, p. 210) noted that “the form of a nonreferential mention is not responsive to the presence or absence of a prior mention”, and as this study will show, it is important to exclude non-referential forms (NPs) from studies of how grammar is shaped by information status. For example, the form *escalera* (‘ladder’) in (8) below is used non-referentially (a detailed discussion of the different categories of discourse referentiality and their grammatical manifestations is presented in Chapter 5), and as such, is not responsive to the effects of information status, as seen in the coding as a full NP in three mentions over four clauses (and is more of a “prop” than a central figure to the action), as opposed to the subject (clearly an “actor”), which has been established and receives no overt linguistic coding here.⁴

(8) Non-referential use of a NP

ø se vuelve a subir la escalera,
ø se --
ø se limpia su ropa,
y se ø vuelve a subir a la escalera.
 .. (H) *ø se sube a la escalera,*
 ‘(he) goes back up the ladder,
 (he) --
 (he) cleans his clothes,
 and (he) goes back up the ladder.
 .. (H) he goes up the ladder,’ (PS100:59–63)

In addition, several scholars have noted that non-referential forms exhibit different morphosyntactic behavior than referential expressions. Hopper and Thompson (1984) show that referential uses of NPs take different forms than non-referential uses of NPs in that they show the more overt morphological marking when used to introduce referents, whereas NPs used non-referentially lack some or all of the morphological marking associated with nouns. Thompson (1997) finds that core arguments tend to be tracking (i.e., referential), whereas obliques tend to be non-tracking (i.e., non-referential). Torres Cacoulllos and Aaron (2003) find that a significant amount of bare, lone English origin nouns in New Mexican Spanish tend to appear in non-referential expressions, but they do not show that there is a general tendency for nouns used non-referentially to be bare. Du Bois (1980, p. 212) and Ono and Thompson (1994) also note the correlation between non-referentiality and lack of determiner in English.

While the difference between forms used to code referential and non-referential expressions of third person has been well-noted in the literature

4. Note here the distinction between referential and referring. Most linguists would agree that the NP *escalera* in example (8) is a referring NP, but the important distinction in this work is that it is non-referential in Du Bois’ (1980) sense.

(cf. Du Bois, 1987; Givón, 1981; Hopper & Thompson, 1984; Prince, 1981), the importance of referentiality is often overlooked in studies of third person and it is not apparent that researchers distinguish between the two discourse functions of NPs when conducting quantitative research on NPs, particularly where information flow is concerned.

A survey of ten studies that examine the relationship between information status, NP form and distribution in several languages (Ashby, 1995; Ashby & Bentivoglio, 1993, 1997; Bentivoglio, 1993; Du Bois, 1987; Durie, 2003; Ewing, 2005; Kärkkäinen, 1996; Matsumoto, 2003; Thompson, 1997) reveals that only one researcher (Durie, 2003) states that he excluded non-referential NPs from his data. While some take referentiality into account to some degree, many appear to disregard it. Ashby & Bentivoglio (1997) only include full, specific NPs in the S (subject of intransitive verb) and O (object) syntactic roles in their study; as will be seen in Chapter 5, these criteria correspond with discourse referentiality for Spanish. In her study of the core-oblique distinction in English, Thompson (1997) excludes Predicating NPs, for example *ornaments* in the phrase *put up ornaments*, on the grounds that they are non-referential in function.⁵ She does, however, include expressions in oblique roles and codes those for information status. It should also be noted that although Durie (2003) does mention that non-referring nominals were excluded from his study of Acehnese, and he discusses the differing discourse roles of non-referring NPs, he does not clearly explain how he distinguished between referring and non-referring NPs in selecting his data, and says (2003, p. 193n10) that “the typology of non-referring expressions is too complex to explore here. It remains an important topic for research at the grammar-discourse interface”.⁶

Recall from the introduction that the analogy of a referent to a computer file is often invoked (cf. Du Bois, 1980, pp. 220–23, 1987, p. 817) to explain referentiality. The premise is that there are two main functions of third person (nominal) expressions. On the one hand, *referents* can be understood as having cognitive files – when a new referent is introduced (typically with a NP), a new file is opened, and as that referent is tracked throughout the discourse, new information is added to that file. On the other hand, *non-referential expressions* do not have an associated file, and as they do not have continuity of identity, they are not tracked in subsequent discourse. As Du Bois (1987) and Croft (1991) have claimed, and as I will

5. Following Dixon (1979), grammatical role A refers to the subject of a transitive clause, and S is the subject of an intransitive clause. O is used for Object.

6. The terms referential, tracking, and referring have all been used in the literature in different ways. It is my understanding that the NPs that Durie calls non-referring are the same as the NPs that are labeled non-referential in this work.

argue here, the evidence suggests that speakers attend to new nominal expressions of referents on a separate cognitive track than new (in the sense of not previously mentioned), non-referential expressions. Importantly, unlike activation status (e.g., given and new), where researchers are often very clear in defining how they code these variables, a replicable way of operationalizing the coding for referential versus non-referential expressions has not been proposed. Even when researchers do distinguish between the two types of expressions (e.g., Durie, 2003), they are often vague in describing how they code for referentiality. What is lacking, then, in the research is (1) a replicable and operationalized way to code for referentiality, and (2) research which investigates the linguistic variation involved in the differential coding of referential and non-referential expressions. This investigation offers an operationalization of referentiality based on morphosyntactic and semantic properties identified through empirical analysis, as well as an investigation of other form-function links between expressions used referentially and non-referentially.

2.3.3 Identifiability

Identifiability, or whether or not the speaker believes that the hearer will be able to identify the referent, is often linked to information flow. On the one hand, it seems very simple: given referents within a discourse should all be identifiable, and among new referents, some will be identifiable and others will be unidentifiable (cf. Thompson, 1997, p. 68). However, it is not always the case that once a referent is introduced into the discourse it becomes identifiable to the speaker. In example (9), we can see how identifiability is negotiated between the speakers. After the first mention of *tu pelada* ('your girlfriend') by speaker R, speaker J asks for clarification. At the same time, speaker R has realized that speaker J is not making the connection, so he gives more information to help J identify the referent, he says *la recepcionista* ('the receptionist'). It is at this point that the identity of this referent, which was first introduced as definite, is established. And J goes on to explain that this person is not, in fact, his girlfriend, she is his employee, and therefore the first mention of this person cannot truly refer to her. As was discussed above, the notion of "givenness" is complicated at best, and as this example shows, identifiability is even more difficult to define in a manner which the linguist can use to accurately identify and code referents for identifiability.

(9) Negotiating identifiability

- R: *y tu pelada?*
 ... [*la recepcionista*]?
 J: [*cuál pelada*]?
 R: <@ *no*,

- mentira @>.*
- J: *Ø no es mi pelada,*
Ø es mi --
Ø es mi [empleada].
- R: *[Ø es mi] recepcionista.*
no,
mentira.
- J: *Ø es la que me pasa las llamadas.*
- R: *@@@@*
- J: *@@@@*
 (H) *no=,*
Ø no es mi pelada.
y además=s,
yo no creo en el amor.
- R: *'and your girlfriend?*
... [the receptionist]?
- J: *[what girlfriend]?*
- R: *<@ no,*
just kidding @>.
- J: *(she)'s not my girlfriend,*
(she)'s my --
(she)'s my [employee].
- R: *[(she)'s my] receptionist.*
no,
just kidding.
- J: *she's the one that transfers my calls.*
- R: *@@@@*
- J: *@@@@*
 (H) *no=,*
(she)'s not my girlfriend.
and anyway,
I don't believe in love.' (Jugo:1089–1106)

Identifiability is often, but not always, linked to the use of articles (cf. Lambrecht, 1994, p. 79). For languages like English and Spanish which have definite and indefinite articles, the indefinite article is said to be used for unidentifiable mentions, and the definite article (and by extension, other definite markers such as possessive adjectives and relative clauses (Du Bois 1980, p. 273)) are thought to mark identifiability. However, Du Bois (1980) shows that this is not always the case; in his (English language) data he finds first mentions with the definite article and second or subsequent mentions with the indefinite article, and either article with non-referential uses of NPs. Similar examples were found in the data used in this study (see (10) and (11)). It is here that the notion of emergent grammar, as

opposed to a priori grammar, is particularly useful in explaining these supposedly anomalous uses of the definite article for the first mention of a referent. The way in which the narrative data were gathered for this study put speakers in an unusual speech situation – they were told to act as though there was no shared knowledge. It is likely, however, that many of them realized that the linguist recording the narrative probably did have knowledge of the video, if not from a direct viewing, then presumably from recording other people talking about it. The way in which the Pear Stories were collected also draws into question the coding for identifiability as the speakers were asked to relate the story to the investigators as if they were telling the story to someone who had not viewed the film, when in actuality, the researchers had seen the film.⁷

Another reason that identifiability is a potentially troublesome category is that unlike discrete categories, such as definite or indefinite article, where the linguist only has to note if it is or is not present, identifiability may be a more scalar category (cf. Lambrecht 1994: 84), where there are degrees of identifiability, therefore rendering it difficult to assign values.

In an emergent model of grammar, in which grammar arises from speech, we can make sense out of apparent form-function mismatches that are evident in the data. Example (10) illustrates a first mention with the definite article, and through a more thorough understanding of the circumstances under which this token was produced, we have a better framework for understanding the coding of this first mention with the definite article. Example (11) shows a second mention of the boy occurring with the indefinite article. Again, a focus on the communicative context allows us to provide a possible explanation for this apparent deviation from the normal pattern of definite markers with recently mentioned referents. The shift from the narrative task to a reflective task, evidenced in the phrase *no entendí mucho* ('I didn't understand much'), and the subsequent return to the narrative task provides insight into possible motivations for the second indefinite mention. This attention to the online editing of a narrative and focus on the discourse allow us to see how establishing identifiability can be viewed more as process than as a singular act.

(10) First mention with definite article

*primero **el= señor** recogía sus manzanas,*
 'first **the= man** was collecting his apples,' (PS36:17)

7. In later fieldwork conducted by the author in 2015 with Spanish-speaking children in the US, the Pear Story narratives were collected without instructions to tell the story to someone who had not viewed it before and a higher rate of first mentions with definite articles were observed. This suggests that the Ecuadorian speakers were indeed treating the referents as less identifiable than the Spanish-speaking children in the US.

(11) Second mention with indefinite article⁸

... primero pasa ... un niño,
 no entendí mucho,
 pero pasó **un niño** con un= --
 ... con un becerro,
 '... first along comes a boy,
 (I) didn't understand much,
 but along came **a boy** with a= --
 ... with a calf,'

(PS12:20–23)

Many researchers report coding referents for identifiability (see, for example, Bentivoglio, 1993; Ewing, 2005; Thompson, 1997), yet a standard method for coding for identifiability is not widely employed. It would seem that coding for identifiability would require some insight into a speaker's thought process, or at least knowledge of what a speaker considers their interlocutor capable of identifying. As linguists do not have the privilege of accessing speaker's thoughts, nor can they predict with any great certainty what two referents will be mutually identifiable to two speakers, it is very difficult to objectively code for identifiability. In addition, the method of data collection used in collecting the narratives for this study complicates the ability to determine what a speaker might judge to be identifiable.⁹ While the notion of identifiability is very useful to understanding the differential treatment of new referents, for this study, identifiability was not coded independently, instead the discussion focuses on how observations of the speakers' use of definite and indefinite articles with the introduction and subsequent mention of referents (see Chapter 6) informs our understanding of how identifiability relates to linguistic coding.

8. Some may argue that this is a repetition of the first mention and therefore not a true subsequent mention. It should be noted that there are no prosodic clues (i.e., truncation) that this the second mention is a repair intended to replace the first mention, and in addition, it is not an exact repetition, as the tense changes from present in the first mention to past in the second mention. Du Bois (1980, p. 260) refers to these types of repetitions as 'digression-and-repeats' and notes that they are responsible for 'late indefinites' in English language Pear Film narrations as well.

9. Participants narrated the Pear Film to the investigators, whom they knew had seen the Pear Film. However, they were instructed to narrate the film as if they were telling the story to someone who had not viewed the film. This differs from the method of collection described in Chafe (1980, p. XV), in which narrators were told that they were narrating the film to someone who had not previously viewed the film. However, as mentioned previously, there was a higher rate of the definite article with first mentions among speakers in a different study who did know that the person they were narrating to had seen the film.

2.3.4 Specificity

Another dimension linked to information flow is specificity. Linguistic expressions designating third person (as is the case for second person) can fall into two types of classes: general and specific. Forms that refer to a class of entities, or a member of a class which is interchangeable with others, as in (12), are said to be general.¹⁰ NPs or other forms of unique reference are said to be specific, they are not interchangeable with other members of the class, such as (13).

(12) General NP

- B: *inclusive en los aviones hay --*
hay un --
un gran cuidado de las --
(H) de las azafatas,
- B: 'including in (the) airplanes there is --
 there is a --
 great care from the --
 (H) from the flight attendants,' (Hermanos:358–361)

(13) Specific NP

- R: *no ves ø que está súper descargado ese teléfono?*
- R: 'don't (you) see that that phone is super
 uncharged?' (Jugo:819)

Specificity has also been linked to anaphoric coding. Lambrecht (1994, p. 80) points out the relationship between specificity and anaphoric reference: "If the referent is specific, the anaphor must be a definite pronoun or noun phrase ... If the referent is non-specific, the anaphor must be an indefinite expression". Ewing (2005) shows that specific referents are more likely to be referred to with lesser coding (i.e., pronouns, unexpressed mentions) than are forms with a generalizing function. This is likely related to the relationship between specificity and discourse referentiality. As most general expressions of third person are also non-referential, their linguistic coding is not subject to influence by information status, and therefore speakers tend not to use pronouns and unexpressed mentions for these expressions. Conversely, specific expressions of third person tend to be referential, and therefore may be coded as pronouns and unexpressed mentions when given.

Ashby and Bentivoglio (1993) incorporate specificity into their analysis of PAS in French and Spanish and find that it has significant correlations with two

10. A terminological distinction between general and generic should be made here. General NPs are all NPs that are not specific; generic NPs are a type of general NP, but there are other general NPs which are not generic.

phenomena – newness and syntactic role. They find that new participants are more likely to be general than specific (generalizing and particularizing in their terms), and that in Spanish, specific uses of NPs are more likely to be found in the S role than in the O role. Ashby (1995) reports additional findings for new referents in French, with specific referents more likely to appear in locative, subject of copular verb, S, oblique, *avoir*-cleft and *avoir*-simple syntactic roles, while dispreferred roles for specific uses of NPs include A (subject of a transitive verb), O, and two structures with *y'a*. These last two studies do not exclude non-referential uses of NPs from their data, and it is likely that this has resulted in a skewed look at the influence of specificity on syntactic role of new referents.

2.4 Genre differences

As more and more linguists work with corpus data, a fuller understanding of the scope of genre differences is essential. In Biber and Conrad's (2009, p. 23) words:

Given the ubiquity of register/genre variation, and understanding of how linguistic features are used in patterned ways across text varieties is of central importance for both the description of particular languages and the development of cross-linguistic theories of language use.

In addition to cross-linguistic theories, it is also important to understand the effects of genre and the extent of variation produced from one genre to another in order to better interpret results of different studies. When the results of one study are compared to the results of another, differences in rates of linguistic form are sometimes cited as evidence of dialectal differences or language change (cf. Cameron, 1993; Flores-Ferrán, 2002; and Travis, 2007 for discussion of the history of the study of subject personal pronoun expression in different varieties of Spanish; Gutiérrez, 1995, for a comparison of the rates of the synthetic and periphrastic futures in two dialects of Spanish). These changes are sometimes attributed to phenomena such as language contact, however often there are multiple causes for differences in rates of linguistic forms. This study contributes to the advancement of our understanding of how usage shapes grammar via the observance of and subsequent interpretation of the links between rates of occurrence of linguistic forms and genre-specific communicative goals and actions.

Variation in distribution of linguistic form between genres has been noted in several studies (cf. Biber, 1988; Biber et al., 2006; Travis, 2007), but genre differences have not been well documented with respect to information flow, although most studies of information flow have focused on spoken language, and different genres are represented. The bulk of the literature has focused on narratives

and monologic speech (e.g., Ashby & Bentivoglio, 1993, 1997; Du Bois, 1987; Durie, 2003; Givón, 1983, *inter alia*), although some have examined conversation (e.g., Ewing, 2005; Kärkkäinen, 1996; Kumpf, 1992; Matsumoto, 2003), and other genres (Kumpf, 2003; O'Dowd, 1990).

Oral narratives and spoken conversations are both genres of spoken language and are similar in that they usually represent the online production of speech. There are marked differences between the two genres, however.¹¹ Narratives tend to be largely monologic, with a focus on relating one person's experience, and this can affect distribution of linguistic forms (for example, the finding that there is greater continuity of subject in narratives than in conversations (cf. Travis, 2007, p. 130)). Conversations are multi-party, and speakers are challenged with simultaneously balancing cognitive (i.e., information flow) pressures with social/interactional pressures (i.e., turn-taking) when choosing linguistic forms.

Another way to think about the differences between the genres is in terms of information flow pressures. Both Du Bois (1987) and Durie (2003) discuss information flow pressure (i.e., the greater the number of referents, the higher the information flow pressures). Du Bois (1987, p. 835) predicts that in texts with higher information flow pressures (i.e., narratives), there will be a stricter adherence to Preferred Argument Structure (as evidenced in the alignment of S and O roles for containing lexical mentions), and that the proportion of intransitive predicates (i.e., NPs in the S role) used for introducing new mentions will be higher than in texts with lower information pressures (i.e., conversations). He proposes two measures of information pressures: the ratio of new humans to clauses, and the ratio of new entities of any kind to clauses (or IUs) (Du Bois, 1987, p. 834). Durie (2003) finds that, in Acehnese, the greater the information pressure, the greater the tendency for the S role to fragment, i.e., to pattern in some cases like O (object) and sometimes like A (subject of a transitive verb), rather than show a more ergative alignment (pattern like O), as predicted by Du Bois (1987).¹²

11. It should be noted that the terms 'narrative' and 'conversation' are used here to describe the two datasets, but that in practice, there are different kinds of narrative and conversation. The narratives of the Pear Film used in this study are one specific type of narrative and the communicative goals and distribution of forms (for example, occurrence of first person) may be markedly different than other types of narratives, such as personal narratives, as are frequently described in Labov's work (cf. Labov, 2014). Conversations also vary in their communicative practices and can contain narrative sequences.

12. Recall that in ergative languages, A and O share similarities, whereas in nominative/accusative (non-ergative) languages, A and S are similar, and O receives different linguistic treatment.

A few studies (Kärkkäinen, 1996; Kumpf, 2003; O'Dowd, 1990) do look specifically at genre, and have tested Du Bois' hypotheses. O'Dowd (1990) studied PAS in the high-information flow pressure genre of an orally-delivered paramedic training session. She finds some of the principles of PAS do apply (e.g., avoid lexical A, one lexical argument per clause), but that in this genre S does not align with O in introducing new information, but it does align with O in maintaining topic continuity. Kärkkäinen (1996) looked at PAS in a low information pressure genre, American English conversation. Like O'Dowd, she found that the S and A role pattern similarly (non-lexical NPs, given information) in non-narrative portions of conversation, and that new information is generally introduced in the O or Oblique roles. She concludes that her observations of the patterns of argument structure in American English are not due to genre differences, rather that there is no evidence of ergative patterning in nominative/accusative languages like English, the focus of her study, nor other nominative/accusative languages which have been studied such as French (cf. Ashby, 1995), Tamil (Herring, 1989) and Finnish (Helasvuo, 1993). Kumpf (2003) noted conformity with the principles of PAS in teacher's classroom discourse. She also finds in her data that free (unattached) NPs can be used for introducing new information, although the preferred information status of free NPs appears to be given (85%) in her data.

In sum, there is substantial evidence that PAS principles (one lexical argument, avoid lexical A, etc.) are fairly universal, but that the S role is subject to different pressures, including information pressure and, perhaps, genre. This study contributes to this body of literature by direct comparison of two different genres.¹³

2.5 Summary of research questions pertinent to information flow

To recapitulate, while there has been extensive research on how information flow affects NP form and distribution, gaps in the literature remain. This study examines how different information flow properties influence the coding of third person references. Specifically, one important contribution of this study is the methodological importance of distinguishing referential from non-referential expressions. This methodology can be applied to future studies to exclude referents whose form is not affected by information status.

13. See also Dutra (1987) for an account of the split S role in Brazilian Portuguese, where she reports that the preverbal S aligns with A (also strongly preverbal) in having a strong tendency to code given information and postverbal S aligns with O in both word order and preferred information status (new).

Among those third person references that have forms subject to the influence of information status, this study seeks to see how they are treated differently, which general tendencies are cross-genre-neutral and which are genre specific, and how identifiability is evidenced in the correlations with linguistic form.

This study also investigates the factors that contribute to the form and distribution of third person expressions which are not affected by information status (i.e., non-referential uses of NPs), and seeks to provide a better understanding of the discourse roles associated with non-referentiality.

This then allows us to better understand the different dimensions of information flow and how they each affect the grammar of third person. While this case study of the different issues related to information flow provides a much needed synthesis of information flow factors, it was found that information flow factors alone are insufficient for accounting for the form and distribution of NPs in spoken Spanish. The next section introduces other influential factors that have often been overlooked in the literature – the role of dialogic interaction, as illustrated in (14) below, where activation status fails to account for the realization of the NP *Gracie*.

(14) NP realization not explained by activation status

J: *solo está la Gracie,*

Ø te das cuenta?

R: *.. solo **la Gracie** viene.*

J: 'just Gracie is there,
do (you) realize?

R: *.. just **Gracie** is coming.'*

(Jugo:516–518)

2.6 Other factors influencing the linguistic coding of third person

While much research has focused on the cognitive constraints (i.e., information flow) that influence the linguistic coding of third person, less research has focused on other factors that may influence the forms used for references to third person. There is, however, a growing body of literature that points out the ways in which factors unrelated to information flow influence the linguistic coding of third person references. These studies help inform our understanding of the range of factors that influence third person reference.

One of the studies that challenges the models of anaphoric coding centered on activation status (and the underlying cognitive motivations) is found in Fox (1987). In this study, Fox points out that when a stretch of speech is examined, not all anaphoric references can be accounted for strictly in terms of information flow or referential distance. In fact, there are often references whose form

is markedly different than what would be predicted by activation status, and she suggests that speakers may be using unexpected forms in anaphoric reference for interactive purposes.

One of the main differences between Fox's (1987) work on anaphoric coding and traditional (e.g., Givón, 1983) models of anaphoric coding lies in the conceptualization of discourse units. Some researchers (e.g., Chafe, 1987; Du Bois, 1987; Givón, 1983; Croft, 1995) take the stance that the basic unit of discourse is the clause or the intonation unit. Fox (1987), following a Conversation Analysis convention, uses the notion of the adjacency pair (cf. Levinson, 1983) in her analysis.¹⁴ In her study she finds the following: "The first mention of a referent in a sequence is done with a full NP. After that, by using a pronoun the speaker displays an understanding that the proceeding sequence has not been closed down" (Fox, 1987, p. 18). In addition, full NPs are not found only for the first mention of a referent, but they are also used to signal that a preceding sequence has been closed. Fox also reports non-structural uses of full, non-new NPs, including disagreements, assessments, and referent negotiation.

Another important contribution toward a fuller understanding of how interactional factors shape discourse is seen in the role of increments. Increments are "constituents that are added to turns that, a just prior point, are interpretable as possibly complete syntactically and prosodically" (Ford, Fox and Thompson, 2002, pp. 15–16), cf. Schegloff, 1996). Building on previous work on turn-taking (cf. Sacks et al., 1974 and Ford & Thompson, 1996), Ford et al. (2002) find that speakers use increments to resolve speaker change problems. Specifically, they argue that speakers use unattached NPs to bridge gaps between the end of one turn and the beginning of another when the initial (syntactic, prosodic and pragmatic) cues to end a turn failed to result in speaker change.

The literature on the interactive work of list construction has also contributed to a better framework for understanding the ways in which speakers use different structures for specific communicative and interactional goals. Jefferson (1990) first drew attention to list-construction in conversation, noting that lists tend to have three parts, which tend to be structurally similar (e.g. three NPs, three infinitival verbs); Sánchez-Ayala (2003) confirmed the similarities in list construction for Spanish. Their findings imply that while lists are under construction, participants in the conversation rely on their previous experience with lists in order to project what will come next. Helasvuo's (2001a) work on lists has also contributed to the understanding of list-construction by highlighting the interactive functions of list

14. Adjacency pairs are not found in all genres; they are a product of dialogue and are absent in monologic data.

construction as a way of managing discourse sequences. It has also been noted that each item in a list tends to occupy its own intonation unit (see Sánchez-Ayala, 2003, for a prosodic analysis of lists in English and Spanish). Helasvuo (2001a) also observes interactional uses of full NPs in predicate nominal constructions, as speakers collaborate in co-constructions; however this is not a very common occurrence in her data. Dumont (2006) also finds evidence of interactional motivations for linguistic coding in her study of third person subjects in Spanish, in which the repetition of a full NP over two consecutive clauses is linked to interactional factors – repetition and contrast. Tannen's (2007) work on repetition in conversation also highlights the interactionally motivated uses of repetition in discourse to facilitate linguistic “production, comprehension, connection, and interaction” (2007, p. 101).

Drawing from these insights, the present study employs an integrative approach to the relationship between discourse structure, interactional concerns and linguistic coding. It builds upon previous research by considering the many different ways in which speakers attend to clauses, IUs, turns, lists, co-constructions and sequences in their speech, for a more comprehensive understanding of the linguistic coding of third person.

2.7 Summary

The following research questions have been developed after taking into account the research mentioned above.

- What are the different and “competing” factors that influence the linguistic expression of third person?
- How can we operationalize referentiality in order to improve our understanding of how this form-function link is manifested in Spanish discourse?
- How do information flow constraints show different manifestations under different genres (i.e., different levels of information pressure)?

Chapter 4 discusses the coding and Chapter 5 examines the methodological question of how to distinguish between referential and non-referential expressions. Chapter 6 examines the form and distribution of referents, with special attention to the effects of information flow and information pressures. Chapter 7 shows the ways in which non-referential expressions respond differently to particular discourse pressures than referents, as reflected in their form and distribution. Observed genre differences in linguistic form and distribution are discussed in Chapter 8, and concluding thoughts are found in Chapter 9.

CHAPTER 3

Data

3.1 Data collection

The data for this investigation come from a large corpus of Ecuadorian Spanish that I collected in Quito between June and August 2006. I worked with an Ecuadorian colleague and three local assistants to collect 35 narrations of the Pear Film (Chafe, 1980), 31 spontaneous conversations, and 12 sociolinguistic interviews. All speakers are middle class, high school or college educated, and range in age from 18 to 84.

Ecuadorian Spanish, or more specifically Quiteño Spanish, contains a number of morphosyntactic traits that are not found in all other varieties of Spanish. Some of those are verbal, such as the mirative and evidential functions of the Present Perfect (cf. Dumont, 2013) and the imperative use of the synthetic future (cf. Haboud, 1998). Others are nominal and include *léismo*, or the use of the pronoun *le* for both masculine and feminine direct objects (the pronoun *le* is only used for indirect objects in some other varieties of Spanish) and the use of the definite article before proper names. In addition to morphosyntactic differences, there are also lexical and phonetic differences from other varieties of Spanish, such as loanwords from Quichua (*naño* for ‘brother’, *muchar* for ‘to kiss’) and assibilated /r/, especially in the older generations. Haboud (1998), and to a lesser extent, Lipski (1994) detail many of the uniquely Ecuadorian features of this variety. There is no reason to believe, however, that the results presented in this volume are only indicative of Ecuadorian Spanish, rather it is very probable that similar results would be obtained from data obtained from other regions of the Spanish-speaking world.

Two genres are included in this investigation: the elicited monologic narratives and spontaneous conversations. Each genre provides its own advantages to the study of NPs, and the inclusion of both genres allows for comparison between the two. In one genre, the narratives, the information pressures are high due to the large number of new referents that are introduced over a relatively short span of speech (cf. Du Bois, 1987), yet the social/interactional dynamic is

low, as one speaker has the floor and there is no competition for taking a turn at talk. Therefore, the study of narratives allows for an in-depth analysis of how speakers maximize syntactic resources while introducing and tracking several referents throughout a narrative in a short span of time. In the other genre, the conversations, the information pressure is low (as observed in the number of referents), yet the social/interactional dynamic is high. A comparison between the two genres provides an ideal circumstance in which to compare the relationships between information flow and interactional concerns and linguistic coding (realization of the different forms and their distribution in these two types of discourse).

3.2 Pear film narratives

The Pear Film was designed as a tool for linguists to elicit narrations of the same experience from different speakers. There is no dialogue in the film, but other sounds were included. It has been used to collect data from many different languages (cf. Chafe, 1980) and the narratives have been used to investigate many linguistic phenomena. It was chosen for this project because it is a reliable means of collecting comparable narratives across speakers and the design of the film is such that it presents the speakers with many opportunities to track referents, thus giving us valuable insight into the different linguistic resources that speakers draw upon to introduce and reintroduce referents into discourse.

The Pear Film is a six minute film that was made at the University of California at Berkeley in 1975. The film begins with a man picking pears in a rural setting. A man and a goat pass by, and then a young boy on a bicycle approaches and takes one of the baskets of the pears that the man has collected. He rides away, passes a young girl on a bike, loses his hat, hits a rock while distracted and topples over, spilling the basket of pears. Three young boys come into the scene to help the boy with the bicycle pick up the pears and load the basket back onto the bike. The boy starts to continue on his way, when one of the boys whistles to him to come retrieve his hat. He collects his hat from the boy and gives him three pears in return. The three boys set off in the direction of the man who was picking pears, who has just noticed that he is missing one of the baskets. The boys pass by the man while eating the pears, at which point the film ends.

Fifteen narrations of the Pear Story were selected for this investigation, from eight male speakers and seven female speakers. Each recording lasts between one minute, thirty seconds and five minutes, eight seconds, with most lasting between two to three minutes. Details of the narrations are presented in Table 1.

Table 1. Pear stories

PS #	Speaker	Age	Duration (minutes.seconds)	# of words	# of IUs
12	male	23	2.58	552	139
13	male	50	2.05	287	66
14	male	19	1.36	271	77
29	female	25	1.58	338	81
30	male	27	1.54	375	101
33	female	30	1.42	291	58
34	male	23	1.44	239	57
36	female	43	2.24	441	99
38	male	58	2.23	404	105
40	female	33	2.38	539	130
41	male	37	1.30	277	53
42	male	66	3.02	379	82
43	female	26	3.02	542	122
52	female	57	3.26	434	97
100	female	26	5.08	1061	246
Totals			37.30	6430	1513

The participants viewed the Pear Film on a laptop computer with headphones, and after watching the movie were instructed to narrate to one of the investigators what they had seen in detail, as if they were re-telling the story to a friend who had not seen the film. Each digital recording was made in the presence of one of the corpus investigators, but interaction from the investigator was kept to a minimum. In addition, the principal investigator did not know most of the participants prior to making the recordings and is not a member of the speech community. In a few cases there are a minimal number of backchannels produced by the investigator during the narration (saying *mhm*, laughing, etc.), but aside from this these narratives are entirely monologic.

3.3 Spontaneous conversations

Fifteen participants were also selected for this investigation from the spontaneous conversations, 7 males and 8 females, and the first 250 tokens were extracted from each speaker's speech. Three assistants were hired in Ecuador to make digital

recordings of routine conversations with their spouses, families, friends and co-workers. The entire speech event was recorded, beginning with greetings and typically ending when the speakers shifted to another activity. Topics found in the conversations range from topics related to school and work to family, mutual friends, memories, and plans for the future. The conversations are summarized in Table 2. The names of speakers are pseudonyms, as are all names contained within examples.

Table 2. Summary of conversational data¹

Recording title	Speakers included	Age	Duration (hours. minutes.seconds)	# of words	# of IUs
Clases	Amanda, Carolina	26, 21	12.40	2814	754
Birthday	Amanda, Francisco	26, 28	14.46	3006	762
Jugo	Rodrigo, Julio	19, 19	25.08	5225	1425
Café	Gerardo, Paca, Señora Manuelita	50, 46, 84	1.28.52	19,031	4941
Chicas	Adán	19	17.58	4603	1335
Fumar	Adrián	18	24.02	5139	1261
Viejos Amigos	Eva	19	40.03	9069	2459
Hermanos	Beatriz	57	1.13.45	17,535	4684
Deportes	Alicia	23	14.56	3765	933
Food	Lucy, Eduardo	28, 30	24.55	4486	1107
Totals	15 participants		5.57.05	74,673	19,661

3.4 Transcription method

All data were transcribed according to the conventions of Du Bois, Scheutze-Coburn, Cumming and Paolino (1993). In this method, each intonation unit

1. The duration, number of words and number of IUs is given for each transcript. For the conversational data, the sample from which the tokens in this study were taken did not include the entire transcript but were limited to the first 250 third person references produced by each speaker.

(IU) occupies its own line in the transcription, and a number of production and interactional features (pauses, laughter, overlap, inhalation, etc.) are captured in the transcription. An IU is identifiable as “a stretch of speech uttered under a single coherent intonation contour” (Du Bois et al. 1993, p. 47). The transitional continuity of the IU, that is, “the marking of the degree of continuity that occurs between one intonation unit and the next” (Du Bois et al. 1993, p. 53) can be realized as continuing, final, appeal, or truncated, and is noted for each IU. It is important to note that final intonation does not always occur at the end of a clause or the end of a turn, continuing intonation does not necessarily mean that the speaker will continue, appeal intonation is not exclusively reserved for questions, and truncated IUs will not always be followed by a completion or repair of the previously truncated IU. In other words, IUs are defined prosodically, not functionally, although there is some functional correlation.

According to Chafe (1994, p. 69), “Intonation units are hypothesized to be the linguistic expression of information that is, at first, active in the consciousness of the speaker and then, by the utterance of the intonation unit, in the consciousness of the listener.” IUs can range from small, one or two word units such as back-channels (*mhm, oh*), discourse markers (*pues ‘well’*), and free NPs (as illustrated in Chapter 1), to clauses and parts of clauses. Croft (1995) finds for English that nearly all intonation units are also grammatical units (clauses and phrases), but that not all grammatical units are intonation units (that is, an IU can be comprised of more than one grammatical unit). Chafe (1994) notes a correlation between substantive IUs and clauses in English, but also points out that substantive IUs can also be comprised of just a referent (especially free NPs).² A number of linguists (Chafe, 1994; Croft, 1995; Du Bois, 1987:812; *inter alia*) would argue that IUs constitute the basic unit of the grammar of spoken speech. Chafe (1994) proposes a constraint of one new idea per IU, linking the IU to memory, cognition and information flow. By studying how speech and grammatical structures are distributed by speakers across different IUs, we come closer to understanding what constitutes a meaningful “unit of speech” to a speaker at the time that an utterance is constructed. The appendix provides a summary of the symbols used in the transcription.

The Pear Stories were transcribed by myself, and five were checked for accuracy by a native speaker of Ecuadorian Spanish. The conversations were all roughly transcribed by local (Ecuadorian) assistants, in most cases the project assistants themselves, then checked by myself, sent back to a native speaker (most often a

2. See, however, Matsumoto (2003) for a discussion of the correlation between clauses and IUs in spoken Japanese.

participant in the conversation) for clarification, and finally transcribed in the final format by myself. Refined methods of transcription such as this are advantageous to those who pursue discourse analytic approaches to the study of language as they facilitate a deeper level of linguistic analysis that includes important extralinguistic cues.

Coding

All linguistic forms designating 3rd person were extracted from the 15 Pear Story Narratives. This includes pronominal, nominal and unexpressed (verbal) mentions from all syntactic roles yielding 1565 tokens for this genre. The first 250 third person references produced by each of the 15 speakers in the spontaneous conversations (3750 tokens) were extracted.

Excluded from the analysis were forms that could not clearly be analyzed (truncated NPs, NPs for which syntactic role cannot be determined due to truncation of the IU, (partially) unintelligible NPs) and certain expressions that did not exhibit variation (subjects of relative clauses that are coreferential with the head, headless relative clauses, NPs embedded within other NPs, infinitival verbs, 3rd person verbs with impersonal subjects, and fixed clause length phrases such as *Dios le pague* ‘thank you’ (lit. ‘God pay you’, where neither *Dios* (‘God’) nor *le* (‘you’) were included) and *que el Señor te bendiga* ‘may God bless you’). After these exclusions 4418 (1228 from the narratives and 3190 from the conversational data) tokens remained for analysis and were coded for the factors described in the following sections. The excerpts in (15) and (16) illustrate exclusions and the discussion that follows will clarify which tokens were included or not included in this study and why. The underlined forms that are not bolded (and are italicized in the English translations) were not included, the bolded forms were.¹

(15) Excerpt 1

J: ... (2.0) (1) *ø* entro en (2) tu [*<X cocina X>*]?
 R: [(3) *ø* estoy] con (4)
el Julio.
 ... (5) *qué?*
 J: y (6) ***otra vez*** (7) *me* *toca* (8) *firmar* (9) ***la hoja?***

1. Remember that NPs embedded within a larger NP were excluded. In the case of (52) below, *un sombrero. que* (53) *le pertenecía* (54) *al niño*, *que* – .. *que sustrajo* (55) *la* – (56) *la canasta*, ‘(52) a hat. that (53) (le) belonged to (54) the boy, that – .. that stole (55) the – (56) the basket,’ the entire NP *un sombrero... la canasta* (a hat... the basket’ is included, but the NPs embedded within the relative clause (*le, el niño, la canasta*) are excluded.

- R: .. *no*,
 ya,
 solo (10) *una vez*.
- J: .. *ah*,
 fresco.
 .. (11) *esa* *es*.
 ... [qué (12) *chévere* <X *grabadora* X>].
- R: [siéntate,
 (13) *Julio*].
- J: (14) *Dios* (15) *te pague*.
- R: (16) *ø* *quieres* (17) *juguito*?
- J: .. *bueno*.
 .. *gracias* (18) *Rodriguito*.
 (19) *se* (20) *te agradece con* (21) *toda el alma*.
- R: ... (2.0) *toma* (22) *ø*.
- J: .. *gra=cias*,
 @@
 (23) *qué más*,
 (24) *Rodriguito*?
 qué tal --
 (25) *qué hiciste* (26) *ø* *hoy*?
- R: ... (H) *hoy --*
 (27) *ø* *estaba pasando* (28) *las... conversaciones*.
 ... (2.0) (H) *y ya*.
 (29) *tú*?
- J: *cómo es* (30) *ø*?
 como (31) *un diálogo*?
- R: .. *sí*,
 más o menos.
- J: .. (32) *o sea*,
 (33) *tú pones* (34) *Julio* *dice* (35) *esto*,
 (36) *Roddy* *dice* (37) *esto*.
 .. (38) *Junior* *dice* (39) *esto*.
 ... (40) *Junior* *es vago*.
- J: '... (2.0) (1) *do I go into* (2) *your* [<x *kitchen* x>]?
- R: [(3) *I'm with*
 (4) *Julio*].
 ... (5) *what*?
- J: *and* (6) *this time* *do* (7) *I* (8) *have to sign* (9) *the*
 paper?
- R: ..*no*,
 yeah,
 just (10) *one time*.
- J: .. *ah*,
 cool.

.. (11) that's it.
 ... [what (12) a cool <X recorder X>].
 R: [sit down,
 (13) Julio].
 J: (14) God pay (15) you.
 R: do (16) you want (17) juice?
 J: .. ok.
 .. thanks (18) Rodrigo.
 (19) (20) you are appreciated with (21) all my heart.
 R: ... (2.0), (22) $\emptyset$ take.
 J: .. thanks,
 @@
 (23) what else,
 (24) Rodrigo?
 how --
 (25) what did (26) you do today?
 R: ... (H) today --
 (27) I was transcribing (28) the.. conversations.
 ... (2.0) (H) yeah.
 (29) you?
 J: how is (30) it?
 like (31) a dialogue?
 R: .. yes,
 more or less.
 J: .. (32) so,
 (33) you put (34) Julio says (35) this,
 (36) Roddy says (36) this,
 .. (38) Junior says (39) this.
 ... (40) Junior is lazy.

(Jugo:59-96)

(16) Excerpt 2

es (41) lo que hizo (42) el niño,
 (43) $\emptyset$ siguió,
 ... pero,
 ... (44) $\emptyset$ siguió (45) su --
 (46) su camino,
 no?
 (H) (47) los tres niños,
 .. también comenzaron a --
 a --
 (H) a regre- --
 a regresar.
 por donde (48) el ca- --
 por donde (49) el niño había venido.
 (H) y (50) $\emptyset$ encontraron (51) un= --

(52) un sombrero.
que (53) le pertenecía (54) al niño,
que --
.. que sustrajo (55) la --
(56) la canasta,
no?
entonces,
(57) los niños,
al encontrar (58) el sombrero,
... (59) lo --
(60) lo recogieron,
y (61) le --
y (62) ø (63) le llamaron (64) la atención (65) al niño,
que ya estaba --
.. ya estaba (66) algunos metros.. lejos de (67) ellos.
entonces (68) ellos (69) le devolvieron muy amablemente
(70) su sombrero,
'that's (41) what (42) the boy did,
(43) (he) went,
... but,
... (44) (he) went on (45) his --
(46) his way,
no?
(H) (47) the three boys,
.. also started to --
to --
(H) to re- --
to return.
from where (48) the ro- --
from where (49) the boy had come.
(H) and (50) (they) found (51) a= --
(52) a hat.
that (53) (le) belonged to (54) the boy,
that --
.. that stole (55) the --
(56) the basket,
no?
so,
(57) the boys,
upon finding (58) the hat,
... (59) (it) --
took (60) it,
and (61) (le) --
and (62) (they) (63) (le) caught (64) the attention of (65)

the boy,
who was already --
.. was already (66) a few meters.. away from (67) them.
 so (68) they very nicely returned (69) his hat to (70)
him, ' (PS12:74–103)²

Included in the analysis were:

- full NPs ((4), (6), (9), (10), (13), (17), (18), (21), (24), (28), (31), (34), (36), (38), (40), (46), (47), (49), (52), (57), (58), (64), (65) and (70));
- pronominal NPs (including (60), (63), (68) and (69));
- unexpressed NPs, such as (43), (44), (50) and (62);
- interrogative pronouns, see (5), (25);
- demonstrative pronouns, i.e., (11), (35), (37) and (39);
- and other pronouns (such as *algo* ‘something’, *alguien* ‘someone’, etc., not illustrated above).

Excluded from this study were:

- references to first and second person ((1), (3) (7), (15), (16), (20), (26), (27), (29), and (33));
- NPs that were unclear to the transcriber, as in (2) and (12);
- infinitival subjects, such as (8);
- impersonal subjects, for example, (19) and (30);
- NPs occurring in set phrases and discourse markers ((14), (22), (32)),
- headless relative clauses (41);
- truncated NPs (see (45), (48), (51), (55), (59) and (61));
- and NPs embedded within other NPs (42), (53), (54), (56), (66) and (67).

4.1 Syntactic and semantic factors

4.1.1 NP realization

Referent realization is often said to be related to information flow (Ariel, 1990; Ashby & Bentivoglio, 1993; Du Bois, 1987, 2003b; Givón, 1983; Prince, 1981).

2. PS stands for ‘Pear Story’, the number refers to the speaker, and the number following the colon refers to the line number in the transcription. Examples taken from the conversational portion of the corpus include the title of the conversation and transcript line number(s) for that example.

More specifically, it has been proposed that there is a link between the availability/ accessibility of a referent and the realization – more available/accessible referents (often recently mentioned in the discourse) show a tendency toward being realized pronominally or as unexpressed mentions, and conversely, less available/ accessible referents tend towards showing more explicit marking (full NPs, cleft and focus constructions, etc.). In other studies, referent realization has also been linked to interactional factors (Ford & Fox, 1996; Fox, 1987; Helasvuo, 2001a). Six different types of NP realization were coded in the data in order to test hypotheses about proposed form-function links:³

- full NP⁴
- personal pronoun (subject *él* ‘he’ and *ella* ‘she’, direct object *lo*, *la*, *le*, *los*, *las*, *les* and indirect object *le* and *les*), demonstrative pronoun (*eso*, *esa*, *ése*, *esos*, etc.), other pronoun (*algunos* ‘some’, *otros* ‘others’, *alguien* ‘someone’, *todos* ‘all, everyone’, *nada* ‘nothing’, *una* ‘one’, etc.), unexpressed mentions, interrogative pronouns (*qué* ‘what’, *quién* ‘who’, etc.).^{5,6}

4.1.2 Modification

In order to look for evidence of form-function links and patterns as they relate to information flow and referent modification, all full NPs were coded for type of determiner: none (bare), definite article (*el*, *la*, *los*, *las*), indefinite article (*un*, *una*, *unos*, *unas*), demonstrative adjective (*este*, *ese*, etc.), descriptive adjective (*otro*, *buen*, *mala*), possessive adjective (*mi*, *tu*, *su*, etc.), quantifier (*todo*, *mucho*, etc.), interrogative adjectives (*cuál*, *qué*, etc.), and whether or not the NP was modified by a relative clause.

3. Three additional realizations were found but excluded: subjectless phrases (*ø es que* ‘it’s that’, *ø es cierto* ‘it’s true’), headless relative clauses (*lo que ...*), and truncated NPs.

4. This includes proper names, which, like other full NPs, contrast with pronouns and unexpressed mentions. Other researchers (e.g., Prince (1981) and Ewing (2005)) have pointed out that proper names often differ from other full NPs with respect to information flow properties (accessibility, referentiality, etc.); these properties are captured in other aspects of the coding for this study.

5. Ecuadorian Spanish is a *leísta* variety in which direct object pronouns often take the forms *le* and *les*, which are only used as indirect objects in other varieties.

6. There are different ways of coding for these factors. One way which may be fruitful for future studies would be to separate the object pronouns that occur as part of a construction (particularly when they appear with the full NP) from those which do not.

Definiteness in Spanish (i.e., modification by a definite article, demonstrative adjective, possessive adjective or quantifier) is thought to be associated with identifiability of a referent (cf. Bosque, 1989; Ortiz Ciscomani, 2009), indefiniteness with newness and unidentifiability (Garachana Camarero, 2009), while lack of modification is associated with non-referentiality (Torres Cacoullos & Aaron, 2003). Du Bois (1980) has shown a similar relationship between the use of definite and indefinite markers and identifiability of referents in English. Other uses of modifiers not associated with information flow are explored more thoroughly in Chapter 7.

Fox and Thompson (1990) show that in American English conversation there are distinct patterns in the distribution of grammatical roles of NPs that are heads of relative clauses and the grammatical role of the coreferent within the relative clause, which they attribute to speakers' attention to information flow.⁷ While NPs embedded within relative clauses were not extracted for this study, it was noted which NPs are heads of relative clauses in order to better examine their role in managing information flow. An example of an NP modified by a relative clause follows in (17).

- (17) NP modified by relative clause
pero decidió al final,
*coger **todo el canasto que estaba lleno.***
 'but (he) finally decided,
 to take **the whole basket that was full.**' (PS12:41)

4.1.3 Syntactic role

Numerous studies have pointed out the relationship between syntactic role and information flow (Ashby & Bentivoglio, 1993; Du Bois, 1987, 2003b). Studies of Preferred Argument Structure (i.e., Ashby & Bentivoglio, 1993; Du Bois, 1987, 2003b) indicate that speakers tend to limit the amount of new information in a clause to just one new argument per clause, and that new information is most often found in the S (subject of intransitive verb) or O (object) roles. It has also been consistently found that referents in the A role are almost exclusively given, and therefore generally unexpressed or pronominal.

7. For example, in the sentence 'probably the only thing you'll see is like the table' from Fox and Thompson (1990:302), the NP 'the only thing' is the subject of the main clause, but the object of the relative clause.

There have also been studies demonstrating links between particular syntactic role and interactive tasks not related to information flow (Ford, Fox & Thompson, 2001; Helasvuo, 2001a). Eight different syntactic roles were distinguished in this study. Core arguments include subject of transitive verb, subject of intransitive verb (which includes single arguments of verbs that are not truly transitive, such as (18) and (19)), direct object, indirect object; non-core arguments include object of preposition, adverbial, predicate nominal, and vocative.^{8,9}

(18) Single argument of *haber*

siempre hay alguien que nos va a ayudar.

'there is always **someone** that is going to help us out.'

(PS12:63)

(19) Single argument of *hacer*

pero está haciendo basta=nte so=l.

'but it is really sunny [lit. it makes a lot of sun].'

(Clases:7)

The morphosyntactic coding of some direct objects in Spanish is done with the *a personal* (personal *a*), and indirect objects are also marked with *a* in Spanish.¹⁰ These arguments, exemplified in (20) and (21), were coded based upon their relationship to the verb (that is, they are directly or indirectly affected by the action), and not as objects of prepositions, as the NP *Baños* following *a* in (22) was coded (locative in function).

(20) Direct object with *a*

ahora sí le puedo ver a la Sonia.

'now I can see Sonia.'

(Jugo, 490)

(21) Indirect object with *a*

y le entrega el sombrero al niño.

'and (he) gives the hat **to the boy**.'

(PS100:209)

(22) Object of preposition with *A*

porque no me fui a Baños,

'because I didn't go **to Baños**.'

(Deportes, 61)

8. Note that all subjects of 'Complement-taking predicates', i.e. *decir* 'say', *pensar* 'think', *ver* 'see' were coded as subjects of intransitive verbs (cf. Thompson, 2002).

9. Although vocatives functionally overlap with the category of second person (see Donati, 2010, for discussion of vocatives as a marker of second person), formally they most closely resemble third person and for that reason were include here.

10. The use personal *a* with direct objects is variable and is related to semantic properties (animacy, specificity) (cf. Butt & Benjamins 2004:33), but is also correlated to other social and discursive factors (see Balasch 2011 for a variationist account).

A separate code was given to those interrogative pronouns that do not fulfill a clear argument function, but whose role is merely grammatical (see (23)).¹¹ Another code was created for those NPs which are an extension of a previous syntactic frame, as in (24), where *una parte* (*en el aeropuerto*) is interpreted as directly following from the verb *ver* (which has a separately expressed direct object), as opposed to being a syntactically isolated (free) NP. Lastly, there were a number of NPs that did not fit into any other of the above roles, including appositives as in (25) and syntactically isolated (“free”) NPs as in (26), which were put into a heterogeneous category of “Other”.

- (23) Interrogative/grammatical role
qué dijo tu mami ayer,
 ‘**what** did your mom say yesterday,’ (Jugo:386)
- (24) Extension of previous syntactic frame
 S: .. vi una parte,
 en mi casa,
 F: ya.
 S: de ahí **una parte** en el aeropuerto,
 S: ‘.. I saw part [of the soccer game],
 at my house,
 F: yeah.
 S: and then **another part** at the airport,’ (Deportes:83–86)
- (25) Appositive
 (H) *pasaron los tres niños,*
los que le ayudaron al niño a recoger las --
las peras,
 ‘(H) the three boys passed by,
 the ones that helped the boy collect the --
 the pears,’ (PS36:56–58)
- (26) Syntactically isolated NP
pero a la una,
de una a dos hay supletorio.
.. supletorio de productos del banco.
 ‘but at one o’clock,
 from one to two there is a make-up exam.
 .. **make-up of [the exam on] banking products.**’
 (Birthday:158–60)

11. In traditional analyses, many of these instances would be assigned traditional syntactic roles, i.e., direct object for (23). For the purposes of this study, they were given a separate code because the discourse functions interrogative pronouns perform are quite different from discourse functions of direct objects that are directly related to the semantics of the verb.

4.1.4 Semantic class of verb

There are correlations between different types of verbs and different discourse functions. Motion verbs, for example, have been shown to often introduce new referents in Spanish (Bentivoglio, 1993; Dumont, 2006) when compared to other verb classes. Scheibman (2001) shows that relational predicates perform specific discourse functions. In her study of American English, a large number of third person singular subjects are nonhuman referents that occur with relational predicates (mostly copular clauses, e.g., “that is a good deal”), which permit “speakers to index their opinions and attitudes” (2001, p. 85). Accordingly, in this study all subjects and objects were coded for the semantic class of the verb, which were categorized as follows- copula; motion; cognitive process; perception; speech act; “other”; *haber* (‘there is’); and stative, impersonal and experiencer verbs. Table 3 exemplifies the kinds of verbs found in each class.

Table 3. Verb classes

Class	Examples
Copula	<i>estar</i> (‘to be’), <i>parecer</i> (‘to seem’), <i>ser</i> (‘to be’)
Motion	<i>andar</i> (‘to go’), <i>bajar(se)</i> (‘to go down’), <i>entrar</i> (‘to enter’), <i>ir</i> (‘to go’), <i>pasar</i> (‘to pass’), <i>regresar</i> (‘to go back’), <i>subir(se)</i> (‘to go up’), <i>venir</i> (‘to come’), <i>volver</i> (‘to return’), etc.
Cognitive process	<i>aprender</i> (‘to learn’), <i>creer</i> (‘to believe’), <i>darse cuenta</i> (‘to realize’), <i>decidir</i> (‘to decide’), <i>entender</i> (‘to understand’), <i>enterarse</i> (‘to find out’), <i>pensar</i> (‘to think’), <i>recordar</i> (‘to remember’), <i>saber</i> (‘to know’), etc.
Perception	<i>escuchar</i> (‘to listen’), <i>mirar</i> (‘to watch’), <i>observar</i> (‘to observe’), <i>oír</i> (‘to hear’), <i>oler</i> (‘to smell’), <i>sentir</i> (‘to feel’), <i>ver</i> (‘to see’), etc.
Speech Act	<i>decir</i> (‘to say’), <i>agradecer</i> (‘to appreciate’), <i>pedir</i> (‘to ask for’), <i>sugerir</i> (‘to suggest’), etc.
Other	<i>abrir</i> (‘to open’), <i>dar</i> (‘to give’), <i>hablar</i> (‘to speak’), <i>hacer</i> (‘to make/do’), <i>poner</i> (‘to put’), <i>recoger</i> (‘to pick up’), <i>silbar</i> (‘to whistle’), <i>tener</i> (‘to have’), <i>terminar</i> (‘to finish’), <i>tomar</i> (‘to take’), <i>trabajar</i> (‘to work’), <i>vivir</i> (‘to live’), etc.
Haber	<i>haber</i> (existential verb, ‘there is/are’)
Other stative, impersonal and experiencer verbs	<i>encantar</i> (‘to love’), <i>existir</i> (‘to exist’), <i>gustar</i> (‘to like’), <i>importar</i> (‘to matter’), <i>llamarse</i> (‘to be named’), <i>molestar</i> (‘to bother’), <i>pertenecer</i> (‘to pertain’), <i>quedar</i> (‘to stay, remain’), <i>suceder</i> (‘to happen’), etc.

4.1.5 Summary

In sum, based on insights from previous studies of NPs in English, Spanish and other languages, the tokens in these data were coded for four syntactic and semantic factors (realization, modification, syntactic role, semantic class of verb) in

order to investigate how the linguistic coding of these tokens reflects relationships between syntactic, semantic and other factors (specifically information flow and interactional factors, as described in the following sections).

4.2 Factors related to information flow

The effects of information flow on referent realization and distribution have been well documented (Ariel, 1990; Ashby & Bentivoglio, 1993; Du Bois, 1987, 2003b; Givón, 1983; Prince, 1981), as was discussed in Chapter 2. Many ways of measuring information flow and topic continuity have been proposed; the coding of factors related to information flow used in this study draws from several previous studies and attempts to provide a comprehensive yet objective manner of coding for different parameters of information flow in a way which can be reproduced in future studies.

4.2.1 Activation State

Different models of activation state have been proposed; perhaps the two most influential are those found in Chafe (1987) and Prince (1981).¹² In Chafe's (1987) model, recall that there is a three-way distinction between active, semi-active, and inactive referents. In Prince's (1981) paper, she discusses seven different activation states: brand new unanchored, brand new anchored, new-unused, noncontaining inferrable, containing inferrable, textually evoked and situationally evoked. However, because grammars reflect a binary system (cf. Lambrecht, 1994, p. 100) a binary distinction between discourse new (first mention of a referent in the transcript) and given (referent which has previously been mentioned in the discourse) was used here.¹³ There is a subset of referents (general, non-referential referents, described in detail below) for which activation state does not apply. The following example, (27), taken from the beginning of one of the narrative transcripts (the first two lines contained instructions from the investigator), illustrates this coding (note that activation status was not recorded for expressions that were found

12. But see also Prince (1992) for another model. One major addition to her revised schema is the distinction between *discourse* new and old NPs and *hearer* new and old.

13. Many factors (distance from previous mention, previous knowledge, schema, and episode boundaries among others) have been proposed to influence the accessibility of a referent, and these are difficult to separate from one another and to code independently. See Ewing (2005, pp. 122–124) for a detailed discussion of why a binary distinction between given/new referents proves to be a practical and reliable way of coding activation status.

to be used non-referentially, explained further in Chapter 5). First mentions are coded as discourse new (N), the second and subsequent mentions as given (G). The labels “new” and “given” refer to the relationship of the referent to the discourse (new to the discourse, previously mentioned in the discourse) and do not assume that the referents are indeed brand new entities to the hearer nor that they are known, in the case of “given”.

(27) New/given coding

```

.. eh=,
estaba un campesino (N) recogiendo=.. unas=,
eh,
manzanas (N),
de= su árbol (N).
ya Ø (G) tenía.. (H) dos canastas (N) llenas de manzanas,
eh,
es- --
.. era --
(H) ... (2.0) pero=,
... en el momento de.. seguir recogiendo,
y subirse a su árbol (G),
eh,
vino un --
un --
primero pasó un --
.. un burrito (N).
un --
otro señor con un burrito (N).
.. (H) y= --
y Ø (G) lo (G) vio,
y= luego él (G) subió a recoger.. más de --
de su árbol (G).
'\.. eh=,
there was a peasant (N) gathering.. some,
eh,
apples (N).
from his tree (N).
(he) (G) already had.. (H) two baskets full of apples (N),
eh,
es- --
.. was --
(H) ... (2.0) bu=t,
... at the moment to.. keep picking,
and climb the tree (G),

```

```

eh,
along came a --
a --
first came a --
.. a donkey (N) .
a --
another man with a donkey (N) .
.. (H) a=nd --
and he (G) saw him (G) ,
a=nd then he (G) climbed up to pick.. more from --
from his tree (G) .'

```

(PS29:3-25)

4.2.2 Specificity

The distinction made here is between general and specific expressions; general expressions do not name a specific entity, as in *guitarra* in the example below, but specific referents do, such as *Tebi*, the subject of (28). It should be noted that *generic* references are a type of general expression, but not all general expressions are generic.

(28) Specific versus general referents

```

... pero el Tebi ya sabe tocar guitarra?
'... but Tebi already knows how to play guitar?

```

(Café:1699)

In the lengthier excerpt found in (29), a larger range of type of general expressions (G) is seen as they contrast with specific (S) expressions. Of the three different third person human expressions, two types are used to refer to specific people (Ricardo and his boss, *el señor*), but the non-specific third person is used to talk in more general terms about “the people at Ricardo’s work who are in charge of giving permission in these types of cases”. Given speaker G’s relationship with Ricardo, it is possible that he does not know enough about Ricardo’s job situation to be able to speak in more specific terms about who gives permission for days off, but it appears that speaker M knows that he reports to a specific man (*el señor*) for this types of situations. For the non-human general NPs, we see *mañana* (‘morning’) being used to talk about the time of day in relationship to the adverbial expression *mañana* (‘tomorrow’), and not a specific time, such as *a las ocho de la mañana* (‘at eight in the morning’). Likewise, *permiso* (‘permission’) is used to talk about the act of giving permission in general, and no specific type of permission is mentioned. The NP *estos casos* ‘these cases’ was coded as general in that it refers to the verbal situation ‘these cases (in which someone has special circumstances that were alluded to previously)’ or perhaps ‘cases like these’ and not an entity within

the discourse that has been named. Lastly, a specific inanimate NP is exemplified in *el trabajo*.

(29) Coding for specificity

G: y **el Ricardo** (S) no viene?

M: pueda que $\emptyset$ (S) ya venga.

G: oh.

M: porque $\emptyset$ (S) quedó a venir,
 porque **él** (S) me va a llevar mañana de **mañana** (G)..
 ahm,

G: ah=,

y $\emptyset$ (G) **le** (S) darían **permiso** (G) en el trabajo (S)?
 ... (2.0) $\emptyset$ (G) **le** (S) darían **permiso** (G) en el
trabajo (S) al **Ricardo** (S)?

M: **eso** (G) dijo $\emptyset$ (S),

que $\emptyset$ (S) **le** (S) va a pedir al.. **señor** (S).
 cómo no $\emptyset$ (G) **le** (S) van a dar en **estos casos** (G),
 no.

G: .. ah,

no,
 pues,
 a no $\emptyset$ (G) **le** (S) den,
 nosotros estamos prestos a acompañarle.

G: 'and **Ricardo** (S) isn't coming?

M: maybe **he**'ll (S) come.

G: oh.

M: because **he** (S) agreed to come,
 because **he** (S) is going to bring me tomorrow in **the**
morning (G).. ahm,

G: ah=,

and will **(they)** (G) give him **permission** (G) at **his**
job (S)?

... (2.0) will **(they)** (G) give **him** (S) **permission** (G)
 in his **job** (S) to **Ricardo** (S)?

M: **(he)** (S) said **that** (G),

that **(he)** (S) is going to ask **the.. man** (S).

how would **(they)** (G) not let **him** (S) in **these cases**
(G),

no.

G: .. ah,

no,

well,

if **they** (G) don't give (it) to **him** (S),

we are willing to accompany you.' (Café:390-406)

4.2.3 Distance from previous mention

It is often thought that the number of clauses between two mentions (Du Bois, 1987, p. 816; Givón, 1983) can give insight into the different degrees of “old” or “non-new” information in discourse. Specifically, it is hypothesized that the greater the distance between two (co-referential) referents in discourse, the more likely a previously introduced referent will receive linguistic coding associated with less continuous/accessible referents.

The number of intervening clauses was noted for those tokens that had a coreferential mention within the preceding 10 clauses; tokens with a coreferential mention at a distance greater than 10 clauses after the previous mention were given the same code. In (30) there are five mentions (four unexpressed) of the same referent (Mónica), with no intervening clauses between them. In (31), there are three intervening clauses (marked by arrows in the margin) between two mentions of the referent “he/him.”

(30) No intervening clauses

R: *la Mónica?*

A: *ah=.*

.. Ø iba corriendo y shu=,

Ø se resbala,

y Ø se pega una matada,

y Ø se raspó la rodilla,

R: *'Mónica?*

A: *ah=.*

.. (she) was running and shu=,

(she) slips,

and (she) takes a huge spill,

and (she) scratched her knee.'

(Fumar:226–231)

(31) Three intervening clauses

M: *... sus abuelitos también tienen que haber^{le} apoyado*
mu=cho,

en fin,

→ *no es cierto (1)?*

G: *[cla=ro,*

→ *así es (2)].*

M: *[la familia],*

la familia unida,

→ *que es lo mejor (3).*

G: *... pero,*

él,

qué va a seguir ahí?

M: *'... his grandparents also must have supported him a*
lot,

- after all,
 → isn't that right (1)?
 G: [of course,
 → that's the way it is] (2).
 [the family],
 the united family,
 → that's the best (3).
 G: ...[but,
 he,
 what is (he) going to study there?'] (Café:100–110)

4.2.4 Realization of previous mention of the referent

This variable was included in the study in order to investigate the relationship between activation state and referent form; the hypothesis was specifically that newer referents are coded with relatively longer forms, and given referents receive relatively less coding (i.e., a pronoun is more likely to be preceded by a coreferential mention as a full NP than as an unexpressed mention (cf. Givón, 1983)).

Seven types of forms used in the previous realization were found here: personal pronoun, demonstrative pronoun, other pronoun, unexpressed, truncated; a full NP that was the same as the target token (e.g. *el hombre ... el hombre*); and a full NP that was a different lexeme than the target token (e.g. *Sara ... mi hermana*), a kind of re-naming of the same referent.

4.2.5 Discourse referentiality

The importance of discourse referentiality, that is, whether a form is used to refer to an entity that plays an important role in the subsequent discourse, has been acknowledged by many scholars (Du Bois, 1980; Givón, 1981; Hopper & Thompson, 1984). It is very clear that speakers make different choices regarding the linguistic coding of referential uses of NPs when compared to non-referential uses of NPs in speech – both in terms of form and syntax. It is also clear that the discourse referentiality intersects with other dimensions of information flow. For example, it makes little sense to speak of the activation status of a NP used non-referentially. In example (32) below, the NP *profesora* ('teacher') is non-referential, occurring as a predicate nominal which is being used to classify a person. It is the first mention of *profesora* in this conversation, yet upon hearing *profesora*, the interlocutors will not create or access a "file" for this NP. Instead, *profesora* is information that is added to the existing information file about the subject NP and referent (*Erika*, *ella*). Similarly, in (33) in the verb-object compound *tener vacaciones* there is no specific vacation that is activated in the minds of the participants, rather the verbal predicate *have vacation* (or 'be on vacation') is evoked.

(32) Non-referential use of a NP

- P: ... (2.0) y la [Erika],
 M: [tengo] --
 P: ella si está trabajando?
 M: ... ahora ya en vacaciones,
 ya no tiene,
 pues.
 ya --
 .. ya salen.
 G: [hm=].
 P: [oh=].
 ... pero ella es **profesora**,
 P: '... (2.0) and [Erika],
 M: [(I) have] --
 P: is she working?
 M: ... now (she's) already on vacation,
 (she) doesn't have,
 well.
 already --
 .. (they) are already out.
 G: [hm=].
 P: [oh=].
 ... but she is a **teacher**.' (Café:843–853)

(33) Non-referential use of a NP

- C: .. tenemos [**vacación**].
 A: [no,
 pero si] tienen ca- --
 no tienen **vacaciones** por dos meses,
 o cómo es?
 C: no,
 tenemos **vacació=n** cada seis meses.
 C: '.. we have [**vacation**].
 A: [no,
 but if] you have ca- --
 don't you have **vacation** for two months,
 or how is it?
 C: no,
 we have **vacation** every six months.' (Clases:40–46)

However, discourse referentiality is also perhaps the most difficult parameter of information flow to operationalize for a quantitative study, as the term has loosely been used in the literature to apply to different situations, and has not been well defined in a way that is easily replicable. As Ewing (2005, p. 149) points out,

discourse referentiality involves speaker intention, and “cannot be established by a single quantitative criterion such as, for example, number of mentions”.

One of the most satisfying descriptions of the differences between discourse referential and non-referential uses of NPs is found in Du Bois and Thompson (1991), who use the terms *tracking* and *non-tracking*. Tracking referents are those which have “continuity of identity” within the discourse (cf. Du Bois, 1980); non-tracking NPs are found to fulfill a variety of other discourse functions, including predicating, orienting and classifying. Predicating NPs name events, situations or activities, as illustrated in (33) above and (34) below, and in Spanish tend to occur in V-O compounds. Orienting NPs spatially or temporally situate another NP, as in (35), and are usually found as adverbials or as parts of prepositional phrases. Classifying NPs, as in (32) above and (36) below, are usually found as predicate nominals in Spanish and their function is to help classify tracking referents.

(34) Predicating NP

F: *se llevó unos para aprender inglés,*

F: ‘she got some to learn **English**,’ (Birthday:260)

(35) Orienting NP

R: *.. eso haces todos los días,*

R: ‘... you do that **every day**,’ (Jugo:879)

(36) Classifying NP

C: *una laptop ya es una computadora.*

C: ‘a laptop is **a computer**.’ (Clases:604)

Tokens were coded as discourse non-referential if they fulfilled one of the non-tracking discourse functions described by Du Bois and Thompson (1991) and illustrated above, as well as those which have a non-tracking grammatical role, such as interrogative pronouns and vocatives. The remainder were coded as discourse referential, except for a small number which were coded as unclear (n=2 from Pear Stories, n=11 from conversations) because it was not apparent whether they fulfilled a non-tracking function or not. In the excerpt in (37), a number of different non-referential uses are observed with respect to how they contrast with referential uses of NPs (R). Non-referential predicating are indicated with (P), non-referential orienting appear with (O), non-referential classifying occur with (C), and other uses occur with (Q).

Excerpt (37) contains a total of twelve expressions that were coded as performing a non-referential function. Four of these were found to fulfill grammatical or other functions: two are non-referring interrogatives (*qué* and *cuándo*) and two are vocatives (*Eva*, *Rodriguito*). Another three serve orienting functions: *tal vez* and *dos meses* are temporally orienting NPs and *Estados Unidos* is used

to spatially orient speaker E's sister. Only one NP in this excerpt, *el cumple de tu mami*, performs a classifying function, that is, it helps classify the event *hoy* 'today'. Table 4 summarizes the distribution of the non-referential NPs.

Table 4. Non-referential expression found in excerpt 37

Predicating (P)	Grammatical or other functions (Q)	Orienting (O)	Classifying (C)
la mínima idea	qué	tal vez	El cumple de tu mami
un pastel	cuándo	dos meses	
lo básico	Eva	Estados Unidos	
clases de verano	Rodrigo		
4	4	3	1

The most frequent non-referential function exemplified in this excerpt is predication, that is, those expressions that function to name part of an event, activity or situation. The first of these, in *no tengo la mínima idea* ('I don't have the slightest idea') forms part of the situation speaker E names in claiming not having any knowledge or plans (about what she will do that day). The second and third predication expressions are found in a sequence in which E talks about possible plans for the day. Now that they have established (see excerpt 37) that it is E(va)'s mother's birthday, she puts forth two possible activities that she might do in order to help celebrate her mother's birthday. On the one hand, she might buy a cake (*comprarle un pastel*), but on the other hand she might make a cake (with Luisa) (*hacerle un pastel con la Luisa*). The fourth predication expression, *lo básico*, is used to talk about the activity of "studying the basics", a modification of the earlier predicate "studying a lot". And the final expression in this excerpt was coded for a non-referential function because it names part of the situation that is offered as an explanation for why E's sister is returning soon from the United States.

(37) Coding for functions associated with referentiality

- R: .. **qué (Q)** vas a hacer ahora,
Eva (Q) ?
- E: .. no sé=,
Rodrigo (Q) ,
 .. no tengo **la mí=nima idea (P)** .
 ... [**tal ve=z (O)**] ,
- R: [XX] --
 [2hoy no es2] **el cumple de tu mami (C)** ?
- E: [2XsalirX2] .
 .. (H) oh=,

- sí=.
- ... sí=,
- voy a salir,
- entonces,
- a comprarle (R) un pastel (P).
- .. o creo que vamos a hacer un pastel (P) con la
- Luisa (R).
- R: ... oh=.
- y tu ñaña (R)?
- .. cómo está Ø (R) en Estados Unidos (O)?
- E: ah=,
- le (R) va bien=.
- .. Ø (R) está estudiando basta=nte.
- ... pero=,
- ya,
- lo básico (P),
- no más.
- R: .. [en serio]?
- E: [@@@]
- <@sí@>.
- R: .. cuándo (Q) regresa Ø (R)?
- E: .. (H) en dos meses (O) regresa Ø (R) ya,
- porque Ø (R) fue por clases de verano (P).
- R: `.. what (Q) are you going to do now,
- Eva (Q)?
- E: .. I don't know=,
- Rodriguito (N),
- .. (I) don't have the sli=ghtest idea (P).
- ... [ma=ybe (lit. such a time (O))],
- R: [XX] --
- [2isn't today2] your mom's birthday (C)?
- E: [2Xgo outX2].
- .. (H) oh=,
- yes=.
- ... yes=,
- I am going to go out,
- then,
- to buy her a cake (P).
- .. or I think we are going to make a cake (P) with
- Luisa (R).
- R: ... oh=.
- and your sister (R)?
- .. how is (she) (R) in the United States (O)?

E: ah=,
 (it) is going well for **her= (R)**.
 .. **(she) (R)** is studying a lot.
 ... but=,
 yeah,
the basics (P),
 no more.

R: .. [seriously]?
 E: [@@@]
 <@yes@>.

R: .. when does **(she) (R)** come back?
 E: .. (H) in **two months (O)** **(she) (R)** comes back already,
 because **(she) (R)** went for **summer classes (P)**.'

(Viejos Amigos:1-32)

4.2.6 Summary of factors related to information flow

This section described how tokens were coded for five factors related to information flow and how they are important. Three (activation state, distance from previous mention and realization of the previous mention) have to do with other mentions of the same referent in the surrounding discourse, and the inclusion of these factors in this study will help us understand how discourse at large shapes the grammar. The other two, specificity and discourse referentiality, are also known to correlate with the linguistic expression of third person and for this reason are included.

Discourse referentiality

The separation of referential from non-referential forms is an essential step in this comprehensive study of third person, yet it proves to be a daunting task, with previous studies painting a fuzzy picture of what referential forms look like as opposed to non-referential forms. In this chapter, a number of hypotheses regarding distribution of linguistic form and non-referential functions are tested with the aim of developing a clear, comprehensive and replicable way of coding third person expressions for referentiality. Recall from Chapter 4 that all tokens in this study were coded for referentiality following Du Bois and Thompson (1991). Here, the data are examined in order to look for correlations between non-referential functions and formal properties (i.e., presence of modifier, core or oblique role, persistence, etc.) that have been linked to referentiality in the literature. Then, the data were analyzed in order to determine which combination of formal factors may be used as a model of the formal distinction between discourse referential and non-referential uses of NPs.

5.1 Factors related to discourse referentiality

5.1.1 Persistence

The relationship between persistence (i.e., continuity of reference in discourse) and discourse referentiality has often been discussed in the literature (Du Bois and Thompson, 1991; Helasvuo, 2001b; Torres Cacoullos and Aaron, 2003), although the correlation between persistence and discourse referentiality is not straightforward. Torres Cacoullos and Aaron (2003) attempted to use persistence as a measure of tracking status of NPs in their study of lone English-origin nouns in Spanish, but found this to be unsatisfactory as speakers “seem to present NPs as tracking and then do not track them” (2003, p. 324), i.e., the NPs do not persist after the first mention.

In fact, when one examines spoken language, examples such as (38) below contradict the alleged link between persistence and referentiality. In this example, *sol* (‘sun’) is mentioned several times but is clearly performing the discourse

non-referential function of predicating. This shows that NPs may persist, or be repeated in the discourse, while performing non-referential functions, and that persistence is not always linked to referentiality.

(38) Persistent discourse non-referential NP

- A: *qué más,
Caro?
.. te vas a clases ya?*
- C: *.. sí,
sí me voy ahora.*
- A: *.. en serio?
pero está haciendo **basta=nte so=l**.
.. qué te vas a poner?*
- C: *sí,
todos estos días ha hecho **mucho sol**,
pero ahora,
me voy a poner un poco abrigado,
porque estoy un poco con gripe.*
- A: *.. (H) uh=,
pero te vas a calcina=r,
sí está haciendo **un sol**,
creo que estamos a unos treinta y cinco grados.*
- C: *.. sí=,
debe se=r.
.. [ayer] también hizo **un sol** --*
- A: *[uh].*
- C: *.. bastante fuerte.*
- A: *'what else,
Caro?
.. are you going to class already?*
- C: *.. yeah,
yeah I'm going now.*
- A: *.. really?
but **the sun** is really strong.
.. what are you going to wear?*
- C: *yeah,
these last few days there has been a lot of **sun**,
but I'm going to bundle up,
because I've got a little cold.*
- A: *.. uh,
but you're going to fry,
yeah there's really a lot of **sun**,
I think it's around 35 degrees.*
- C: *.. yeah,*

must be.
.. yesterday also **the sun** was --
A: uh
C: .. really strong.' (Classes:1-22)

In Givón’s (1995) sense, an NP is deemed “persistent” if it occurs again in at least one of the following ten clauses.¹ For this study, this measure was combined with Givón’s measure of “lookback,” whether or not a referent occurred in the preceding ten clauses. Therefore, tokens in this study which were also mentioned in the preceding or following ten clauses were coded as persistent. In combining Givón’s persistence and “lookback” all occurrences of both previous and subsequent mentions in the immediately surrounding discourse are captured, whereas if persistence alone were used, only the subsequent mentions would be included. It should also be noted that Givón (1995) assigns referents a Topic Persistence value based upon the number of times that a referent occurs again in the following ten clauses; here only the binary distinction between persistent and not-persistent was made, as one subsequent mention is sufficient for persistence. There are certain NP types and uses (vocatives, *nada* ‘nothing’, *nadie* ‘no one’, *alguien* ‘someone’, *todo eso* ‘all that’, *cuenta* in *darse cuenta* ‘to realize’ etc.) for which persistence is not applicable. Results for persistence and discourse (non-)referential function are given in Table 5.

Table 5. Discourse Referentiality and Persistence in both genres (N=4405)

Persistence	Non-referential		Referential		Total	
	N	%	N	%	N	%
Persistent	634	25	1869	75	2503	100
Not persistent	830	81	189	19	1019	100
N/A	855	97	28	3	883	100
Total	2319	53	2086	47	4405	100

Chi-square = 1779, df 2, *p* < 0.0001

The strongest correlation aside from the N/A and non-referential correlation is found between non-persistence and non-referential discourse functions (81%). There is also a correlation between referential uses and persistence (75%), but

1. Givón first introduces the notion of persistence in the Introduction to his 1983 volume, where he states that ‘There is no maximal value assigned by *definition...*’ (1983, p. 15). It is in his 1995 work that he specifies that persistence is ‘the number of times the referent recurs within the next (cataphoric) 10 clauses following its present occurrence’ (1995, p. 79).

25% of persistent NPs are still non-referential, and 19% of non-persistent NPs are referential in function. In the following sections, the relationships between other properties of NPs and discourse referentiality are examined in order to determine if there are stronger indicators of referential uses than persistence alone.

5.1.2 NP realization

In a study of Javanese conversation, Ewing (2005) found a correlation between the form of NPs and discourse referentiality. Specifically he found that the larger nominal expressions, such as headless relative clauses and full NPs, are more commonly associated with non-referential (non-tracking in his terms) referents, and that pronouns and evoked (unexpressed) mentions are more commonly associated with discourse referential NPs. Proper names tend to be non-referential, especially when used as vocatives, but may be tracking in other grammatical roles. Ewing’s findings for Javanese (all grammatical persons, referring expressions only; his Table 4.6, 2005, p. 150) are copied here for convenience in Table 6:

Table 6. Discourse Referentiality and Expression type in Javanese (Ewing 2005: 150)

Expression	Non-referential		Referential		Total	
	N	%	N	%	N	%
Headless Relative Clause	7	88	1	13	8	100
Lexical	100	65	55	35	155	100
Proper Name	51	76	16	24	67	100
Pronoun	10	15	55	85	65	100
Unexpressed	1	1	121	99	122	100
Total	169	41	248	59	417	100

The coding scheme used in the study of Ecuadorian Spanish is slightly different and consists of four main types of expressions (after consolidation): lexical, personal pronouns, other pronouns, and unexpressed mentions. Unlike Ewing’s data, two categories were used for pronominal forms of expression for these data: personal pronouns and others. The category “other pronouns” includes demonstrative and interrogative pronouns, as well as others (*ninguno* ‘none’, *todo* ‘all’, *alguien* ‘someone’, etc.). This two way distinction was made because these three types of “other pronouns” were found to primarily perform non-tracking discourse functions, as opposed to “personal pronouns”.

Prior to looking at the data, this decision to include the demonstratives with interrogative and other pronouns may seem counterintuitive. While there were

clear cases of demonstrative pronouns that served to track referents, the majority of the uses (77%, 76/99) were found to occur with a primarily non-tracking function, some of which are illustrated in (39). In this excerpt, A is asking her younger sister, C, about her course of study at the university and the functional polysemy of *eso* becomes apparent. The first two uses of *eso* seem to perform a deictic function of pointing back to *física*, or perhaps on a broader level point back to the idea of what A mistakenly believes C's course of study to look like. In terms of discourse-non-referential function, it can be viewed as non-tracking predicating, in Du Bois and Thompson's (1991) terms, in that it names a situation or activity (the course of study). Subsequent uses include *todo eso* ('all that') to close lists, and *eso sí* ('that's for sure' (lit. 'that yes')) to perform the interactional function of expressing agreement. These are also non-referential functions in that they fulfill a discourse function unrelated to tracking a referent, which is the primarily role of a referential NP.

(39) Non-referential uses of the demonstrative pronoun

- A: *y cuánto más te falta?*
 C: *.. falta=n .. cuatro más.*
 A: *.. [cuatro]?*
 C: *[hasta décimo] hacemos.*
 A: *oh=.*
 chévere,
 pero te gusta lo que estás siguiendo,
 que es lo importante.
 C: *.. sí,*
 sí me gusta,
 pero a veces también me aburro un poquito.
 A: *.. es que es pura fórmula,*
 pura matemática,
 pura --
 C: *no.*
 no es .. matemática,
 es a los primeros ... semestres [no más].
 A: *[pero yo he]*
 escuchado siempre a los que siguen sistemas,
 full matemáticas,
 ... quí- --
 y cómo --
 no,
 es física,
 cómo es?
 C: *... no.*
 para nada,
 → *eso es --*

- **eso** es los .. pr- --
el primer semestre,
nada más.
no.
es más= ... programación,
es más --
... desarrollo de sistema,
→ todo **eso**.
- A: .. hm=.
pero qué es lo que más te gusta,
o sea,
de todas las materias que tienes?
- C: ... (3.0) hm=,
.. qué te puedo decir?
programar.
- A: [programación]?
- C: [programar],
redes,
... hm=,
→ todo **eso**.
- A: pero chévere,
no?
.. porque imagínate,
ahorita para cualquier trabajo,
necesitan siempre alguien que sepa de [computación,
de --]
- C: [oh,
→ **eso** sí].
- A: de sistemas,
en todo.
- C: claro,
→ **eso** sí,
pero es de --
pero tienes que saber desde abrir una máquina,
hasta configurarle y todo.
- A: .. pero les enseña también a limpiar por dentro,
a hacer mantenimientos,
- C: [claro],
- A: → [2**todo eso**2]?
- A: 'and how much do you have left?
- C: .. there are still .. four (semesters) to go.
- A: .. [four]?
- C: [up until ten] is what we do.
- A: oh=.

- cool,
but you like what you're doing,
which is the important thing.
- C: .. yes,
yes I do,
but sometimes I get a little bored.
- A: .. it's just that it's pure formulas,
pure math,
pure --
- C: no.
it's not .. math,
(that's) the first ... semesters [is all].
- A: [but I have] always
heard from those who study software
engineering,
a lot of math,
... chem- --
and what --
no,
it's physics,
is that right?
- C: ... no.
not at all,
→ **that** is --
→ **that** is .. fir- --
the first semester,
and nothing more.
no.
it's more... programming,
it's more --
... system development,
→ all **that**.
- A: .. hm=.
but what is the part that you like the most,
I mean,
of all the subjects you have?
- C: ...(3.0) hm=,
.. what can I tell you?
programming.
- A: [programming]?
- C: [programming],
networks,
... hm=,
→ all **that**.
- A: but cool,

- no?
 .. because imagine,
 now for whatever job,
 they always need someone that knows [computation,
 of --]
- C: [oh,
 → **that's** for sure].
- A: of systems,
 in everything.
- C: right,
 → **that's** for sure,
 but um --
 but you have to know everything from opening up a
 machine,
 to configuring it and all.
- A: .. but it teaches them to clean the inside too,
 to do maintenance,
- C: [of course],
- A: → [**all that**]?' (Clases: 76–144)

Unlike Ewing's data, for this study no separate category was designated for proper names. Instead, these NPs were included in the lexical category. Other differences from Ewing's study are the inclusion of non-referring expressions and the exclusion of headless relative clauses. Despite some methodological differences, the results for both genres of Spanish, shown in Table 7, are similar to those Ewing obtained for Javanese. Ewing finds that both personal pronouns and unexpressed (Ewing's "evoked") mentions tend to be referential in Javanese, whereas in the Ecuadorian data personal pronouns are 84% referential and unexpressed mentions are 79% referential in function. The higher proportion of unexpressed non-referential mentions in the Spanish data can be attributed to the use of general (non-specific) third person plural, as in example (40), which is part of a predicating construction.

(40) Non-specific unexpressed third person plural²

F: ... (3.0) *y tú necesitas ir a --
 pasantía,*

2. It can be argued that non-specific third person plural forms could be excluded on the grounds that they function as impersonal statements and therefore have no subject; they were included in this study because as Lapidus and Otheguy (2005) point out, there are different degrees of nonspecificity, that is, in certain contexts, the subject does in fact refer to a set of possible referents or a designated group of people, and in some contexts, these 'impersonal subjects' are in fact expressed in Spanish.

y contactarte con un asesor.

A: [ya].

F: [para] ver cuántos negocios **Ø cierran** de cada
cuántas personas.

F: 'and you need to go to --
internship,
and get in contact with an assessor.

A: yeah.

F: to see how many deals **(they) close** out of each
number of people.' (Birthday:115–119)

Table 7. Discourse Referentiality and Expression type in both genres (N=4405)

Expression	Non-referential		Referential		Total	
	N	%	N	%	N	%
Lexical	1704	64	961	36	2665	100
Personal pronoun	83	16	426	84	509	100
Other pronouns	355	90	38	10	393	100
Unexpressed	177	21	661	79	838	100
Total	2319	53	2086	47	4405	100

The correlation between “other pronouns” and non-referential functions is very strong, with 90% of these pronouns coded as fulfilling non-referential functions. Personal pronouns show a tendency to be referential (84%); however it will be shown later that a stronger correlation is found when combining expression with specificity. For the remaining types of expression (lexical and unexpressed) the relationship between referentiality and form of expression is even less clear, so properties of personal pronouns, lexical and unexpressed mentions were further analyzed.

5.1.3 Modification

In their analysis of bare English-origin nouns in New Mexican Spanish, Torres Cacoullós and Aaron (2003) argue that non-referential (non-tracking) uses of NPs in Spanish correspond to lack of determination. Determiners here include all articles, quantifiers, demonstratives and possessive adjectives.

The results for presence of determiner (for lexical mentions, including NPs derived from adjectives only) and referentiality are displayed in Table 8.

Table 8. Discourse Referentiality and Determiner presence for Lexical mentions in both genres (N=2665)

Modifiers	Non-referential		Referential		Total	
	N	%	N	%	N	%
Bare ³	684	93	54	7	738	100
Determiner	1020	53	907	47	1927	100
Total	1704	64	961	36	2665	100

Chi-square⁴ = 363.988, *p* < 0.0001

Across the two genres bare NPs are almost categorically found to fulfill non-referential functions (see the predicating NP *miedo* in (41)), which is similar to Torres Cacoullos and Aaron’s (2003) findings. However, for those nouns occurring with determiners, approximately half are referential (see *la hoja* in (42), which refers to an entity present in the room) the other half non-referential (see (43), where the predicating NP *las escondidas* occurs in a fixed expression.⁵ In fact, more than half of the lexical NPs that were coded referential appear with a determiner, 60% (1020/1704). So while lack of determiner may be a good indicator that a NP is being used non-referentially, presence of determiner is not a good criterion for identifying referential uses. Modification thus cannot be used as a measure of referentiality, but lack of modification can.

- (41) Bare NP used non-referentially
- A: *no te tengo **miedo**,*
- A: ‘I’m not afraid of you’ (lit. I have no **fear** of you)
(Deportes:599)

3. This category includes some NPs co-occurring with a post-posed descriptive adjectives, e.g., *hacen cosas sucias* ‘they do **dirty things**’.

4. Chi-squared is the most commonly used statistical test in these data, as it tests for independence in the distribution of two different factors as related to groups or sets of data. It is applied to all 2x2 tables in these data but not to larger tables, as a significant difference with multiple factors is more likely but a Chi-squared test does not indicate where between the groups the significant difference lies.

5. All determiners were collapsed here for convenience. However, determiner type was also examined, showing contradictory patterns between the two genres (NPs with demonstratives tended to be non-tracking in narratives, yet tracking in conversation, and NPs with quantifiers were more commonly tracking in the narratives but non-tracking in conversations) or inconclusive patterns (NPs with definite articles, indefinite articles and possessive adjectives were closer to half referential, half non-referential).

(42) NP with determiner used referentially

R: *le diste la hoja?*R: 'did you give her **the paper?**' (Jugo:397)

(43) NP with determiner used non-referentially

A: .. (H) *jugando a las escondidas?*A: '.. (H) playing **hide-and-go-seek?**' (Fumar:220)

5.1.4 Specificity

Du Bois and Thompson (1991) note that NPs that perform generalizing and particularizing functions occur with both tracking and non-tracking functions, although there is a loose correlation between specificity (particularity) and tracking function, and generality and non-tracking function. This relationship was explored for the Ecuadorian Spanish data, and several patterns emerge. Table 9 shows the distribution of general and specific personal pronominal and unexpressed mentions and referential functions. Recall from Table 7 that in the aggregate, personal pronouns and unexpressed mentions mostly perform referential discourse functions (84% and 79%, respectively), which is not the case for lexical NPs, at 36%. General pronominal and unexpressed referents are very strongly associated with non-referential functions (87.5%, see (44) and (45)), while specific pronominal and unexpressed references are largely associated with referential functions (97%, see (46), where *ella* refers to the speaker's co-worker).

(44) General pronominal NP

J: *yo no sé a quien le jurabas,*J: 'I don't know who (**le**) you swore to,' (Jugo:221)

(45) General unexpressed NP

C: *pero creo que igual peligroso.*A: .. *cla=ro,*
ahora [Ø te ven no más],

C: [XXX]

A: *y Ø te pueden robar.*

C: 'but I think that (it's) just as dangerous.

A: .. of course,
now [**(they)** can see you],

C: [XXX]

A: and **(they)** can rob you.' (Clases:536–540)

(46) Specific pronominal and unexpressed NP

F: *una de mis compañeras tiene un hijito de unos*
*cuatro o cinco años.*A: *hm,*
ya.

F: entonces **ella** se llevó o=tro,
Ø ya no se llevó este Baby Mozart.
F: 'one of my colleagues has a little son about three
or four years old,
A: hm,
yeah.
F: so **she** took one,
(**she**) didn't take this Baby Mozart.' (Birthday:252-257)

Table 9. Discourse Referentiality and Specificity in both genres (among personal pronouns and unexpressed mentions; N=1347)

Specificity	Non-referential		Referential		Total	
	N	%	N	%	N	%
General	231	87.5	33	12.5	264	100
Specific	29	3	1054	97	1083	100
Total	260	19	1087	81	1347	100

Chi-square = 974.992, $p < 0.0001$

This distinction is less clear cut, however, for (non-bare) lexical mentions, as seen in Table 10. While the same pattern is seen for general referents (strongly non-referential), specific referents are much more divided (only 62% are referential). For this reason, (non-bare) lexical forms are further examined in the next section.

Table 10. Discourse Referentiality and Specificity in both genres (lexical NPs only; N=1927)

Specificity	Non-referential		Referential		Total	
	N	%	N	%	N	%
General	500	89	63	11	563	100
Specific	520	38	844	62	1364	100
Total	1020	53	907	47	1927	100

Chi-square = 408.916, $p < 0.0001$

5.1.5 Grammatical role

Thompson (1997) found a relationship between core arguments and tracking referents. Specifically, she found the majority of core arguments to be tracking, and the majority of obliques to be non-tracking. For this reason, the discourse

referentiality and grammatical role of lexical NPs (excluding bare nouns) are examined here in Table 11 and Table 12, where the genres are examined separately because the proportion of NPs performing referential functions in the O role is quite different between the two. Across all syntactic roles, there was a strong correlation between pronominal and unexpressed mentions regarding specificity (cf. Table 9). As Table 11 and Table 12 show for lexical mentions, specific NPs were largely referential and non-specific NPs were non-referential. It was therefore not necessary to look at grammatical role separately for these NPs. The data present similar results for the core vs. oblique distinction in Spanish in the narratives (95% of A, 82% of S, and 80% of objects are used referentially, with figures ranging from 23% to 69% for arguments in non-core roles), but the distinction is less obvious in the conversational data (87% of A, 69% of S, 52% of objects were used for non-referential functions, and figures for arguments in non-core roles ranging from zero (vocatives) to 66%). Indirect objects are included with core arguments here because they show similar properties to core arguments: they are often human, as are A and S (Du Bois 1987), whereas obliques are most often non-human (89% of objects of prepositions in the conversations [620/698] and the narratives [206/231] were non-human). They are collapsed with direct objects because they pattern similarly.⁶

Table 11. Discourse Referentiality and grammatical role in Spanish Narratives (lexical mentions only; N=634)

Role	Non-referential		Referential		Total	
	N	%	N	%	N	%
Core roles						
A	2	5	41	95	43	100
S	30	18	140	82	170	100
Objects (Direct and indirect)	38	20	151	80	189	100
Non-core						
Prepositional phrases, Adverbial phrases and Predicate nominal	150	77	45	23	195	100
“Other”	9	31	20	69	29	100
Extension of previous syntactic frame	5	63	3	38	8	100

6. For instance, all nine indirect objects in the narratives were found to be referential, and 79% (142/180) of the direct objects were found to be referential. In the conversational data, 51% of direct objects were non-referential (114/224) and 28% (5/18) of indirect objects were non-referential.

Total	234	37	400	63	634	100
-------	-----	----	-----	----	-----	-----

Table 12. Discourse Referentiality and grammatical role in Spanish Conversation (lexical mentions only; N=1293)

Role	Non-referential		Referential		Total	
	N	%	N	%	N	%
Core roles						
A	5	13	34	87	39	100
S	81	31	184	69	265	100
Objects (direct and indirect)	115	48	127	52	242	100
Non-core						
Prepositional phrases, Adverbial phrases and Predicate nominal	485	88	68	12	553	100
“Other”	73	47	81	53	154	100
Vocative ⁷	12	100	–	–	12	100
Extension of previous syntactic frame	15	54	13	66	28	100
Total	786	61	507	39	1293	100

In order to look for stronger trends in the Spanish conversational data, in which grammatical role did not pattern as expected, grammatical role and (non-)referential function were examined in conjunction with specificity. The results are given in Table 13 for the conversational data.

We can see from Table 13 that when specificity and grammatical role are combined for the conversational data, the patterns are much stronger than when either specificity or grammatical role is examined alone.⁸ A more clear core-oblique distinction is observed among the NPs that refer to specific entities, with the core arguments correlating with referential discourse functions and NPs in non-core roles strongly associated with non-referential functions. There is no such distinc-

7. Vocatives are categorically non-referential; while the NPs in vocatives are often referring (i.e., they refer to a real world entity), functionally they are not referential.

8. A Chi-square test of the totals for general/specific and referential/non-referential shows a significant difference (Chi-square = 446.886, $p < 0.0001$).

Table 13. Specificity, Grammatical role and Discourse Referentiality of lexical (non-bare) NPs in Spanish Conversation (N=1293)

Specificity	Role	Non-referential		Referential		Total	
		N	%	N	%	N	%
General	A	4	50	4	50	8	100
	S	57	70	25	30	82	100
	Objects (direct and indirect)	85	88	12	12	97	100
	Prep. phrases, Adv. phrases and Predicate nominal	231	97	8	3	238	100
	“Other”	43	81	10	19	53	100
	Extension of previous syntactic frame	10	91	1	9	11	100
General total		430	88	60	12	490	100
Specific	A	1	3	30	97	31	100
	S	24	13	159	87	183	100
	Objects (direct and indirect)	30	21	115	79	145	100
	Prep. phrases, Adv. phrases and Predicate nominals	254	81	60	19	314	100
	“Other”	30	30	71	70	101	100
	Extension of previous syntactic frame	5	29	12	71	17	100
Specific total	Vocative	12	100	–	–	12	100
		356	44	447	56	803	100
Grand total		786	61	507	39	1293	100

tion for the general NPs, which pattern the same (non-referential) regardless of core or oblique status. Recall that when objects were examined independently of specificity, nearly two thirds were non-referential and one third referential. When specificity is taken into consideration, 88% of general objects are found to be used for non-referential functions (see the predicating NP *la atención* in (47)) and 79% of specific objects are found to be used for tracking referents (see *el sombrero* in (48)). This is a very important finding and supports the approach taken here that looking at factors in conjunction with one another yields better results than look-

ing at factors independently. Overall, 31% of S tokens were coded as fulfilling non-referential functions and 69% were coded referential, but 70% of general NPs in the S role fulfill non-referential functions, and 87% of specific S are used for referential functions (see the predicating NP *el tiempo* in (49) and the referent *la Gracie* (50)). Specific A is very strongly associated with referential uses (97%), whereas for general A the tokens are only 50% non-referential in function. This is likely due in part to the extremely low token count (only 8 general, lexical non-bare A), and the fact that lexical A itself is somewhat unusual, given the “non-Lexical A Constraint” (Du Bois 1987:829).

- (47) General, non-referential object
.. lo= llamaron la atención con un .. silbido,
'.. they called his attention with a .. whistle,' (PS43:88)
- (48) Specific, referential object
y Ø le botó el sombrero.
'and (she) knocked off his hat.' (PS29:42)
- (49) General, non-referential S
A: *(H) qué rápido que pasa el tiempo,*
A: *'(H) how quickly time passes,'* (Clases:70)
- (50) Specific, referential S
R: *.. solo la Gracie viene.*
R: *'.. only Gracie is coming.'* (Jugo:518)

Now that NPs in core roles (A, S and O), vocatives, objects of prepositions and adverbial phrases have been accounted for, we are left with NPs in the role “Other” and “extension of previous syntactic frame”. Interestingly, NPs in the “Other” syntactic role appear to pattern more like core arguments, with the majority of general “Other” referents being associated with non-referential functions (81%) and the majority of the specific “Other” NPs being associated with referential functions (70%). This is an important finding because these NPs have often been left out of analyses, and when they have been studied it has been largely in isolation rather than in terms of how they might pattern like other NPs. “Extension of a previous syntactic frame” NPs pattern like those in the “Other” syntactic role in the conversations, but as will be seen below, do not follow the same patterns in the narratives.

The core-oblique split related to specificity observed in Table 12 for conversational data is also seen in the narrative data, shown in Table 14. Specific NPs in the core argument roles (87–98%) and “Other” role (74%) are strongly correlated with a referential function; and the majority of the non-core specific arguments are still associated with non-referential uses (71%). Again, all general NPs have a very strong tendency across syntactic roles to be used non-referentially (100%

in all roles except for objects, where 86% are non-referential). The notable differences between the genres are (a) a higher proportion of specific NPs that are used referentially in the narratives (71% of specific NPs in the narratives are found to have referential functions, versus 56% in the conversations), (b) general NPs are over three times as common in the conversational data (38%, 490/1293) than in the narrative data (12%, 73/634), and (c) the different patterning of NPs in the “extension of a previous syntactic frame” role.

Table 14. Specificity, Grammatical role and Discourse Referentiality of full NPs in Spanish Narratives (N=634)

Specificity	Role	Non-referential		Referential		Total	
		N	%	N	%	N	%
General	A	1	100	–	–	1	100
	S	9	100	–	–	9	100
	Objects (direct and indirect)	19	86	3	14	22	100
	Prep. phrases, Adv. phrases and Predicate nominals	39	100	–	–	39	100
	“Other”	2	100	–	–	2	100
	Extension of previous syntactic frame	–	–	–	–	–	–
General total		70	96	3	4	73	100
Specific	A	1	2	41	98	42	100
	S	21	13	140	87	161	100
	Objects (direct and indirect)	19	11	148	89	167	100
	Prep. phrases, Adv. phrases and Predicate nominals	111	71	45	29	156	100
	“Other”	7	26	20	74	27	100
	Extension of previous syntactic frame	5	63	3	37	8	100
Specific total		164	29	397	71	561	100
Grand total		234	37	400	63	634	100

5.2 Summary of grammatical manifestations of discourse referentiality

When referentiality is viewed in terms of discourse function, these data show that a combination of factors (realization, specificity and syntactic role) provides a way

to measure or predict discourse referentiality of a given expression. In addition to highlighting the complexity of the category of referentiality, this marks a significant advancement toward being able to code for referentiality in a way that best captures all of the associated features, instead of a narrower approach in which all tokens are expected to correspond to one associated feature, an approach which fails to capture the nuances of referentiality. For some tokens, realization alone proved to be a sufficient criterion – all expressions realized as demonstrative, interrogative and other pronouns were strongly found to correlate with non-referential functions. For others, it was a combination of realization and specificity – all specific personal pronouns and unexpressed mentions were strongly associated with referential uses, all general personal pronouns and unexpressed mentions were strongly correlated with non-referential functions. For the remaining lexical mentions, it was a combination of modification (or more precisely, lack of determiner), syntactic role and specificity that showed the strongest correlations. Bare NP were most frequently found to display non-referential functions. For the remaining non-bare lexical NPs, specific lexical mentions as core arguments or in the syntactic role “Other” were strongly employed referentially, specific lexical mentions that were not core arguments were largely non-referential in use, and all general lexical mentions were very strongly oriented toward non-referential functions. Note that the only specific, non-referential NPs are those that are not core arguments or in the syntactic role “Other”.

The NPs in the grammatical role “extension of previous syntactic frame” did not reveal clear patterns with respect to referential uses. For this reason, it was decided to recode their grammatical role according to the role they performed in the previous syntactic frame, for example, in example (24), repeated here as example (51) for convenience, *una parte* was coded as general and as a direct object. Recall from 4.2.2 that the specificity of an expression (specific or general) was coded according to whether or not an expression named a specific entity. In this example, *una parte* is used to talk about the activity of watching a part (as opposed to a whole) of a game, and for this reason was coded as general. In the first instance, this expression is found in a DO role (*vi una parte*), and the second instance it is found as an extension of this frame. For this reason, the NP in *de ahí una parte en el aeropuerto* and is recoded as a non-referential expression, as the general pattern for non-referential NPs was found to perform non-referential functions.

(51) Extension of previous syntactic frame

S: .. *vi una parte*,
 en mi casa,
 F: *ya*.

S: *de ahí una parte en el aeropuerto,*
S: *'... I saw part [of the soccer game],*
at my house,
F: *yeah.*
S: *and then another part at the airport,' (Deportes:83–86)*

Based on the patterns that were observed in this chapter, for the remainder of the study, the data were recoded for referentiality according to the table presented in Table 15.

Table 15. Summary of coding for discourse referentiality

Discourse Referential	Discourse Non-referential
A subset of non-bare lexical NPs as described below	Bare lexical NPs
Specific subjects realized as (non-bare) lexical NPs, personal pronouns and unexpressed mentions	All other subjects
Specific objects (D and IO) realized as (non-bare) lexical NPs or personal pronouns	All other objects
Specific (non-bare) NPs in syntactic role “Other”	Predicate nominals
Extension of previous (referential) syntactic frame	Objects of prepositions
	Adverbials
	Vocatives
	General NPs in syntactic role “Other”
	Grammatical words (<i>qué, cuánto</i> , etc.)
	Extension of previous (non-referential) syntactic frame

It is through the identification of the factors that correlate with discourse referential functions that the tokens in this study can be re-coded for referentiality in an objective and replicable way. This is relevant to the present study in the following chapters, which explore the relationships between form and function as related to referential uses (and how they relate to notions of activation and identifiability, etc.) separate from the non-referential uses. This may also be useful in future studies because it allows other scholars to code for referentiality in the same way in order to more readily compare across investigations. It builds upon previous work in which many of these factors were associated with referentiality and lays a framework for future investigations that need to account for referential and non-referential expressions separately.

Linguistic patterning of referents

Referents are often what we think of as prototypical nouns – the characters in a story, the parts of the landscape that paint the setting, or a topic of conversation such as the new car or house someone has purchased. In this chapter, linguistic coding of referents is explored with respect to the relationships between information flow, form and syntactic role in order to account for the cognitive and other pressures that influence the form and distributions of referents in each genre. The set of linguistic expressions examined here are those summarized in Table 15, e.g., all specific non-bare lexical and personal pronominal and unexpressed referents in core roles and specific non-bare referents in the syntactic role “Other”. As anticipated, the data show many patterns that are directly related to information flow, yet there are many notable exceptions, which are discussed in detail. New and given referents show significantly different morphosyntactic patterns, which are then further examined. The sections on given referents are of particular significance to theories of anaphora.

6.1 Formal characteristics of discourse referents

As outlined in Chapter 4, all discourse referential tokens were assigned an information status code of either new (to the discourse) or given (previously mentioned within the transcript), as opposed to non-referential expressions, for which information status does not apply. In addition, according to the coding scheme in this study, they are also all specific. The aims of this section are to examine and contrast the forms that new and given referents take (realization and definiteness), what syntactic roles are most conducive to presenting new referents in Spanish, and which syntactic roles are productive for given referents.

6.1.1 Discourse referentiality and realization

6.1.1.1 *New referents and realization*

Unsurprisingly, as Table 16 shows, new referents are introduced almost categorically as full NPs (95% in the narratives, 91% in the conversations, the difference between the genres is not statistically significant). Of the few that were coded pronominally, they are either highly marked (and may be analyzed by some as full NPs, or redundant (see (52)), and the unexpressed first mentions are highly inferable from the context. These figures are similar to what Green (2012, p. 49) reported for four genres of written Australian English, with 4.3% of personal pronouns conveying new information. In these data, 7.7% (17/221) conversational personal pronouns are used for first mentions, and in the narratives 3.1% (5/160). The genre difference supports the assertion that there is a greater degree of speaker confidence regarding shared information in the conversations than in other genres.

Table 16. Discourse new referents and realization (N = 428)

	Narratives		Conversation	
	N	%	N	%
Full NPs	123	95	270	91
Pronouns	5	4	17	6
Unexpressed ¹	2	1	11	4
Total	130	100	298	100

In the narratives, all five pronominal discourse new referents were part of the double object construction or “redundant pronouns” (cf. Butt & Benjamin, 2004), in which the referent is also referred to with a full NP, as in (52). This structure is also common in the conversations (accounting for 10/17 new pronouns). Another five in the conversation were modified pronouns, as in (53), and two were “numeric pronouns” (which may or may not be modified), as in (54). The remaining one was part of a joke in which the speaker made a play on words. Note that subject personal pronouns (*él, ella, ellos*, etc.) were not found to introduce new referents, which is not surprising.

1. This only refers to unexpressed subjects, as unexpressed objects were not included (cf. Chapter 4). Subjects = 44% (537/1228) of all narrative tokens and 32% (1026/3190) of conversational NPs. Of these, 54% (291/537) are unexpressed in the narratives and 52% (538/1026) in the conversations.

(52) Redundant pronoun²

Ø **le** encuentra a **unos chicos**,
 '(he) (**le**) meets **some boys**,

(PS42: 37)

(53) Modified pronoun³

R: y **los de la basura** nunca arreglaron la puerta,

R: 'and **the ones from the garbage** (i.e., the garbage
 men) never fixed the door,

(Jugo: 143)

(54) New, numeric pronoun⁴

A: (H) *ahí está una de mis mejores amigas*,

A: '(H) there is one of my best friends,' (Deportes:125)

2. Co-referential clitic pronouns and full NPs were coded as two distinct mentions to capture the broad range of uses of the pronouns. In many contexts, as in this example, the use of the pronoun is variable. For example, in the conversational data, the following example is found, also with a specific, human direct object, yet the redundant pronoun is not found.

(i) *viste el Guillermo?*
 'did you see Guillermo?'

(Jugo:237)

In this example, the accusative *a* is also lacking, and it is found that the use of accusative *a* is also variable when the direct object co-occurs with the pronoun *le*, as in:

(ii) *y Ø le encuentro ahí la Carolina*,
 'and (I) (*le*) find there la Carolina,'

(Fumar:290)

Note that in (52) there is a lack of agreement between the pronoun and the co-referential full NP (*le* is singular, *unos chicos* is plural), which suggests that the pronoun is not just a copy of the full NP, or an agreement marker, rather it has a separate function, perhaps to signal to the interlocutor that new information will be presented. This use is more common with new indirect objects (for the conversations, 50% [4/8] new IOs are accompanied by a pronoun) than with direct objects (10% [11/110] conversational DOs and 9% [4/44] new DOs in the narratives were accompanied by a pronoun, there were no new IOs in this genre).

Regarding the lack of agreement, an anonymous reviewer suggests that this is not unusual for speakers of Spanish, although it was found infrequently in these data (lack of agreement was found in 2/162 pronominal objects [direct and indirect] in the conversations and 4/110 in the narratives). As noted earlier, the use of *le* for a direct object is typical of this dialect (and others).

3. *Los* was coded as a pronoun, however it could also be considered to be the definite article standing alone (i.e., not modifying a head noun). For the purpose of this analysis it is considered a pronoun, albeit an infrequent form, which shows that pronouns that introduce new information are different from pronouns that signal given information.

4. All cases of *uno/una/unos/unas* that did not modify a NP as an indefinite article were coded as pronouns in this study. These cases could be argued to be part of a morphological scale between a determiner and a pronoun, much like the head noun in (53) could be argued to be a pronoun or a determiner.

Both of the unexpressed mentions of new referents in the narratives were produced by the same speaker, who may have been struggling with the high information pressures of introducing so many new referents into the narrative or perhaps with the task of narrating the story itself. One of these is illustrated below, in (55). In this example, taken from the beginning portion of the transcript (the first 15 lines include eight IUs of instructions from the investigator and a description of the scenery present at the beginning of the film), the speaker first introduces the man who picks pears. He then goes on in the twelfth line to refer to the young man who walks past with a goat (*ø estaba halando una oveja* '(he) was pulling a sheep'). In line 12 of this example and again in lines 16 (*.. ø estaba halando por ahí* '(he) was pulling over there') and 17 (*ø asoma* '(he) appears') the subject is unexpressed. It is not until line 18, a full 6 IUs after the first mention, that the referent is realized as a full NP (*y asoma otro joven* 'and another young man appears').

- (55) Unexpressed first mention in narrative
- 1 ... (H) *ahí asoma un campesino,*
 - 2 ... (2.0) (H) *que está subiéndose la escalera,*
 - 3 .. *a cosechar la fruta.*
 - 4 ... *esa fruta es= .. peras.*
 - 5 ... (3.0) *entonces,*
 - 6 *ø está cosechando las frutas,*
 - 7 *ya ø cogió una canasta muy= ... grande.*
 - 8 *la canasta es grande.*
 - 9 .. *y ø está cosechando,*
 - 10 *ya ø cogió dos,*
 - 11 .. (H) *asoma --*
 - 12 → .. *ø estaba halando una oveja.*
 - 13 ... *o un chivo.*
 - 14 .. *que es de la misma .. raza,*
 - 15 *no?*
 - 16 → (H) .. *ø estaba halando por ahí,*
 - 17 → *asoma ø,*
 - 18 → *y asoma otro joven,*
 - 19 ... *otro campesino ahí asoma con la oveja.*
-
- 1 '... (H) and a peasant appears,
 - 2 ... (2.0) (H) that is climbing up the ladder,
 - 3 .. to harvest the fruit.
 - 4 ... that fruit is= .. pears.
 - 5 ... (3.0) so,
 - 6 (he) is harvesting the fruits,
 - 7 (he) already got a really big basket.

- 8 the basket is big.
 9 .. and (he) is harvesting,
 10 (he) already got two,
 11 .. (H) appears --
 12 .. (**he**) was pulling a sheep.
 13 ... or a goat.
 14 .. that is the same .. species,
 15 no?
 16 (H) .. (he) was pulling over there,
 17 (he) appears,
 18 and **another young man** appears,
 19 ... another peasant appears with the sheep.'
- (PS52: 16–24)

For the conversational data, in all the instances of new referents being “introduced” as unexpressed mentions, the referent is clearly inferable from the context. Interestingly, of the 11 tokens of new unexpressed mentions, six refer to someone’s parents, as in (56). Note in the example that using an unexpressed mention for the first reference to someone’s parents does not cause a problem, as evidence by Rodrigo’s response in (56).

(56) Inferable unexpressed mention

- A: *mañana=na no va a haber nadie,*
 Ø se van a Cuenca.
 R: *.. pero no Ø saben que Ø van a hacer la fiesta?*
 A: *sí sabe la mamá.*
 A: *`tomorrow no one is going to be there,*
 (they) are going to Cuenca.
 R: *.. but (they) don’t know that (they’re) going to*
 throw the party?
 A: *yeah the mom knows.'*
- (Fumar: 431–434)

In summary, the introduction of discourse new referents in both genres is usually done with a full NP but there are exceptions to this pattern. When exceptions occur, the surrounding discourse provides clues as to why the speakers may have chosen a lesser form (i.e., highly inferable referent, high information flow pressures, etc.).

6.1.1.2 Given referents and realization

Given referents, by contrast, have been mentioned previously in the discourse, so it follows that they are, to some degree, topical. It is often claimed that continuous/accessible referents will receive less grammatical coding than their newer, discontinuous/inaccessible topics (cf. Givón, 1981).

Table 17. Given referents and realization (N = 1379)⁵

	Narratives		Conversation	
	N	%	N	%
Full NPs	219	33	177	24
Pronouns	155	24	204	28
Unexpressed	282	43	342	47
Total	656	100	723	100

It is clear, however, from Table 17, that a significant proportion of given referents (33% in narratives and 24% in conversation) receive substantial grammatical coding (i.e., full NPs).⁶ Givón's (1981) claim is partially true; when the distribution of given referents expressed as pronouns and unexpressed mentions is compared with that of new referents (cf. Table 16), it is clear that a much higher proportion of given referents receive less grammatical coding than new referents. Yet, while the number of given referents which are encoded as a full NP is much smaller than the proportion of first mentions encoded as full NPs (in the narratives, 33% given vs. 95% new, and in the conversations, 24% vs. 91%, or approximately one quarter to one third in each), it is still higher than what may be predicted by theories of information flow and referent realization. It was hypothesized that a combination of referential distance from previous distance (as a measure of accessibility) and activation state (operationalized as first or non-first mention) might yield more information about the role information flow plays in referent realization. In order to determine if these given NPs are being reintroduced into the discourse after a period of time and are therefore being treated as "new" again, the distance from the previous coreferential mention was measured in clauses, and the results are presented in Table 18 (note that figures in this Table represent a previous mention in all syntactic roles, not just subjects).

5. The counts for given referents do not include referents that were mentioned previously as a group ('member of a set') or those plural referents whose members were previously introduced individually. These referents are examined separately below.

6. The proportion of given referents which are full NPs in the narratives is significantly higher than the number of given referents that are full NPs in the conversations (Chi-square = 12.886, $p < .01$).

Table 18. Distance in clauses from previous mention for given full NPs (N=394)⁷

	Narratives		Conversation	
	N	%	N	%
0	67	31	75	43
1–2	54	25	41	23
3–4	22	10	15	9
5–10 clauses	37	17	14	8
10+	39	18	30	17
Total	219	100	175	100

As Table 18 shows, NPs with co-referential mentions in the previous three clauses (i.e., 0–2 clauses) account for a large portion (56% in narratives, and 66% in conversation) of the given referents that are realized as full NPs.⁸ Even more surprising is that nearly half (43%) of the given full NPs in conversations were produced in the following clause, as in (57), and approximately a quarter (23%) at a distance of 1–2 clauses, as in (58). In (57) it is the same speaker who produces the same full NP in two consecutive clauses; it appears as though the emphasis as evidenced by the use of *mismo* “herself” may have motivated the use of the expression (versus unexpressed) for the second mention. In (58), three speakers are collaboratively narrating an event and additional factors such as repetition as and narrative structure techniques to signal change in setting show clear influence on the form of the NP *Martín*. It is interesting that two speakers (Rodrigo and Gerardo) produce identical IUs almost simultaneously, both of which contain a given full NP, even though the previous mention was realized as a full NP at a distance of just one clause away. Clearly, distance from previous mention, and by extension, topic accessibility/continuity is not the only factor related to the realization of given referents (as noted in Fox, 1987). Additional factors that contribute to referent realization not predicted by information flow certainly merit further investigation.

7. Two referents were excluded from the conversational data because they were third person references to participants in the conversation and the previous realization was first or second person.

8. It is interesting that these data show higher topic continuity for 3rd person subjects in the conversations; this is contrary to what Travis (2007) found for 1st person singular subjects in a comparison of conversational and narrative data. This may be an artifact of the Pear Stories, which have five salient human actors, as opposed to the narratives data in Travis’ study, in the stories are mostly told from a first person point of view.

- (57) Full NP followed by another full NP with no intervening clauses
 R: .. pero no $\emptyset$ saben que $\emptyset$ van a hacer la fiesta?
 A: sí sabe **la mamá**.
 ... (2.0) **la mamá mismo** le dijo.
 R: '... but (they) don't know that (they're) going to throw the party?
 A: yeah **the mom** knows.
 ... (2.0) **the mom herself** told her.' (Fumar: 433–435)
- (58) Given full NP followed by another given full NP at a distance of 1 clause
 G: no=,
 el **Martín** no más.
 ese día estábamos aquí viendo el fútbol,
 pues.
 R: ajá,
 y justo [llegó el **Martín**].
 G: [y justo llegó] el **Martín**,
 que venía de la --
 M: .. de la fiesta.
 así fue.
 G: 'no=,
 just **Martín**.
 that day we were here watching soccer,
 anyway.
 R: aha,
 and [**Martín** arrived just then].
 G: [and **Martín** arrived] just then,
 he was coming from the --
 M: .. from the party.
 that's what happened.' (Café: 1137–1146)

A number of studies (Cameron, 1994; Flores-Ferrán, 2002; Travis, 2005, 2007) have found that priming, or parallel syntactic structures, exhibits a consistent influence upon variable first and second person subject expression in Spanish. Priming effects for realization of third person referents have been less studied in Spanish (however, see Cameron, 1994, and Flores Ferrán, 2002, for analyses that include third person), and is perhaps more complicated, as third person referent realization has three forms (full NPs, pronouns and unexpressed) and is subject to different constraints (particularly information flow) than first or second person subjects. The distribution of the realization of all given full referents (not just subjects) by form of previous mention is given in Table 19. As Table 19 shows for the conversations, when there is a lexical mention of the referent in the immediately preceding clause (cf. (57)), the majority of full NPs were also previously realized

as a full NP (62%). Pronouns and unexpressed mentions also show a tendency toward mirroring the same previous realization (45% of pronominal given mentions that were referred to in the previous clause were also realized pronominally in the previous mention, and 50% of unexpressed mentions that were mentioned in the previous clause were also realized as unexpressed mentions in the previous clause). The trend goes beyond just the immediately preceding clause for lexical mentions and unexpressed mentions – regardless of distance, if the previous mention was a full NP, a subsequent full NP is always preferred, and if the previous mention of the referent was unexpressed, it is also likely that the subsequent mention will be unexpressed, even at a distance of 3–4 clauses. Table 20 shows similar trends in the narratives, where there is no distance between two mentions of the same referent, 58% of full NPs were previously realized as a full NP, 37% of pronominal mentions were previously realized pronominally, and 59% of unexpressed mentions were previously unexpressed.

Table 19. Realization of Given NP, Distance in Clause from Previous Mention and realization of previous mention in conversations (N=703)⁹

Realization	Distance in Clauses	PM Full NP		PM Pronoun		PM Unexpressed		Totals
		N	%	N	%	N	%	N
Full NP	0	46	62	15	20	13	18	74
	1–2	29	71	3	7	9	22	41
	3–4	13	87	–	–	2	13	15
	5–10	9	64	1	7	4	29	14
	10+	19	63	5	17	6	20	30
Total		116	67	24	14	34	20	174
Pronoun	0	34	28	55	45	32	26	121
	1–2	10	24	13	31	19	45	42
	3–4	5	33	4	27	6	40	15
	5–10	2	22	5	56	2	22	9
	10+	5	50	1	10	4	40	10
Total		56	28	78	40	63	32	197

(Continued)

9. 20 NPs were excluded because either the NP referred to a participant in the conversation (3rd person mention, not a vocative), therefore coding for previous mention is irrelevant as the previous mention was either first or second person, or because the recorded portion of the conversation began after the referent had already been introduced, and the previous mention was therefore absent from the transcript, i.e., if target occurred in first 10 clauses.

Table 19. (continued) Realization of Given NP, Distance in Clause from Previous Mention and realization of previous mention in conversations (N=703)

Realization	Distance in Clauses	PM Full NP		PM Pronoun		PM Unexpressed		Totals
		N	%	N	%	N	%	N
Unexpressed	0	76	36	30	14	108	50	214
	1–2	27	34	17	21	36	45	80
	3–4	3	14	4	19	14	67	21
	5–10	3	43	1	14	3	43	7
	10+	2	20	3	30	5	50	10
Total		111	33	55	17	166	50	332
Grand total		283	40	157	22	263	37	(703)

In addition, in nearly all cases, regardless of form of preceding mention, the greater the distance between two mentions, the more likely a referent is to be realized as a full NP. However, even when there is a distance of more than ten clauses, there is still variation in form of expression. The figures in Table 19 and Table 20 suggest that there is a lexical priming effect for full NPs, but can neither confirm nor deny that structural priming is a relevant factor influencing third person referent realization in Spanish. In order to consider structural priming, a more comprehensive analysis that separates the different syntactic roles in order to account separately for unexpressed mentions, which here only refer to subjects, would have to be undertaken, which is beyond the scope of the current study. However, the figures in the above tables do indicate that information flow alone, particularly the notion that given and topical referents (here operationalized as those which have been previously mentioned in recent discourse) receive lesser grammatical coding, is not able to explain all instances of referent realization. However, interactional reasons often explain the coding. In the following six examples of such referents, the role of factors beyond information flow are discussed.

Repetition of a full NP is a recurrent type of “unexpected full NP” in the data. Full NPs that are repeated for interactional purposes, as *un rico cafecito* in (59) show the same realization. This repetition occurs after a pause and immediately precedes a shift in topic, suggesting that the repetition is functioning to avoid a lull in the conversation, indicate participatory listenership (Tannen, 2007), or perhaps as an acceptance of a contribution (cf. Clark & Schaefer, 1989).

Table 20. Realization of Given NP, Distance in Clause from Previous Mention and realization of previous mention in narratives (N=642)¹⁰

Realization	Distance in Clauses	PM Full NP		PM Pronoun		PM Unexpressed		Totals
		N	%	N	%	N	%	N
Full NP	0	39	58	14	21	14	21	67
	1–2	32	63	10	20	9	18	51
	3–4	12	57	1	5	8	38	21
	5–10	26	70	7	19	4	11	37
	10+	31	82	3	8	4	11	38
Total		140	65	35	16	39	18	214
Pronoun	0	30	31	36	37	32	33	98
	1–2	12	30	9	22	19	48	40
	3–4	4	50	3	38	1	13	8
	5–10	1	33	1	33	1	33	3
	10+	1	100	–	–	–	–	1
Total		48	32	49	33	53	35	150
Unexpressed	0	60	27	30	14	130	59	220
	1–2	22	45	6	12	21	43	49
	3–4	2	33	1	17	3	50	6
	5–10	1	33	1	33	1	33	3
	10+	–	–	–	–	–	–	–
Total		85	31	38	14	155	56	278
Grand total		273	42	122	19	247	38	642

(59) Full NP in repetition.

M: *vamos a servirnos un rico café [ahora] que nos hace frío.*

G: [XX]

un rico cafecito.

M: *.. un rico cafecito.*

G: *y el Ricardo no viene?*

M: *pueda que ø ya venga.*

M: 'let's have a nice cup of coffee [now] that it's cold.

10. Fourteen tokens were excluded because the previous realization was truncated

- G: [XX]
a nice little cup of coffee.
 M: **a nice little cup of coffee.**
 G: and Ricardo isn't coming?
 M: it's still possible that (he) will come.'
 (Café:386-391)

Full given NPs also occur if the speaker feels that the referent was not activated the first time it was mentioned, as occurs when speakers' turns overlap with each other. In (60), Rodrigo and Adrián are jointly constructing a list of their childhood friends from the neighborhood. Their turns frequently overlap in this excerpt precisely where friends are mentioned for the first time. As a result, several names (Eloy, Darwin, and *un orejón* ("a big-eared kid")) are repeated as they may not have been activated the first time they were mentioned. Note that it is the same speaker that produces the full NP twice. The NP *Favio* was also produced in overlap, but here Rodrigo repeats the NP, acknowledging that he did in fact hear the name and is was activated as part of the list.

(60) Full given NP due to overlap

- R: ... *pero quién más vivía aquí?*
 .. *habíamos --*
éramos bastantes,
de chiquitos.
 A: *el [Eloy].*
 R: *[es que el Darwin] también vivía aquí.*
 A: **el Eloy.**
 R: .. **el Darwin,**
nos llevábamos con el Darwin.
...(2.0) el Juan Carlos.
.. vivía el Edwin,
el Flaco.
...(2.0) (H) también [salía]--
 A: *[el Favio].*
 R: **el Favio.**
y salía el Santiago,
hermano de la Solange,
también [jugaba] con nosotros.
 A: *[ah].*
.. también --
...(2.0) también= .. salían a jugar --
sabía jugar el Mauro,
el Ricardo,
 R: .. *[claro].*

- A: [el hermano] del Darwin --
[2del Mario2].
- R: [2y había un orejón2],
y --
.. **un orejón había,**
- R: '... but who else lived here?
.. we were --
there were a lot of us,
little kids.
- A: [Eloy].
- R: [it's that Darwin] also lived here.
- A: **el Eloy.**
- R: .. **Darwin,**
we hung out with Darwin.
...(2.0) Juan Carlos.
.. Edwin lived here,
Flaco.
...(2.0) (H) also [there was]--
- A: [Favio].
- R: **el Favio.**
and Santiago would come out
Solange's brother,
also [played] with us.
- A: [ah].
.. also --
...(2.0) also .. would play --
Mauro would play,
Ricardo,
- R: .. [of course].
- A: [the brother] of Darwin --
[2of Mario2].
- R: [2and there was a big-eared kid2],
and --
.. **there was a big-eared kid,'** (Fumar:340-369)

It has been suggested (Givón, 1983; Hochberg, 1986) that subjects may be expressed in order to avoid ambiguity that arises when there are two semantically compatible nouns in the context that could potentially compete for the subject role. The full NP *Pauli* in the last line of (61) does appear to fulfill this function. In this example, there are two third person feminine animate referents (Natalie and Pauli) which could both potentially be the subject of the verb *había dicho* ('had told'). In this case, pronominal or unexpressed realization could result in confusion, so the choice of a full NP provides a clear benefit for referent tracking

in this instance. In addition, note that the immediately preceding mentions of the referent were as indirect object (*le, a la Pauli*) and object of a preposition (*con la Pauli*); work by Paredes Silva (1993) on Brazilian Portuguese has indicated that change in syntactic role may influence subject expression. It is also possible that contrast, as indicated in the use of the conjunction *pero* (cf. Silva-Corvalán 2001, p. 154), is a contributing factor to the coding of the NP as full. Note that the change in setting may also influence the referent form.

(61) Given full NP in different syntactic roles

M: ... (2.0) (TSK) *la Natalie debería hablar con la Pauli.*

P: ... (H) *sabe que sí le ha dicho,*

creo,

alguna vez ella a la Pauli,

verá.

.. (H) *pero **la Pauli** le había dicho que ya no hay*

como,

M: '... (2.0) (TSK) *Natalie should talk with Pauli.*

P: ... (H) *you know she has said something to her,*

I think,

one time she (did) to Pauli,

see.

.. (H) *but **Pauli** told her that there is no way,'*

(Café:707–712)

While some subjects may be expressed in order to resolve ambiguity, this is not always the underlying motivation behind the expression given referent as a full NP, as evidenced in the following examples. In (62), the NP *el Flaco* occurs as a full NP upon first mention and again as a full NP two clauses later. In this example, as in the previous, the NP does change syntactic role, but it appears that the change of conversational action (i.e., listing all the people who have left to talking about missing one of those people) is more influential in the choice to express the referent as a full NP. This instance of realization of *el Flaco* as a full NP could also be partly attributed to potential inference with the semantically compatible NP *el Edwin*, who also left, and is therefore a candidate for being missed. The significance of repetition of a form in establishing structural boundaries has also been noted by Fox (1987) and Tannen (2007).

(62) Full given NP shift in conversational action

R: .. *y=,*

los demás=s,

.. ya no hay muchos,

pues,

- sólo el Mario y yo quedamos de los .. que vivíamos
en la célula.
- E: .. de ley.
[y qué es del Edwin]?
- R: [porque el Edwin ya] se fue=,
el Flaco se fue=,
.. todos se [fueron].
- E: [oh=],
yo le extraño **al Fla=co**.
.. ya quisiera que venga.
- R: `.. and,
everyone else,
.. there aren't many anymore,
well,
just me and Mario remain of those .. that lived on
the block.
- E: .. for sure.
[and what about Edwin]?
- R: [because Edwin already] left,
Flaco left,
.. everyone [left].
- E: [oh=],
I miss **Flaco**.
.. I wish he could come back.' (Viejos Amigos:154–166)

Example (63) shows a different pattern than the previous examples, in which the referent was expressed in precisely the same form as the previous mention. Rather than a repetition of a full NP, this example shows the introduction of a referent with a full NP (*tu ex pelada* 'your ex-girlfriend'), followed by five unexpressed subject mentions and two mentions as an object, then a return to a full, but different, NP, *la man* ('the chick'). The reason for the switch is not entirely clear, but it likely involves subjectivity/expression of stance. On the one hand, it could be argued that the speaker expressed the subject because the verb form *hubiese llamado* ('would have called') is ambiguous for person, so with no expressed subject it could be either first person singular or third person singular. However, given the context, with a first person singular object pronoun (i.e., *me hubiese llamado* 'would have called me'), the case for ambiguity is weak. It is doubtful that the speaker would feel it necessary to further resolve the potential for ambiguity by expressing the subject, *la man*. In addition, if the case were to disambiguate the subject, the speaker easily could have selected another lesser form of coding, i.e., the personal pronoun *ella* ('she'). A more plausible explanation arises when we consider the fact that the speaker uses a different full NP, *la man* ('the chick'), than the first mention (*tu ex*

pelada 'your ex-girlfriend'). It appears that by using a full NP Adrián wants to create some distance between himself and his ex-girlfriend and indicate his feelings about her to Rodrigo and therefore refers to her with a less personal term than ex-girlfriend. Lexical forms carry more information than lesser forms of coding, and pronoun or unexpressed mention would not be able to convey the stance that the full NP does in this example.

(63) Full NP after unexpressed mention

- R: ... (3.0) *que pasó con tu= --*
con= --
tu ex pelada que le dijiste que venga,
ya no ø viene?
- A: .. *no,*
ø no vino ya.
mejor.
 ... [le] --
- R: [pero],
le llamaste a decir que no ø venga?
- A: *no.*
- R: .. *ø no vino?*
- A: .. *no.*
 .. *es que ya salí de mañana.*
 ... *porque de ley **la man** me hubiese llamado de mañana*
acá a la casa y ø me hubiese dicho ya voy.
 ... *entonces ø ya no me encontró.*
- R: '... (3.0) what happened with your --
 with --
 your ex-girlfriend that you told to come over,
 (she's) not coming?
- A: .. *no,*
 (she) didn't come.
 better.
 ... [she] --
- R: [but],
 did you call her to tell her not to come?
- A: *no.*
- R: .. (she) didn't come?
- A: .. *no.*
 .. *it's that I left this morning.*
 ... *because for sure **the chick** would have called this*
morning and (she) would have told me she's
coming.
 ... *so (she) didn't find me.'*

(Fumar:527-541)

The NP *la silla* ('the chair') in (64) presents an interesting case. At this point in the conversation, the wheelchair has already been the topic of conversation for quite some time. In this excerpt, it first occurs expressed as an object pronoun (*le* in *le utilicé* 'I used it'), then as a full NP as an object of a preposition (*con la silla de ruedas* 'with the wheelchair'), immediately followed by an unexpressed mention as a subject ($\emptyset$ *me ha sido muy útil* ('(it) has been very useful to me'). After a change in syntactic role, we may have predicted that *silla de ruedas* would be expressed in $\emptyset$ *me ha sido muy útil* ('(it) has been very useful to me'). Even more curious, the very next occurrence of *silla*, in the following clause, is expressed, although here there is continuity of syntactic role between the two mentions. It should also be noted that the intonation of the IU where *silla de ruedas* appears as an object of a preposition and an unexpressed subject is interesting – a more typical parsing of IUs would occur with *con la silla de ruedas* in its own IU, as an increment to the previous IU, followed by the clause $\emptyset$ *me ha sido muy útil* on the next line, as the beginning of a new action, a characterization of the wheelchair. There is, however, a plausible interactional explanation. Just prior to the IU in question (*con la silla de ruedas* $\emptyset$ *me ha sido muy útil*), Beatriz's turn was interrupted by her brother. Beatriz then reaches a possible CTRP (Complex Transition Relevance Place, cf. Ford and Thompson, 1996), or place where it is likely for a change of speaker to occur, in the IU following the overlap (*también* $\emptyset$ *asistí* '(I) also attended'), yet does not want to give up her turn at talk, so she speeds up her speech in the next IU (a technique known as "rush through", see Schegloff (1982, 1987)), producing what would normally be two IUs under one intonation contour. Perhaps she realized that the resulting IU sounds awkward, with the same referent (wheelchair) occurring in two syntactic roles and later attempts to repair this by expressing *silla* as a full NP in the IU *la silla es muy agradable* ('the chair is very nice'), then completes the list with another unexpressed mention. The resulting IUs: *también* $\emptyset$ *asistí*, *con la silla de ruedas* $\emptyset$ *me ha sido muy útil*, '(I) also attended. with the wheelchair (it) has been very useful,' present a combination of two clauses that is similar to what has been recognized in the literature as syntactic amalgams (Lambrecht, 1988). It should also be noted that the syntax of this three-part list deviates from the norm described above, where expressions in a list tend to exhibit repetition of the same form. In this list, however, perhaps due to the previous interactional problems, we see the first part realized with an unexpressed subject ($\emptyset$ *me ha sido muy útil*), an expressed subject in the second (*la silla es muy agradable*) part, and a return to unexpressed in the third part ($\emptyset$ *es muy buena* '(it) is very good').

(64) Full given NP after change in syntactic role

B: *sí,*
sí.
 $\emptyset$ *le utilicé para el velorio de --*

- (H) *de= mi mami,*
el tra- --
el traslado de mi mami.
.. (H) también fui a la misa de --
- G: *... [cla=ro,*
a la misa de mamá].
- B: *[a la misa de un año de fallecimiento] de mis*
padres,
también ø asistí.
con la silla de ruedas ø me ha sido muy útil,
... (H) mhm.
la silla *es muy agradable,*
ø es muy buena.
- B: *'yes,*
yes.
(I) used it for the wake of --
(H) of my mom,
the fu- --
the funeral of my mom.
.. (H) I also went to the mass for --
- G: *... [of course,*
to mom's mass].
- B: *[the mass for the one year anniversary of the*
death] of my parents,
(I) also attended.
with the wheelchair (it) has been very useful,
... (H) mhm.
the chair *is very nice,*
(it) is very good.' (Hermanos:172-186)

To summarize, in this section we have examined six referents that occur as full NPs in contexts in which theories of information flow would predict lesser coding. These examples exhibit linguistic coding which deviates from patterns related to information flow seen repeatedly in the discourse, and it is only by examining each token in its natural context that the meaning of the deviation is unveiled. It is evident from these examples that factors beyond activation status and distance from previous mention contribute to the choice of encoding a referent. Speakers choose to code given referents as full NPs to accomplish different interactional goals – indicating to their interlocutor that they are constructing repetition, managing turn-taking problems, marking an referent via choice of lexeme to indicate attitude toward or psychological distance from that referent, and to clarify a referent. Although it is not possible to quantify the scope of each of these interactional factors that contribute to referent realization, as each instance must be examined in the context in

which it occurs, it is important and worthwhile to note that overall, interactional factors contribute in important ways to the realization of referents in conversation.

6.1.2 Discourse referential NPs and modification

One of the most frequent types of modification of full NPs in Spanish is with an article, either definite or indefinite. The terms “definite” and “indefinite” have two different, but related, meanings. In the discourse-pragmatic sense, there is a difference between definite referential nouns, or those which are “assumed by the speaker to be uniquely identifiable to the hearer” (Givón, 1978, p. 296) and indefinite referential nouns, which are “not so assumed” (Givón, 1978, p. 296). In the descriptive sense, some languages are thought make such a distinction in the morphosyntax of the NP, as reflected in the terms definite and indefinite articles. Spanish (like English) has definite and indefinite articles, which may correspond to a degree to the discourse-pragmatic meaning, yet also have different uses. In this section, the occurrence of new NPs with different modifiers is examined with an attempt to find links or patterns between form and function.

Articles in Spanish have been linked to a number of different functions. The definite article is commonly said to be used with entities (referents) that the speaker assumes are known or identifiable to their interlocutor (Bosque, 1989; Klein, 1976; Ortiz Ciscomani, 2009), as well as with generic (referring to an entire class, i.e. they are a sub-class of non-specific; see Torres Cacoullos & Aaron, 2003, for evidence for different behavior from other non-specific NPs) entities (Butt & Benjamin, 2004; Ortiz Ciscomani, 2009). Ortiz Ciscomani (2009) lists the following ways in which a speaker may assume that a referent is known to the hearer (1) previous mention, (2) previous mention of an associated entity or an evoked frame, (3) parts of a whole, including body parts, (4) identifiable through “shared cultural world”, (5) nouns of unique reference, (6) presence in the physical environment of the speech occasion. In addition, Butt and Benjamin (2004:31) note that an article (definite or indefinite) is required before nouns modified by a qualifier, as in (65), presumably because the modification is a pathway to identifiability or serves to mark the NP as a unique referent. Ortiz Ciscomani (2009) also notes that the definite article may be used to convey subjectivity and topicality. In addition, several dialects of Spanish use definite articles before proper names (De Mello, 1992), which is extremely common in this variety of Ecuadorian Spanish.

(65) Definite article preceding modified NP

G: *creo que tiene que divorciarse por **las leyes de=**
Estados Unidos,*

G: ‘I think that she has to get divorced because of **the**
laws of the United States,’ (Gerardo: Café, 1416)

The indefinite article in Spanish seems to share some uses of indefinite article in English, although there are important differences (cf. Klein, 1976; Torres Cacoullos & Aaron, 2003).¹¹ Klein (1976) points out the retention of the numerical meaning of the indefinite article, showing that *un* (the masculine singular indefinite article) is often more functionally similar to ‘one’ in English than it is to the English indefinite article, ‘a’, in the sense that it serves to individuate referents. The Spanish indefinite article has also been said to mark new, unidentifiable information (Garachana Camarero, 2009). Unlike English, Spanish makes use of a plural indefinite article, as in (66).

(66) Plural indefinite article

B: *unos cinco tornillos están ahí.*

B: ‘some five screws are in there.’ (Hermanos: 102)

Garachana Camarero (2009, p. 425) sums up the difference between the definite and indefinite articles as essentially being tied to the identifiability of the entity – the use of the definite article implies identifiability, the use of the indefinite article with referents implies unidentifiability. In the following sections, usage of articles and other modifiers is examined with reference to discourse functions, as they related to identifiability and accessibility, among other functions.

6.1.2.1 *New referents and modification*

The distribution of lexical first mentions with different types of modifiers is presented in Table 21.

The same proportion of new referents is introduced with an article (definite or indefinite) between the two genres (76%, see Table 21), however the rates of each article are quite different. The definite article occurs at a higher rate in the conversations as compared to the narratives (63% versus 33% of all first mentions), which may suggest that speakers treat more referents in the conversations as identifiable, supporting the hypothesized link between identifiability and the definite article (Garachana Camarero, 2009). In the narratives, the most frequent form for a first mention is with the indefinite article (43%), but first mentions with the definite article are not uncommon (33%).

11. Among the notable differences between the languages is the Spanish tendency to omit the indefinite article preceding nouns referring to profession, occupation (*Es Ø abogado* ‘s/he is a lawyer’), sex (*Soy Ø mujer* ‘I am a woman’) and other complements of *ser* and the verb *tener* (*¿Tienes Ø carro?* ‘Do you have a car?’).

Table 21. New referents and determiner type (lexical NPs only; N = 393)

	Narratives		Conversation	
	N	%	N	%
Definite article	41	33	171	63
Indefinite article	53	43	35	13
Possessive adjective	14	11	41	15
Demonstrative adjective	1	1	10	4
Descriptive adjective ¹²	6	5	5	2
Quantifier	8	7	8	3
Total	123	100	270	100

One factor that contributes to the higher proportion of new referents with indefinite articles in the Pear Stories is related to genre-specific communicative actions. Of all the new human referents in the Pear Stories, 60% (32/53) are introduced with an indefinite article; only 17% are introduced with a definite article (9/53). On the other hand, in the conversations, only 7% (7/105) of the new human referents are introduced with an indefinite article, as in (67), and 62% (65/105) are introduced with a definite article. Of the new human referents in the conversations, over one third (38%, 40/105) are first names, as in (68), which are categorically modified by a definite article (100% [40/40] of first mentions of names are modified with a definite article in these data).¹³ First names (and other proper nouns) are not used in the Pear Stories, which helps to explain why fewer overall new referents are introduced with a definite article in this genre. However, when the first names are excluded from the analysis, as Table 22 shows, there is still a significantly higher proportion of new referents marked with the definite article in the conversations (57%) versus the narratives (33%). The proportion of non-human referents introduced with indefinite articles is nearly twice as high in the narratives than in the conversations (30% [21/70] in the narratives, 17% [28/165] in the conversations).

12. Recall from the Coding chapter that some nouns occur with both descriptive adjectives and other determiners (e.g., *mi mejor amiga* 'my best friend'). These cases (n=30 in the narratives and n=65 in the conversations) were coded for the other determiner (e.g., possessive pronoun) and not the descriptive adjective. These tokens all refer to the adjective *otro* or an ordinal number (e.g., *le están viendo otro colegio* 'they are looking into another school'), as no descriptive adjectives occurred without another determiner.

13. Note that first names can be modified by other definite determiners in this variety of Spanish: demonstrative and possessive adjectives, e.g., "*pero yo le veo muy cagado a este Diego*" but I see that Diego as really screwed (Jugo: 1366) and "*ahí quería ponerle a mi Ricardo*" I wanted to put my Ricardo there (Café: 693).

- (67) Discourse new referent modified by indefinite article
- F: *eh,*
una de mis compañeras tiene un hijito de unos cuatro
o cinco años.
- F: *'eh,*
one of my colleagues has a son about four or five
years old.' (Birthday: 251–253)
- (68) First name with definite article
- J: *.. qué más que quieres que te cuente Rodriguito?*
... (2.0) ah,
ahora también me llamó a la oficina la .. Claudia.
- J: *'.. what else do you want me to tell you Rodriguito?*
... (2.0) ah,
today Claudia called me at the office.' (Jugo: 300–302)

Table 22. Proportion of new referents introduced with definite article (excluding proper names) (N = 353)

	Narratives		Conversations	
	N	%	N	%
Definite article	41	33	130	57
Other forms of modification	82	67	100	43
Total	123	100	230	100

Chi-square = 16.338, $p < 0.0001$

It should be noted that the use of the definite article with first names is not limited to first mentions; nearly all first names in the conversational data (90%, 129/144, excluding 58 vocative uses, where articles do not occur) are accompanied by the definite article, regardless of activation status or referentiality.¹⁴

Other common uses of the definite article for first mentions in the conversations include qualified NPs ($n = 27$, see (69)), the introduction of a relative of an already mentioned speaker (i.e., anchored referents, cf. Prince 1981) ($n = 12$, see (70)), body parts ($n = 9$, see (71)), lists, entities of unique reference ($n = 9$, see (72)). Of the remaining discourse new NPs modified by a definite article, the majority were frame-evoked NPs, as in example (73) below, where the topic of discussion is a soccer game. Of these, only body parts were found in the narratives ($n=2$), and they were also introduced with the definite article. Du Bois (1980, p. 237)

14. Of the remaining 10%, 2 of these are modified with a demonstrative adjective.

reports that body parts are the most frequent semantic class of definite mentions in American English Pear Story narratives.

(69) Qualified first mention

E: *oh=,*
te acuerdas cuando jugábamos carnaval con el huevo
podrido?

R: *.. (H) no,*
no recuerdo.

E: *qué a=sco,*
Rodrigo,
[pusieron] --

R: *[y yo te puse]?*

E: *no=,*
le lanzaron a la Mónica Jacome,
el balde de agua con el huevo podrido.

E: *'oh=,*
do you remember when we played carnival with the
rotten egg?

R: *.. (H) no,*
I don't remember.

E: *how disgusting,*
Rodrigo,
[you put] --

R: *[and I put (it) on you]?*

E: *no=,*
you (pl.) threw (it) at Mónica Jacome,
the bucket of water with rotten egg.'

(Viejos Amigos: 278–288)

(70) First mention of a relative

P: *... y **el abuelito de él,***
qué es?

P: *'... and **his grandfather,** (lit. 'the grandfather of*
him')
what does he do?'

(Café: 766–767)

(71) First mention of body part

J: *... me duele **el pie.***

J: *'... **my foot** hurts.'* (lit. 'the foot hurts me')

(Jugo: 23)

(72) NP of unique reference

P: *entonces usted por ver **el mundial** no se ha ido.*

P: *'so you didn't go because you were watching **the***
World Cup.'

(Café: 1150)

(73) Frame-evoked NP

- R: *pero,*
sí nos salvamos algunas,
yo metía la --
la pierna justito,
*y le quitaba **la pelota**.*
- R: 'but,
 we did save some,
 I just stuck out my --
 my leg,
 and I took **the ball** away from him.' (Jugo: 1204–1208)

Like the definite article, possessive adjectives may also relate to identifiability, as suggested by Ortiz Ciscomani (2009), but differ in function. Possessive adjectives often function to give the listener a framework for identifying a particular instance of a referent. The proportion of new nouns introduced with possessive adjectives is similar between the two genres – 11% of all new referents in the Pear Stories, and 15% in the conversations occur with possessive adjectives. In the conversations, possessive adjectives are commonly used (20/41 new nouns modified by possessive adjective) when introducing family members, as in (74). In the Pear Stories, body parts and clothing account for the six new referents modified by possessive adjectives, and the NP *su camino* ('his route') accounts for another four.

(74) Introduction of family member

- A: *... ya quiero saber si ya están aquí para irle a*
*visitar a **mis sobrinos**.*
- A: '... I want to know if they are here yet to go see
my nephews.' (Fumar: 56)

Like possessive adjectives, demonstrative adjectives also function to help identify or situate discourse new referents. New nouns modified by demonstrative adjectives make up a small proportion of the data in each genre (1% in the narratives, and 4% in the conversations). In all instances, they denote referents which are clearly identifiable to the interlocutor, either by presence in the physical environment, as in (75) in which the speaker is referring to a telephone lying on the table, or due to shared information, as in (76), where the participant is describing the movie they just viewed and were asked to narrate.

(75) New referent in surrounding environment

- R: *no ves que está súper descargado **ese teléfono**?*
- R: 'can't you see that **that telephone** is super
 uncharged?' (Jugo: 819)

(76) New “shared information” referent

*esta película comienza muy de mañana,**‘this movie starts early in the morning,’*

(PS38: 5)

New referents modified by quantifiers are somewhat more common in the narratives than in the conversations (7% of new referents in the narratives, 3% in the conversations) – this difference is more likely due to the design of the Pear Film than a true genre difference between narratives and conversation. All of the new referents modified by quantifiers in the Pear stories refer to either the three boys or the two baskets of fruit.

New referents modified by descriptive adjectives that do not co-occur with determiners (e.g., *otro colegio* ‘another school’, not *el otro colegio* ‘the other school’) are a very small proportion of new referents overall (5% in the narratives, 2% in the conversations). However, the descriptive adjectives among the new referents are primarily limited to one token type: all six tokens from the narratives are of the adjective *otro* (‘other’) + NP, and four of the five in the conversations are the adjective *otro*.¹⁵ The remaining conversational token is an ordinal adjective (*séptimo* ‘seventh’). It seems as though these descriptive adjectives occurring before new referents in Spanish serve to contrast these new referents with previously introduced referents in order to disambiguate two similar referents.

The indefinite article, then, when used with referential NPs, appears to function to introduce a new referent that the speaker does not anticipate the interlocutor(s) to be able to identify, or perhaps it is not important for the interlocutor to identify this particular referent, other than recognize it as a new, unique reference and to create a new cognitive file for this referent. The functions of the indefinite article are best seen in how they contrast with uses of other determiners in the first mention of referents. In the excerpt in (77), the indefinite article is used with the introduction of new referents that are not connected to any other entity in the discourse (*un campesino* ‘a peasant’, *un burrito* ‘a donkey’, *un niño* ‘a little boy’). When the second man is introduced, the speaker begins to introduce him with the indefinite article, but then cuts off and repairs the introduction using the determiner *otro* (NP *otro señor* ‘another man’) to contrast this character with the first man who was introduced. The introduction of the baskets, used to hold the already mentioned fruit, is done with the quantifier *dos* ‘two’. Eventually, at the very end of the narrative, the first character is re-introduced (*el agricultor* ‘the farmer’), but the use of the definite article suggests that the speaker expects the referent to be identifiable.

15. This suggests that *otro* is grammaticizing as a determiner rather than functioning as a descriptive adjective.

(77) Identifiability and determiners in Pear Story narrative

.. e=h,
 estaba **un campesino** recogiendo= .. unas=,
 eh,
 manzanas,
 de= su árbol.
 ∅ ya tenía .. (H) **dos canastas llenas de manzanas**,
 eh,
 es- --
 .. ∅ era --
 (H) ... (2.0) pero=,
 ... en el momento de .. seguir recogiendo,
 y subirse a su árbol,
 eh,
 vino **un** --
un --
 primero pasó **un** --
 .. **un burrito**.
un --
otro señor con un burrito.
 .. (H) y= --
 y ∅ lo vio,
 y= luego **él** subió a recoger .. más de --
 de su árbol.
 pero pa- --
 el momento que él estaba distraído recogiendo,
 pasó **un niño**.
 ... (H) con --
 en una bicicleta.
 .. é=l se acer- --
 ∅ se acercó,
 (H) y ∅ quiso coger --
 .. ∅ quiso coger una,
 pero=,
 la tentación le --
 .. le --
 .. (H) le venció y ∅ cogió **todo un canasto** en su
 bicicleta.
 ((40 intervening IUs))
el agricultor vio que ellos tenían la fruta,
 (H) y= ∅ se quedó sorprendido
 \.. e=h,
 there was **a peasant** picking= .. some=,
 eh,

apples,
 from his tree.
 (he) already had .. (H) **two baskets full of apples**,
 eh,
 es- --
 .. was --
 (H) ... (2.0) bu=t,
 ... in the moment to .. keep picking,
 and climb up his tree,
 eh,
 along came a --
 a --
 first came a --
 .. **a donkey**.
 a --
another man with a donkey.
 .. (H) and= --
 and (he) **saw him**,
 a=nd then he went up to pick.. more from --
 from his tree.
 but pa- --
 the minute that he was distracted picking,
 along came **a little boy**.
 ... (H) with --
 on a bike.
 .. **he= came ov-** --
 (he) **came over**,
 (H) and (he) wanted to take --
 .. (he) **wanted to take a**,
 bu=t,
 the temptation to him --
 .. to him --
 .. (H) conquered him and **he** grabbed a whole basket on
 his bicycle.
 ((40 intervening IUs)
the farmer saw that they had the fruit,
 (H) a=nd Ø was surprised. (PS29:1-39, 78-80)

Similar uses of determiners to establish identifiability are found in narrative excerpts from the conversations. In (78), husband (speaker F) and wife (speaker A) were discussing a few videos that speaker F had purchased that day for their nephew's upcoming birthday. Speaker A finds the videos in the car and asks her husband how he acquired *estos videos* ('these videos'), using the demonstrative adjective for the first (and subsequent mention). Then as speaker

F tells the story of how he decided to buy the videos, he talks about the role of his coworkers, who are introduced and later referred to with reference to the speaker – *mis compañeros* ('my coworkers'), and their children are referred to with the possessive adjective *sus* to reference to their relationship with the coworkers, *sus hijos* ('the children'). Later, when speaker F needs to highlight the role of one particular co-worker and to situate her son relative to the age of his other coworkers children, he uses the indefinite article, *una de mis compañeras tiene un hijito de unos cuatro o cinco años* ('one of my coworkers has a son about four or five years old'). Finally, an additional video, purchased by that co-worker is introduced with the determiner *otro*, as it contrasts with the already introduced video, *este Baby Mozart* ('this Baby Mozart').

(78) Identifiability and determiners in conversation

A: y **estos videos que ø ha comprado?**

F: oh=,
qué cheveru=cos.

A: qué li=ndos.
qué son?

F: ... so=n --
para niños.

A: ... oh=,
[qué chévere].

F: [es que como] salí temprano,
ten- --

mis compañeros --
venían justo para el= --

A: ... (2.0) centro comercial?

F: ... sí=,
venían justo para el centro comercial,

A: ... (3.0) y aprovechó y fue con ellos.
(28 intervening IUs)

F: yo realmente,
nunca había visto estos videos,
pero mis compañe=ros dicen que **sus hijos** son
fascina=dos.
.. tienen --
los hijos,
un año y me=dio,
más o [menos].

A: [ah],
es justo para la edad.
pero,
preguntó [ahí más o meno=s]?

- F: [e=h],
 sí,
 en el video [dice] de uno a do- --
- A: [mhm].
- F: a cuatro años.
- A: .. ya.
- F: eh,
una de mis compañeras tiene un hijito de unos cuatro o cinco años.
- A: hm,
 ya.
- F: entonces ella se llevó **o=tro**,
 ya no se llevó este "Baby Mozart".
- A: 'and **these videos that (you) bought?**
- F: oh=,
 how coo=l.
- A: how cute.
 what are they?
- F: ... they are --
 for kids.
- A: ... oh=,
 [how cool].
- F: [it's that since] I got out early,
 ha- --
my coworkers --
 were coming just by the= --
- A: ... (2.0) the mall?
- F: ... yes=,
 they were coming just by the mall,
- A: ... (3.0) and you took advantage of the opportunity
 and went with them.
- ((28 intervening IUs))
- F: I really,
 had never seen these videos,
 but my coworkers say that their kids are fascinated.
 .. they are --
 the kids,
 a year and a half old,
 more or [less].
- A: [ah],
 it's just right for that age.
 but,
 did you ask [there more or less]?

F: [e=h],
yes,
in the video [it says] from one to tw- --
A: [mhm].
F: to four years.
A: .. yeah.
F: eh,
one of my coworkers has a four or five year old son.
A: hm,
yeah.
F: so she bought another (one),
she didn't buy this "Baby Mozart".'
(Birthday:188–204; 235–257)

Not only are nouns modified by determiners, which syntactically precede the noun, but they can also be modified by relative clauses, which follow the noun. The overall proportion of referents that are heads of relative clauses is given in Table 23 below. Note that the occurrence of relative clauses is very low overall (only 4% of all referents are modified by a relative clause). There is a significantly higher overall proportion of referents modified by relative clauses in the narratives (8%) versus the conversations (1%). This may very well be an indicator of the higher information flow pressures of the narratives, as it reflects a more “condensed” form of packaging the information.

Table 23. Overall proportion of referents (lexical NPs and pronouns) modified by relative clauses (N=1160)¹⁶

	Narratives		Conversations	
	N	%	N	%
Not modified by relative clause	464	92	652	99
Head of relative clause	38	8	16	1
Total	502	100	658	100

Chi-square = 16.276, *p* < 0.0001

Looking now specifically at the new referents, as Table 24 shows for the overall figures, there remains a significantly higher proportion of new referents in the narratives that are modified by a relative clause. Again, this likely reflects the higher

16. Excludes unexpressed mentions and subtypes of given referents (members of a set, etc.).

information flow pressures of the narratives and the need for more information to be presented in a smaller unit.¹⁷

Table 24. New referents and relativization (N=415)¹⁸

	Narratives		Conversation	
	N	%	N	%
Not modified by relative clause	113	88	275	96
Head of relative clause	15	12	12	4
Total	128	100	287	100

Chi-square = 7.075, $p < 0.01$

(79) New referent modified by relative clause

existia primeramente un= campesino que estaba cosechando peras.

'at first there was a= peasant that was harvesting pears.'

(PS41:6)

To summarize this section, the data show patterns that suggest that article and other modifier usage is related to information flow and shared knowledge. In addition, the genre differences suggest that the narratives are higher in information flow pressures, and that external situational constraints such as the perceived lack of shared knowledge, or perhaps the formality of the situation in which these data were collected, may be evidenced in the higher rates of first mentions with indefinite articles.

6.1.2.2 Given full NPs and modification

As mentioned earlier, one of the uses of the definite article in Spanish is to signal identifiable information. As seen in Table 25, the high percentages of given referents that are modified by definite articles in these data (79% in narratives, 73% in conversations) confirms that given referents are frequently accompanied by the definite article. When these proportions are compared to the proportion of new referents modified by the definite article (33% in narratives and 63% in conversations, cf. Table 21) the robustness of this use is evident. As opposed to new referents, however, where the definite article is more common in conversations, for

17. Thompson (1984) finds a higher rate of subordination in more formal written versus informal written discourse, although relative clauses were excluded from her study. This suggests that formality may be a significant factor in the rate of subordination.

18. Excludes unexpressed mentions.

the given referents, the definite article is more common among in the narratives. This is likely due to the higher frequency of other definite markers (e.g., possessive and demonstrative adjectives) in the conversations; when these are collapsed, we see that 93% of given referents have some kind of definite marking in the conversations, 88% in the narratives.

Table 25. Given referents and type of modification (N=394)¹⁹

Modifiers	Narratives		Conversation	
	N	%	N	%
Definite article	174	79	129	73
Possessive adjective	14	6	20	11
Demonstrative adjective	6	3	16	9
Indefinite article	9	4	8	5
Quantifier	14	6	3	2
other determiner	2	1	1	1
Total	219	100	177	100

The distribution of given referents modified by a relative clause (6% in narratives and 1% in conversations, see Table 26) is very similar to the distribution of all referents and relativization (8% in narratives and 1% in conversations, see Table 23) above, and shows the same trend (i.e., more modification in the narratives) as the distribution of new referents modified by a relative clause (12% in narratives and 4% in conversations, see Table 24). Again, the higher proportion of referents in the narratives that are modified by a relative clause reflects the higher information pressures of this genre. Not only is it a more condensed form of packaging, but in the examples below, it allows speakers to syntactically treat referents as other given referents (in the sense that they are modified with a definite article), yet give additional information about the referent to remind the interlocutor of the relevance of this particular referent. In (80) and (81) the relative clauses serve to help identify referents that had not been mentioned in recent discourse (both previous mentions were at a referential distance of more than ten clauses).

- (80) Given referent modified by a relative clause
mientras tanto,
el campesino que estaba en la cosecha,
.. (TSK) (H) baja y=,
se da cuenta de que no hay una canasta.

19. As with modification of new referents, only full NPs are analyzed here.

'meanwhile,
the peasant that was in the harvest,
 .. (TSK) (H) goes down and=,
 realizes that one of the baskets is not there.'
 (PS52: 69–72)

- (81) Given referent modified by relative clause
los tres niños siguieron también su camino,
 (H) *y= encontraron **el sombrero que se había .. caído,***
el sombrero del niño.
 'the three boys continued on their way,
 (H) and found **the hat that had .. fallen,**
 the boy's hat.'
 (PS 36: 72–72)

Table 26. Given referents modified by relative clause (lexical NPs and pronouns; N=755)²⁰

	Narratives		Conversation	
	N	%	N	%
Not modified by relative clause	351	94	377	99
Head of relative clause	23	6	4	1
Total	374	100	381	100

Fisher's exact test, two-tailed $p < 0.0001$

6.1.3 Activation status and syntactic role

6.1.3.1 *New referents and syntactic role*

From previous studies of Preferred Argument Structure (cf. Ashby & Bentivoglio, 1993; Du Bois, 1987, 2003b) we know that the preferred grammatical roles for new referents are S and O, but that verbless constructions can also be useful in introducing new reference (Herring, 1989). In this study, 87% of new referents in the narratives are in the S and O roles, and 69% of the new referents in the conversation are in the S and O roles (see Table 27).

One of the interesting trends seen in Table 27 is that a relatively large number of new referents in the conversations are introduced in the "Other" syntactic role, as in (82) and (83). In (82), Karla is a family member, and in (83) Santi and Luis are other friends with whom the speakers frequently play soccer.

20. Excludes unexpressed referents

21. Due to the low number of heads of relative clauses, Fisher's exact test was used instead of Chi-square

Table 27. New referents and syntactic role (N = 428)

	Narratives		Conversation	
	N	%	N	%
A	9	7	20	7
S	68	52	102	34
Direct Object	45	35	105	35
Indirect Object	–	–	9	3
Other	5	4	55	18
Continuation of previous syntactic frame	3	2	7	2
Total	130	100	298	100

(82) New referent in “Other” syntactic role

- P: *mhm=.*
... (H) *y **la Karla**?*
... *ella también ya se graduó este año?*
... (3.0) *y ella,*
en qué?
no va a seguir la universidad?
- P: *'mhm=.*
... (H) *and **Karla**?*
... *did she graduate this year too?*
... (3.0) *and she,*
in what?
(she's) not going to go to college?' (Café: 796–801)

(83) New referents in “Other” syntactic role

- R: *el Diego juega bien,*
*[**el Santi** también].*
- J: *[ese Diego juega bien],*
chucha.
el= --
quien más,
el Guillermo solo.²²
- R: *el Luis no.*
el Luis se [da muchas vueltas].
- R: *'Diego plays well,*
*[**Santi** too].*

22. ‘Guillermo’ had already been introduced previously in the conversation.

J: [that Diego plays well],
 shit.
 the= --
 who else,
 just Guillermo.
 not **Luis**.
 Luis runs around a lot.'

(Jugo: 774–782)

Although these “Other” or free NPs have been largely ignored in other large scale studies of the effects of information flow (cf. Ashby & Bentivoglio, 1993; Du Bois, 1987, 2003b, etc.), several previous works have pointed out that free NPs in a variety of languages can introduce new referents (Bentivoglio, 1993; Croft, 1995, 2007; Herring, 1989; Kumpf, 2003; Matsumoto, 2003; Ono and Thompson, 1994; Tao, 1992). With nearly one fifth of the new referents in the conversation occurring in the “Other” syntactic role, this is clearly a very productive strategy in Spanish for introducing new referents. Bentivoglio (1993, p. 218) shows that most of the “Other” NPs in her Venezuelan Spanish data (a “mixture of narrative, descriptive and expository genres” Bentivoglio 1993, p. 215) are actually used for given referents, but as will be seen in the next section, in Table 28, this trend holds for the narratives in this data (with only 5 new “Other” referents in this genre, compared to 13 given), but not for the conversations (in which there are 55 new referents, a number greater than the 49 given “Other” referents). Herring’s (1989) study of Tamil reported that this strategy, while existent, is not very productive in narratives, with only 2.4% of finite clauses containing the verbless presentational structure. It should also be noted that the much higher proportion of “Other” new referents in the conversations than in the narratives is likely due to the fact that they are highly identifiable to the speakers. Of the 60, nearly one half (N=25) are proper nouns (Prince’s (1981) “unused”), and another eleven are “anchored” (Prince, 1981) with possessive adjectives, as in (84).

- (84) Anchored new referent
- R: 'y **tu pelada**?
 ... la recepcionista?’
- R: 'and **your girlfriend**?
 ... the receptionist?’
- (Jugo: 1089–1090)

6.1.3.2 *Given referents and syntactic role*

The distribution of given referents and syntactic role is presented in Table 28. The two genres show similar patterns. The preferred roles for given referents are subject of intransitive verbs, followed by subject of transitive verb, direct object, and indirect object. Syntactic roles “Other” and extension of a previous syntactic frame

make up less than 10% of the data in each genre. Activation state and syntactic role are considered together in Section 8.2.

Comparing the two genres, in the narratives we see a higher number of given referents as subjects of transitive verbs (26% vs. 15%, Chi-square = 25.711, $p < .0001$) and direct objects (20% vs. 15%, Chi-square = 5.742, $p < .05$), and a lower proportion of subjects of intransitive verbs (40% in narratives, 51% in conversations, Chi-square = 14.082, $p < .001$). This may be another genre difference, namely, that speakers make more use of transitive constructions in the narratives, at least for third person subjects (further discussion of how the genres differ in transitivity is presented in 8.3). Additionally, there are more given referents in the syntactic role “Other” in the conversational data than in the narratives (7% vs. 2%, Chi-square = 54.912, $p < .0001$).

Table 28. Given referents and syntactic role (N=1379)

	Narratives		Conversation	
	N	%	N	%
A	171	26	108	15
S	265	40	366	51
Direct object	129	20	106	15
Indirect object	75	11	92	13
Other	13	2	49	7
Continuation of previous syntactic frame	3	–	2	–
Total	656	100	723	100

6.1.4 Summary of new referents

Not only are there proportionally nearly twice as many new referents in the conversations than in the narratives (29% vs. 17%), but there are also measurable differences between the genres in the way in which new NPs are modified, the number of new referents that are heads of relative clauses, and the preferred syntactic roles of new referents.²³ There are no significant differences between the genres for realization.

23. This alone suggests that there is higher information pressure in the conversations, however it is important to recall that first and second person referents are not included in this study, which would affect the overall proportion of new referents in the conversation, likely showing a much lower proportion of new referents overall.

New referents are overwhelmingly realized as full NPs, and tend to be modified by either a definite or indefinite article. First mentions of referents in the narratives are more likely to be introduced with indefinite articles, and the definite article is more common in conversation, reflecting the higher rate of shared knowledge between participants. The higher rate of new referents that are modified by relative clauses in the narratives likely reflects the higher information flow pressures of the genre. Regarding syntactic roles, the narratives and conversations show approximately the same rate of new referents as subjects of transitive verbs, direct objects, and continuations of previous syntactic frames, but there are more new referents in the syntactic role “Other” in conversations, and more new referents as subjects of intransitive verbs in the narratives, revealing different strategies for referent introduction across genres.

6.1.5 Summary of given referents

Given referents make up 83% of all referents in the narratives and 71% of the referents in the conversations. The conversational figure is comparable to what Sakita (2003) found for American English Conversation, in which 78% of overt NPs were given. Despite the difference in proportion between the two genres, the patterns evidenced in the linguistic coding of similar referent types are very similar across genres. There are slight differences in the rates of modifiers used to encode given referents, rates of relativization and distribution of referents across syntactic role between the narratives and the conversations, which reflect different facets of the communicative actions present in each genre.

As was mentioned earlier, the higher information flow pressures of the narratives contribute to a greater proportion of given referents modified by relative clauses in this genre, and also contributes to the more frequent allocation of referents to transitive roles. This relationship between transitivity and linguistic coding is explored in more depth in Chapter 8, Sections 8.2 and 8.3.

6.2 Additional considerations for referents

This section explores two phenomena that are related to information flow, yet direct comparisons between new and given referents cannot be made. First, the semantic class of verbs that introduce new referents is examined, followed by a section detailing the linguistic coding of referents which are neither truly new or given under the criteria used in this study (i.e., members of a set, re-named mentions, or first group mentions).

6.2.1 New referents and semantic class of verb

For Spanish, Ashby and Bentivoglio (1997) find that “truly new referents” (i.e., not previously mentioned and not identifiable) are commonly introduced with the existential verb *haber* and the verb *tener* (‘to have’). Sánchez-Ayala (2005) finds that *haber* has an almost exclusively presentative function in Spanish, but that *ser* (‘to be’), *tener*, and motion verbs are also commonly used as presentatives.²⁴

The distribution of new referents and semantic class of verb is given in Table 29. In these data, *haber* is scarcely used to present a new referent into discourse. However, copulas (11% in narratives and 19% in conversations) and motion verbs (28% in narratives and 9% in conversations) appear to be rather productive resources for introducing new referents, and perception verbs (14%) are clearly useful in the narratives. *Tener* is more frequently used to introduce new referents in the conversations (7%) than in the narratives (4%). As with preferred syntactic role for new referents, we also different verb types being used in the two genres.

Table 29. New referents (in A, S, DO and IO roles) and semantic class of verb (N=358)²⁵

	Narratives		Conversations	
	N	%	N	%
Other ²⁶	51	42	133	56
Copula	13	11	46	19
Motion	34	28	21	9
Perception	17	14	15	6
<i>Tener</i>	5	4	17	7
<i>Haber</i>	2	2	4	2
Total	122	100	236	100

While the distribution in Table 29 shows the frequency of different verb classes with new referents, this may also reflect the overall frequency of these verb classes in the data. In order to compare which verb classes favor new referents to given referents,

24. Yet Ashby and Bentivoglio (1993) report that copulas strongly disfavor new referents in their data.

25. Only core arguments (A, S, DO and IO) are included. New referents in the ‘Other’ syntactic role do not occur with a verb and are therefore not included in this count.

26. ‘Other verbs’(not to be confused with the syntactic role, ‘Other’), includes cognitive process, speech act and other stative, impersonal and experiencer verbs

Table 30 and Table 31 show the overall frequency of these verbs, along with proportions of given and new referents. As Table 30 shows for the narratives, copulas (Chi-square = 14.204, $p < 0.001$), and perception verbs (Chi-square = 6.009, $p < 0.05$) take significantly higher proportions of new referents in the narratives when compared to the given referents.²⁷ The proportions of *tener* (Chi-square = 0.643, $p = 0.4225$), motion verbs (Chi-square = 2.886, $p = 0.0894$), or *haber* (Chi-square = 2.588) are not significantly different. This is not the case for conversations. As Table 31 shows, only *tener* (Chi-square = 7.673, $p < 0.01$) and *haber* (Chi-square = 5.005, $p < 0.05$) show significantly higher proportions of new referents.

Table 30. Distribution of verb type for Given and New referents in narratives (N=762)

	Given		New		Total	
	N	%	N	%	N	%
Other	431	89	51	11	482	100
Copula	18	58	13	42	31	100
Motion	131	79	34	21	165	100
Perception	44	72	17	28	61	100
<i>Tener</i>	15	75	5	25	20	100
<i>Haber</i>	1	33	2	67	3	100
Total	640	84	122	16	762	100

Table 31. Distribution of verb type for Given and New referents in conversations (N=908)

	Given		New		Total	
	N	%	N	%	N	%
Other	404	75	133	25	537	100
Copula	126	73	46	27	172	100
Motion	92	81	21	19	113	100
Perception	30	68	15	32	44	100
<i>Tener</i>	19	51	17	49	37	100
<i>Haber</i>	1	20	4	80	5	100
Total	672	74	236	26	908	100

27. Chi-square tests in this section compare the distribution of given and new referents among the different verb types between the genres.

It is not surprising that perception verbs, as in (85) would be more productive in the narratives than in the conversations, as part of the task was for the participants to narrate what they saw. For this reason, we see verbs such as *ver* and *escuchar* used in the construction of the narrative, with the participating detailing what he or she saw, or what the film characters did or did not see. In addition to the prominence of perception verbs, motion verbs also play an important role in the narratives. The design of the Pear Film, with new characters constantly arriving, as in the introduction of the three boys in (86), leads to a relatively high proportion of motion verbs in this genre (22%, 165/762, of referents in the narratives occur with a motion verb, whereas only 12%, 113/908, in conversation do).

- (85) First mention with verb of perception

acabo de ver una película,

'I just saw a movie,'

(PS13, 3)

- (86) Referent introduction with motion verb

luego vinieron .. tres niños más que .. pasaban por ahí,

'then three little boys that were passing through

came along,'

(PS41, 28)

This section showed that information flow considerations are relevant not only in the morphosyntax of the referential expression, but there are also correlations between semantic classes of verbs and the communicative action of introducing a new referent. The genre differences likely reflect the most useful types of predicates for introducing various types of referents; actors or characters in a story are referenced in terms of when they arrive upon the scene (motion verbs) or how their presence is first noted (verbs of perception), whereas the types of new referents in the conversation are best introduced in terms of their existence (*haber*, cf. example (87)) or their relationship to another referent (*tener*, see (88)).

- (87) New referent with *haber*

R: *el lunes?*

que yo llevé el atún para comer,

no cierto?

y había otra chica comiendo atún,

y ayer también l- --

la misma estaba otra vez comiendo atún.

R: 'Monday?

that I brought tuna to eat,

right?

and there was another girl eating tuna,

and yesterday too th- --

the same one was again eating tuna.'

(Food:223–228)

(88) New referent with *tener*

P: .. y Ø ya no trabajan como profesores?

M: ya el abuelito ya es jubilado.

P: ..[ah=].

M: [la abuelita] creo que está ya por jubilarse.

P: ... pero ellos tenían **un colegio**,
Ø creo.

.. o no?

o no era de ellos?

P: '.. and (they) don't work as teachers any more?

M: the grandfather is already retired.

P: ..[ah=].

M: [the grandmother] I think is about to get retired.

P: ... but they had **a high school**,
I think.

.. or not?

or it wasn't theirs?'

(Café:775–782)

6.2.2 Other types of “given” referents

A small set of given referents, including individuals who had previously been introduced as belonging to a group (cf. Du Bois' (1980) “member of a crowd phenomenon”), groups whose members had previously been introduced separately (“first group mention”), and individuals who were renamed is examined in this section. As Du Bois (1980) points out, these types of mentions may be treated differently than other given mentions because the individual referent has not been explicitly referred to, and therefore may not be activated individually. In Table 32 the distribution of these referents is presented.

Table 32. Other types of given referents (N=128)

	Narratives	% of referential NPs	Conversation	% of referential NPs
Member of a set	56	6	20	2
Renaming	28	3	12	1
First group mention	4	<1	8	1
Total	88	10	40	4

“Members of a set” and “renaming” referents are more common in the narratives than in the conversations, which is likely due to the design of the Pear Film (in particular, the higher proportion of members of a set is mostly due to the group of three boys that help the main character collect his pears). However, in each genre,

“members of a set” is the most common subtype of given referents, followed by renaming, and lastly, “first group mention”.

As Table 33 shows, individuals whose immediately prior mention had been as members of a group are most likely to be realized as a full NP (66% in the narratives, 80% in the conversations), as in (89). A substantial portion of “members of a set” in each genre (34% in narratives, 20% in conversations) were realized pronominally. Close analysis reveals that all of the pronominal mentions were numerical pronouns (see (90)).

Table 33. Realization of “members of a set” (N=76)

	Narratives		Conversation	
	N	%	N	%
Full NP	37	66	16	80
Pronoun	19	34	4	20
Total	56	100	20	100

Chi-square = 0.775, *p* = 0.3786 (not significant)

(89) “Member of a set” as full NP

(H) *tenía tres canastos.*
.. (H) *por ahí pasó una --*
un señor con una cabrita,
pasaban viendo las peras,
pero siguieron.
(H) *luego viene un niño,*
.. *con una bicicleta.*
(H) *entonces es- --*
al ver a la --
la --
dos canastas *llenas de --*
'(H) he had three baskets.
.. (H) along came a --
a man with a goat,
they went by looking at the pears,
but they continued on.
(H) then a boy comes along,
.. with a bicycle.
(H) so it's- --
upon seeing the --
the --
two baskets full of --'

(PS40: 24–34)

- (90) “Member of a set” realized pronominally
y con todas las canastas,
*.. y parecía que le faltaba **una**,*
 ‘and with all the baskets,
*.. and it seemed that he was missing **one**,’ (PS38:90–91)*

This indicates that speakers do treat these referents differently than other given referents, where pronominal and unexpressed mentions are the preferred method of coding (67% in the narratives, 75% in the conversations, cf. Table 17). As will be seen in Table 34, referents which had previously been introduced as members of a set were most likely to be modified by a definite marker (57%) in the narratives, and in the conversations, they were overwhelmingly (94%) introduced with a definite marker.

Table 34. Modification of “members of a set” (N=53)²⁸

	Narratives		Conversation	
	N	%	N	%
Definite marker	21	57	15	94
Indefinite article	16	43	1	6
Total	37	100	16	100

Fisher’s exact test $p = 0.0098$

It is interesting that in the narratives modification patterns for individuals previously introduced as members of a set are similar to those for new referents (cf. Table 21). With respect to definiteness, the overall rate of definite marking is similar between both given and new referents in the conversations (87% of new referents and 82% of given referents are definite), and the individuals previously introduced as members of a set in the conversations shows and even higher proportion of definite marking (94%). This shows that the genre difference regarding “members of a set” reflects the overall patterns of modification for new referents in each genre.

- (91) Member of a set with indefinite article
.. mientras subía y bajaba con los canastos,
 (H) *eh=,*
y el árbol,
.. que= --
mientras --
*ya cuando llena **una canasta**,*

28. Only those referents realized as a full NP were included in this count.

'.. while he went up and down with the baskets,
 (H) eh=,
 and the tree,
 .. that --
 while --
 when he fills **a basket**, ' (PS43:26–31)

(92) Member of a set with definite article

P: *entonces*,
ya Ø le han matriculado en el Colegio Americano,
pero,
 ... (2.0) *pero **el Roberto** dice que no quiere que vaya*
a ese colegio,
 P: 'so,
 (they've) already registered him in the American
 High School,
 but,
 ... (2.0) but [the] **Roberto** says that he doesn't
 want him to go to that school. (Café: 678–681)

Renaming is moderately productive in the narratives. The majority (19/28) of the “renamed” referents are the pears – referred to sometimes as just fruit (*frutos*, *frutas*), other times as apples (*manzanas*), and at times just as the harvest (*la cosecha*). Another nine “renamed” referents refer to different characters. For example, some speakers vacillate between the labels *campesino* (‘peasant’) and *agricultor* (‘farmer’) when referring to the man who collects the pears and the man pulling the goat is sometimes *un joven* (‘a young man’) and at other times a *campesino*. Two speakers experience uncertainty about what type of animal is being pulled, suggesting both *oveja* (‘sheep’) and *chivo* (‘goat’) or *cabra* (‘goat’) as possibilities. Lastly, one speaker refers to the road as both *camino* (‘road’) and *sendero* (‘path’). In the conversations, renaming often occurs to further specify a referent, as in (93), but giving two different labels to the same referent, as in the narratives, is also found.

(93) Renamed referent

A: *y **tu niño**,*
el --
el --
***el de pelito largo**,*
me --
cómo se lla=ma,
 A: 'and **your brother**,
 the --

the --

the one with the long hair,

me --

what's his na=me,'

(Adán, Chicas, 422–427)

All of the 28 renamed referents in the narratives are full NPs, as are 10 of the 12 renamed referents in the conversations. The other two are modified pronouns as in (93). Nearly all renamed referents are definite (i.e., treated as given), except for new referents for which the speakers had trouble coming up with an appropriate label, as in (94).

(94) Indefinite renamed NP

.. *estaba halando una oveja.*

... o **un chivo.**

`.. he was pulling a sheep.

... or **a goat.**'

(PS 52:27–28)

The last subset of given mentions, those which had previously been introduced individually and are later referred to as a group, is also the least frequent of the three types, with only four tokens in the narratives and eight in the conversations. They tend to be unexpressed, indicating that the speaker feels that their interlocutor(s) will have no problem identifying the referent in question. An example follows in (95). In this example, the speaker had previously introduced the boy who steals the pears (the subject of *está yendo en la bicicleta* 'is going on the bicycle') and the girl on the bicycle (*pasa una niña* 'comes a girl'), and subsequently refers to the two of them as a plural subject (*a lo que Ø están pasando juntos* 'when (they) are passing by each other').

(95) First group mention

Ø *está yendo en la bicicleta,*

eh,

.. *en el que misma --*

en el mismo camino,

pero en el sentido contrario,

pasa una niña,

en otra bicicleta.

.. (H) *a lo que Ø **están pasando** juntos,*

eh,

.. *el sombrero del niño vuela,*

(H) *y el niño,*

por regresar a ver el sombrero,

no ve una piedra grande y se choca.

`(he) is going on the bicycle,

eh,

.. on the same one that --
on the same road,
but on the other side,
comes a girl,
on another bicycle.
.. (H) when (**they**) are passing by each other,
eh,
.. the boy's hat flies,
(H) and the boy,
upon looking back at the hat,
doesn't see a big rock and he crashes.' (PS100 131-143)

In sum, the only genre difference noted is in the modification of members of a set, which mirrors the differential treatment of new referents in each genre. The higher rate of these referents in the narratives has more to do with the design of the Pear Film, where there are several groups (the baskets, the pears, the three boys) that are alternately mentioned as groups or individuals. The “members of a set” show a tendency to be treated as new referents in each genre, whereas renaming and “first group mention” are nearly always treated as given/identifiable referents.

6.3 Summary

In this chapter, the linguistic patterns used in introducing and tracking referents in two registers were explored. First and subsequent mentions were found to differ from each other in form, distribution across syntactic roles, and frequency, showing ways in which speakers attend to information flow. Important genre differences were also uncovered.

Given and new referents differ in overall frequency, with given NPs occurring at a rate of more than five times as often as new NPs in the narratives (83% of all referents in this genre are given), and over twice as frequently than new referents in the conversations (71%). This disparity is best understood with respect to the relative importance of referent tracking, which is most useful in a narrative context than in conversations. Although narrative sequences are present in the conversations, they tend to be shorter and less frequent, resulting in fewer referents to be introduced.

The data also show marked differences in the preferred grammatical roles of new and given referents. Regarding new mentions, referents are unlikely to be introduced as subjects of transitive verbs (only 7% of new referents in each genre occurred in the A role), but as they persist in the discourse, they often go on to occur in this syntactic role (26% of given referents in the narratives 15% of given

referents in the conversations are subjects of transitive verbs). In addition, many more new referents occur as direct objects (35% in both the narratives and the conversations) than do given referents (20% in the narratives and 15% in the conversations), reflecting the presentational use of the direct object role. On the other hand, there are more given indirect objects (11% in the narratives and 13% in the conversations) than new indirect objects (no new indirect objects in the narratives, and only 3% of the new referents are indirect objects in the conversations), which is likely related to information flow constraints that restrict the amount of information contained in each clause.

Lastly, Table 56 and Table 57 compare the proportion of referents according to information status and syntactic role in each genre. When compared to the overall proportion of new referents, three syntactic roles (direct objects, syntactic role “Other” and continuation of previous syntactic frame) show higher than expected proportions of new activation status. On the other hand, subjects of transitive verbs (A) and indirect objects show lower than expected proportions of new referents. A few genre differences also emerge in these tables with respect to information status and syntactic role. In the narratives, subjects of intransitive verbs (S) show a slightly higher than expected proportion of being new, but are less likely to introduce new referents in the conversations. This may reflect the preferred presentational strategies in each genre (specifically, the use of motion verbs to introduce new referents in the narratives, cf. 6.2.1). Additionally, this shows that no syntactic role is exclusive to given or new referents, but that certain syntactic roles (direct object, “Other”, continuation of previous syntactic frame and subject of intransitive verbs [only for narratives]) are more conducive to new referents, whereas given referents are more likely to appear in other syntactic roles (subject of transitive verb, indirect object, and subject of intransitive verbs [in conversations]).

Regarding the subjects of intransitive verbs, there is a clear difference between the genres. There are more new than given subjects of intransitive verbs (S) in the narratives (52% of new referents are S, and only 40% of given referents occupy the same syntactic role), yet this trend does not hold for the conversations, in which there are more given referents in the S role than new (51% versus 34%). This difference in preferred information status for S may be an indication of higher information flow pressures and narrative tasks present in the narratives, where more given referents persist longer and go on to become subjects of transitive clauses (A) in narrative sequences.

Another genre difference is evident in NPs that occur independently. The syntactic role “Other” is twice as common among new referents than given referents. This is related to the presentative function of this syntactic role, which is particularly common in the conversations.

With respect to the modification of new referents that are full NPs, another genre difference is observed. Specifically, the indefinite article is preferred for new referents in the narratives, yet the definite article is preferred in the conversations (Table 21). This difference is thought to be linked to the higher degree of shared knowledge between speakers in the conversations, and subsequent less importance of establishing identity of new entities, because they are already known to the interlocutor. When the modification of given referents is examined, there is no genre difference. The definite article is the most common form of referent modification among given referents, showing how speakers use this determiner to signal that a referent has already been identified. Bare referents (given or new) are highly uncommon in both genres.

While factors pertaining to the flow of information were shown to be of great importance, as borne out in the patterns, there were also counterexamples for almost every pattern, showing that speakers are attending to a number of different competing motivations when selecting the coding and grammar of a third person reference. This in depth exploration of these factors shows how the grammar is a manipulable resource for speakers to draw on in the construction of their message within the particular discourse context in which it occurs.

Linguistic patterns of non-referential expressions

Non-referential expressions are those that less central (yet not less important) roles in a conversation or a narrative. They flesh out the details in a narrative or conversation by helping the speaker to classify or orient referents within the discourse. They do important interactional work such as taking part in agreements, occurring in expressions of courtesy, adding emphasis or marking stance. They appear to bond to verbs in the creation of verbal predicates, where the line between word and phrase is unclear.

Non-referential expressions are particularly useful in conversations, in which they make up nearly two thirds of the data and also form an important constituent of the expressions in the narratives, comprising slightly more than one quarter of the narrative data. Recall from Chapter 5 that non-referential expressions take many forms and syntactic roles – demonstrative, interrogative and other pronouns, as well as represent most non-core arguments and all general NPs (see Table 15). This chapter will show that the linguistic coding of non-referential expressions is markedly different in a number of ways from referents, and in addition, non-referential expressions are not as homogenous of a class as referents are, which is reflected in the distribution of non-referential expressions across more syntactic roles. In order to best understand the patterns as they relate to communicative function, they are examined here according to syntactic role. Most of the comparisons in this are made across genres, but where relevant, comparisons are made between the coding of non-referential expressions and referents.

This chapter marks an important contribution towards understanding the distribution and form of non-referential NPs, which has often been neglected in previous research. As will be seen here, non-referential expressions perform very different functions from referents. This is reflected linguistically in many ways; and a thorough description of these forms is essential to a comprehensive analysis of linguistic expressions of third person.

7.1 Discourse non-referential forms in core roles

Thompson's 1997 study of the core vs. oblique distinction in English indicated that the majority of non-tracking referents fall into oblique roles, and the majority of tracking referents are core arguments. Following her observation, we can expect that non-referential forms in core roles will be less frequent than referents in core roles, and given the fact that core non-referential forms are less frequent, it may be the case that they perform somewhat specialized discourse functions. Table 35 appears to support Thompson's (1997) observation: the majority of non-referential tokens in the narratives (74%) and conversations (66%) are non-core arguments, although this may be an artifact of the coding of referentiality. However, it is important to note that a substantial number of non-referential tokens in each genre are core arguments (i.e., A, S, DO, IO), 26% in the narratives and 34% in the conversations. The proportion of non-referential tokens that are core arguments is significantly higher in the conversations (34%) than in the narratives (26%).

Table 35. Non-referential forms and the core-oblique distinction (N=2483)

	Narratives		Conversation	
	N	%	N	%
Core arguments	91	26	726	34
Non-core arguments	263	74	1403	66
Total	354	100	2129	100

Chi-square = 9.311, $p < 0.01$

As discussed in Chapter 5, non-referential core arguments are also general. An examination of the non-referential core arguments by syntactic role shows several trends. Table 36 shows that the two genres pattern the same in terms of the proportion of non-referential core arguments that fall into each of the syntactic roles (A, S and O); in each genre the highest proportion of non-referential core NPs are Direct Object, followed by S, then A, and Indirect Objects.

Table 36. Non-referential core NPs by syntactic role (N=817)

	Narratives		Conversation	
	N	%	N	%
A	1	1	131	18
S	35	38	276	38
Direct Object	55	61	287	40
Indirect Object	–	–	32	4
Total	91	100	726	100

One noteworthy difference between the genres is the frequency of non-referential A – this is nearly non-existent in the narratives (1 token, and 1% of the non-referential core NPs), yet one in five of the non-referential core arguments in the narratives falls into this role (18%).¹ The majority of non-referential subjects of transitive verbs in the conversations are unexpressed (82%, 108/131) examples of non-specific third person plural, as in (96), where the speaker is talking about being crowned as the winner of a pageant. A smaller number (7%, 9/131) are generic subjects, as in (97).

(96) Unexpressed non-referential A

A: $\emptyset$ *me van a hacer la coronación.*

A: '(they) are going to crown me.' (Deportes:135)

(97) Generic subject of transitive verb

E: *los hombres me hacen sufrir a mí.*

E: 'men make me suffer.' (Viejos Amigos:91)

This genre difference seen in the proportions of non-referential A may be related to the specific discourse needs of the participants when speaking with friends and family, as opposed to narrating. In conversations, people are more likely to make generalizations (cf. Thompson & Hopper, 2001) than they are narrating the Pear Story, and for that reason they make more use of the non-specific third person plural with transitive verbs in the conversational genre. It should also be noted that in each of the above examples, it could be argued that the first person (object) is more topical than the third person (subject). These non-specific third person subjects may be more useful in the conversations, used to relate to the first or second persons, whereas in the narratives of the Pear Story third person specific (e.g., referential) subjects are typical and there is less of a need for non-specific (e.g., non-referential) third person subjects.

Non-referential subjects of intransitive verbs occur at the same proportion in each genre, 38%. As with the non-referential subjects of transitive verbs in the conversations, the non-referential forms in the S role also serve to form generalizations, as evidenced in (98) and (99). In the narratives, a frequent function of non-referential S is to set the stage for the narrative and provide the listener with background information, as in (100).

1. While this appears to indicate that conversations are higher in transitivity than the narratives, it will later be argued that non-referential NPs are not as adequate of a measure of transitivity as referential NPs.

- (98) Non-referential full NP subject of intransitive verb

A: *pero están caras las laptop.*

A: 'but **(the) laptops** are expensive.'

(Amanda, Clases, 562)

- (99) Non-specific unexpressed subject of intransitive verb in conversation

A: *ponte que allá Ø vendan más barato.*

A: 'suppose that **(they)** sell (it) cheaper there.'

(Amanda, Clases, 706)

- (100) Non-specific unexpressed subject of intransitive verb in narrative

el canto de los --

del gallo,

me da la impresión de que Ø es eh,

.. tipo seis de la mañana.

'the song of the --

the rooster,

gives me the impression that **(it)** is eh,

.. like six in the morning.'

(PS43:12–16)

A typical function of the non-referential direct objects in both genres of these data is to form part of naming an event, activity or situation, as in (101) and (102). These direct objects commonly occur with highly frequent yet semantically 'weak' (and highly polysemous) verbs, such as *dar* ('give'), *hacer* ('make, do'), *poner* ('put') and *tener* ('have'). In fact, nearly one half (49%, 142/287) of all non-referential direct objects in the conversational data are objects of one of these four verbs, and a similar proportion (44%, 24/55) in the narratives are objects of these verbs. These are what Thompson and Hopper (2001, p. 46) refer to as "V-O" Compounds, and it could be argued that they are not in fact direct objects, but are rather part of an intransitive predicate.

- (101) Direct object naming an event in conversation

E: *.. y como no hacemos ejercicio,*

E: '.. and since we don't do **exercise**.'

(Eduardo, Food, 315)

- (102) V-O compound in narrative

.. (H) *éste bajó del árbol y Ø se dio cuenta recién en ese momento que le faltaba una cesta.*

'.. (H) he climbed down from the tree and (he) realized (lit. gave himself **account**) at that moment that he was missing a basket.'

(PS41:48–49)

Another interesting genre difference emerges when the realization of subjects is examined. The conversations show proportionally over two times as many unexpressed non-referential subjects as the narratives (19% and 45% respectively, see Table 37). The majority (70%, 128/182) of the unexpressed subjects in the conversations are non-specific third person plural, as in (103). This is a clear genre difference, as the non-specific third person plural is very common in conversation, yet is less so in narratives.

Table 37. Realization of non-referential subjects (N=443)

	Narratives		Conversations	
	N	%	N	%
Full NP	16	44	145	36
Pronoun	13	36	80	20
Unexpressed	7	19	182	45
Totals	36	100	407	100

(103) Non-specific third person plural

- A: ... *cla=ro --*
es que ahora vienen --
en una misma cámara,
o sea es ca- --
es filmadora y [cámara] digital.
- C: [mhm],
claro.
viene con todo.
- A: *pero Ø me han dicho que no es tan bueno,*
- A: \... of course --
it's that now they com --
in just one camera,
well it's ca- --
it's a video recorder and digital [camera].
- C: [mhm],
of course.
it comes with everything.
- A: but (**they**) have told me that it's not so good,'
 (Clases, 447–455)

When the realization of non-referential objects is examined, as in Table 38, we see that full NPs are the preferred form in each genre, 82% in the narratives and 70% in the conversations. This is very different than what was seen for the

non-referential subjects, where the minority were full NPs (44% in the narratives and 36% in the conversations, see Table 37). The genre difference is not significant; it is very likely that most of the non-referential objects are part of V–O compounds, in which they are unlikely to appear in reduced form.

Table 38. Realization of non-referential objects (N=374)

	Narratives		Conversations	
	N	%	N	%
Full NP	45	82	223	70
Pronoun	10	18	96	30
Totals	55	100	319	100

Chi-square = 2.717, $p = 0.09$ (not significant)

The modification of non-referential full NPs in core syntactic roles (Table 39) shows that there are no important genre differences regarding full non-referential NPs. There is a very high proportion of bare non-referential NPs in core syntactic roles (48% in the narratives and 49% in the conversations), supporting Hopper and Thompson’s (1984) observation that NPs that perform discourse functions that are not prototypical (i.e., NPs that do not track entities and are non-referential) will show less of the grammar associated with NPs (i.e., determiners). It is also here that the functional polysemy of the determiners is evident. Recall from 6.1.2 that the determiners function to facilitate the identification of referential NPs, but clearly have other uses with non-referential NPs. These uses include generic references with the definite article and non-specific NPs with the indefinite article, illustrated in (104) and (105).

- (104) Definite article with non-referential NP
A: *es que **el padrino** pone **el licor**,*
A: ‘it’s that **the godfather** supplies **the liquor**,’
(Chicas: 149)
- (105) Indefinite article with non-referential NP
E: *yo ya quiero que me den **un carro**,*
E: ‘I want them to give me **a car** already,’
(Viejos Amigos: 659)

In sum, this section on non-referential forms in core syntactic roles has shown that there are measurable differences in the distribution of tokens in each genre and that the discourse functions performed by these forms are distinct from the discourse functions of referents in the same syntactic roles. Whereas S and O are productive roles for introducing new referents and align in that sense, non-referential

Table 39. Modification of full non-referential NPs in core syntactic roles (N=429)

	Narratives		Conversations	
	N	%	N	%
Bare	29	48	181	49
Definite article	14	23	97	26
Indefinite article	5	8	47	13
Other (adjective, quantifier)	5	8	36	10
Possessive	8	13	6	2
Demonstrative	–	–	1	<1
Totals	61	100	368	100

S and O diverge in their functions. In conversations, there is more alignment in non-referential S and A (e.g., NPs in these roles are useful for formulating generalizations) and the O role is used for V–O compounds. In other words, we see both ergative (referents) and non-ergative (non-referential expressions) patterning. Between the genres, it was observed that non-referential A is nearly non-existent in the narratives, but is productive in the conversations for making generalizations. When the realization of non-referential subjects was examined, it was found that unexpressed subjects are more frequent in the conversations than in the narratives, with the third person plural impersonal subject being a very frequent construction in the conversations. These genre differences reflect the most common communicative actions of each genre – in a conversational setting speakers are more likely to make generalizations, and this contributes to higher rates of non-referential A, unexpressed third person references, whereas the primary communicative action in the narratives is to narrate the plot, with reference to specific characters (referents) and their actions. Lastly, no genre differences emerged when modification of full NPs was examined, but when compared to referential NPs, it is clear that bare NPs are much more common among the non-referential NPs than the referential NPs and the articles perform a different function with non-referential NPs than with referents.

7.2 Objects of prepositions

Of the non-core non-referential tokens, the majority in each genre are objects of prepositions (see Table 40). However, the proportion of non-referential forms that are objects of prepositions is much higher in the narrative data (88% vs. 50%). This

is a reflection of the higher information flow pressures of the narratives, as prepositions are seen as a grammatical resource for speakers to consolidate high amounts of information into a smaller number of units (cf. Biber, 1988; Chafe, 1982, 1985).

Table 40. Non-core non-referential forms and syntactic role (N = 1647)

	Narratives		Conversation	
	N	%	N	%
"Other"	11	4	222	16
Continuation of previous syntactic frame	2	1	37	3
Prepositional Phrase	231	88	698	50
Adverbial Phrase	11	4	118	9
Interrogative	–	–	83	6
Vocative	–	–	120	9
Predicate Nominal	7	3	107	8
Total	262	100	1385	100

These data were examined by preposition to see if different patterns of usage would emerge that would explain the difference in proportion, and the frequency of different prepositions is given in Table 41. As the table shows, the distribution of each preposition is similar between the two genres. It appears then, that the much lower rate of non-referential objects of prepositions in conversations may be attributed to two factors – lower information flow pressures (and therefore, fewer NPs modified by prepositional phrases, cf. Chafe, 1982, 1985) and the fact that non-referential forms are more widely distributed among the different syntactic roles in the conversation than in the narratives.

Table 41. Frequency of prepositions in Narratives and Conversations (N = 929)

	Narratives		Conversation	
	N	%	N	%
<i>a</i> ('to, at')	34	15	133	19
<i>con</i> ('with')	47	20	81	12
<i>de</i> ('from, of')	48	21	164	23
<i>en</i> ('in, on')	73	32	177	25
<i>para</i> ('for')	3	1	29	4
<i>por</i> ('for, by')	18	8	60	9
Other	8	3	54	8
Total	231	100	698	100

Functionally, forms that are objects of prepositions primarily serve for orientation, spatially or temporally, as in (106) and (107), but may also perform other functions, including (but not limited to) indicating accompaniment ((108)) or cause ((109)).

- (106) Spatially orienting object of preposition
*le pone en **la parte de adelante**,*
 '(he) puts it in **the front part**,' (PS13:26)
- (107) Temporally orienting object of preposition
*a **las doce del día** Ø les decimos chau chau.*
 'at **twelve noon** (we) tell them bye bye.' (Birthday:150)
- (108) Object of preposition expressing accompaniment
*.. Ø estoy con **el Rodri**,*
 '.. (I) am with **(the) Rodri**,' (Fumar:555)
- (109) Object of preposition expressing cause
 L: *el Lionel,*
mi compañero,
se ha intoxicado.
.. dos días Ø no ha venido.
*[intoxicado por **la chuleta**].*
 E: *[Ø debe ser por la comida de] ahí,*
 L: *'(the) Lionel,*
my colleague,
got sick.
.. two days (he) hasn't come.
*[poisoned by **the pork chop**].*
 E: *[(it) must be due to the food] there,'* (Cafe:1416)

Grammatically, we can see that objects of prepositions tend to be full NPs (Table 42, 90% in narratives and 88% in conversations) and that of those full NPs, most are modified by a definite marker (see Table 43; 72% in narratives, 59% in conversations). Interestingly, there does appear to be a genre difference in the proportion of full NPs in the syntactic role object of prepositions that are bare (36% in the conversations, yet only 10% in the narratives) and that are modified with an indefinite article (18% in the narratives and only 5% in the conversations).

Table 42. Realization of objects of preposition²

	Narratives		Conversations	
	N	%	N	%
Full NP	207	90	586	88
Pronoun	24	10	83	12
Totals	231	100	669	100

Chi-square = 0.488, $p = 0.4847$ (not significant)

Table 43. Modification of full NPs that are objects of prepositions

	Narratives		Conversations	
	N	%	N	%
Definite marker	167	72	362	62
Indefinite article	42	18	32	5
Bare	22	10	192	33
Totals	231	100	586	100

To summarize, genre differences are evident in the overall frequency of forms that are objects of prepositions (more common in the narratives) and in the modifiers that occur with full NPs.

7.3 Predicate nominals

Predicate nominals occur primarily with the Spanish copula *ser* (the other copula, *estar*, is more frequently found with predicate adjectives), and function to classify, define or evaluate referents, as in (110).

(110) Predicate nominal with *ser*

P: ... (3.0) *pero el Toni también,*
***buen estudiante** ha sido.*

P: '... (3.0) but (the) Toni too,
has been **(a) good student.**' (Café: 793–794)

2. Excluded from Table 42 and Table 43 are 29 tokens of NPs in lexicalized phrases (*de ley* 'for sure' [lit. 'of law'], *por favor* 'please' and *por ejemplo* 'for example'), in which no variation in form or modification is possible.

Predicate nominals are more common in the conversations than in the narratives, but do not occur very frequently overall. Predicate nominals make up only 3% (N=7) of the non-core non-referential expressions in the narratives and only 8% in the conversations (N=107). The higher frequency of predicate nominals in the conversations appears to reflect the more frequent evaluative (as opposed to narrative) communicative acts in the conversational data. Scheibman (2001, p. 81) noted that “in many approaches to linguistic analysis, third person singular past tense is considered the representative site for reporting – for narrating events and conveying information about the world and the people acting in it”; and indeed this narrative function of the third person singular is particularly evident in the Pear Film narratives in this study. She also noted, however, that in conversation, the third person singular is used differently in a corpus of American English conversations. She found that “the most frequent verb types in this corpus are relational predicates, in particular, those with third person singular inanimate subjects. Functionally, these relational clauses index attitude, and, less frequently, place... The frequency of these clause types reflects what speakers are commonly doing in conversation – they are evaluating and situating attitudes, events, and places... using these copular constructions” (2001, p. 83).

Grammatically, we can see from Table 44 that the preferred form of expression for predicate nominals is a full NP (100% in the narratives and 89% in the conversations). In terms of their modification, approximately half are modified and half are bare in each genre, with those that are modified being split nearly equally between definite and indefinite markers (see Table 45).

Table 44. Realization of predicate nominals

	Narratives		Conversations	
	N	%	N	%
Full NP	7	100	95	89
Pronoun	–	–	12	11
Totals	7	100	107	100

Table 45. Modification of predicate nominals

	Narratives		Conversations	
	N	%	N	%
Definite marker	2	29	20	21
Indefinite article	2	29	23	24
Bare	3	43	52	55
Totals	7	100	95	100

In sum, the only notable genre difference with respect to predicate nominals is the frequency with which they occur; they are more common in the conversations than the narratives, reflecting the evaluative actions more characteristic of the conversations. Grammatically, they are similar between the genres, showing a strong tendency to be expressed as full NPs, approximately half of which are bare and the remainder evenly divided into those which are definite and indefinite.

7.4 Adverbials

Forms that are syntactically found in adverbial phrases, such as (111) and (112), make up 4% of the non-core non-referential tokens in the narratives and over twice that amount (9%) in the conversations (cf. Table 40). They are realized almost exclusively as full NPs (Table 46), the majority of which are modified (Table 47).

- (111)

Adverbial NP in conversation

*las dos se fueron **este fin de semana**,*
 'both went **this weekend**,'

(Deportes:73)
- (112)

Adverbial NP in narrative

el momento que él estaba distraído recogiendo,
pasó un niño.
 'the minute that he was distracted picking (pears),
 a boy passed by.'

(PS29:27–28)

Table 46. Realization of adverbial forms

	Narratives		Conversations	
	N	%	N	%
Full NP	9	82	114	97
Pronoun	2	18	4	3
Totals	11	100	118	100

Fisher's exact test $p = 0.0824$ (not significant)

Functionally, these forms orient the listener, and for this reason it is not surprising that they pattern more like objects of prepositions (tendency to be full, definite NPs) than non-referential forms that serve other discourse functions (i.e., predicate nominals, which tend to be bare, or non-referential forms in the core-roles, which are less likely to be full NPs).

Table 47. Modification of full adverbial NPs

	Narratives		Conversations	
	N	%	N	%
Definite marker	5	56	85	75
Indefinite article	2	22	17	15
Bare	2	22	12	11
Totals	9	100	114	100

7.5 Vocatives

Unsurprisingly, vocatives are absent in the narratives, but relatively frequent, at 9% of the non-referential non-core forms (cf. Table 40), in the conversations. Vocatives are uniformly full NPs, and largely bare (Table 48), although 10% are modified by a possessive adjective, as in (113). Note that not all of the vocative NPs modified by possessive adjectives are proper nouns; there are an additional 50 bare vocatives that are not proper nouns, as in (114).

Table 48. Modification of vocatives (conversations only)

	N	%
Possessive adjective	12	10
Bare	108	90
Totals	120	100

(113) Vocative modified by possessive adjective

- F: ... *hola **mi vida**.*
cómo está,
cómo [le fue ahora]?
- A: [*hola **mi amorcito***].
bie=n,
y a ti?
- F: '... hello **my life**.
 how are you,
 how [did it go today]?
- A: [hello **my love**].
goo=d,
and you?'

(Birthday: 4–9)

(114) Bare vocative NP (not proper noun)

A: *y quién llegó primero?*
no me acuerdo,
loco.
 'and who arrived first?
 I don't remember,
dude.' (lit. 'crazy') (Fumar:201–203)

Perhaps one of the most interesting comparisons can be seen when we look not just at vocatives, but at all of the proper names in the corpus. As Table 49 shows, referential proper names are all marked with a definite marker (100%; recall that bare NPs were automatically coded as non-referential), and non-referential proper names are largely bare (75%).³ Examples (115) and (116) exemplify this difference.

Table 49. Modification of proper names (N = 202)

	Referential		Non-referential	
	N	%	N	%
Definite marker	108	100	23	25
Bare	–	–	71	75
Totals	108	100	94	100

(115) Bare non-referential name

J: *.. qué más que quieres Ø que Ø te cuente*
Rodriguito?
 J: '... what else do (you) want me to tell you
Rodriguito?' (Jugo:300)

(116) Definite referential name

J: *hoy Ø le llamé a la Alejandra,*
 J: 'today (I) called (the) Alejandra.' (Jugo:214)

7.6 Summary of non-referential forms

When non-referential forms are examined by syntactic role, it becomes clear that non-referential forms are not a homogeneous class that contrasts with referents, rather the category of non-referential forms is comprised of many subtypes that perform different discourse functions, responding to the communicative needs of the speakers.

3. Definite markers include the definite article and possessive and demonstrative adjectives.

With respect to form, we see differences among the different types. Overall, pronominalization is rare in non-referential expressions, except for subjects, where we see approximately one fifth (20%) to one third (36%, cf. Table 37) are realized as pronouns, and to a lesser degree non-referential objects (18% and 30% pronominal in the narratives and conversations respectively, cf. Table 38). An examination of rates of determiner type and presence also reveal differences between the types of non-referential expressions. On the one extreme are the vocatives, which are fully 90% bare (cf. Table 48), and on the other extreme are the adverbials, which are only 22% (narratives) and 11% (conversations) bare (cf. Table 47), and objects of prepositions (10% bare in the narratives, 36% bare in the conversations, cf. Table 43). The non-referential expressions in other syntactic roles (i.e., subjects, objects and predicate nominals) show a more even distribution between those that occur with and without determiners.

The distribution of non-referential expressions among different syntactic roles is also interesting. Two thirds of non-referential expressions in the narratives are concentrated in one syntactic role, objects of prepositions (65%), whereas the same syntactic role accounts for only one third (33% of the non-referential expressions in the conversations (cf. Table 50). The number of prepositional phrases has been linked to information flow pressures (cf. Biber, 1988; Chafe, 1982, 1985) therefore this is another indication of the higher information flow pressures of the narrative genre. Table 50 also shows that non-referential expressions have a wider range of discourse functions in the conversations (i.e., the larger number of non-referential subjects, vocatives) than in the narratives (where they mainly serve as props or to modify other nouns).

Table 50. Non-referential forms and syntactic role (N = 2483)

	Narratives		Conversation	
	N	%	N	%
A	1	<1	131	6
S	35	10	276	13
O	55	16	319	15
“Other”	12	3	240	11
Continuation of previous syntactic frame	2	<1	37	2
Prepositional Phrase	231	65	698	33
Adverbial Phrase	11	3	118	6
Interrogative	–	–	83	4
Vocative	–	–	120	6
Predicate Nominal	7	2	107	5
Total	354	100	2129	100

Genre differences

As corpus based research increasingly becomes the norm in linguistic investigations, it is important to recognize that not all corpora are created equally, and the study of different speech genres, and subsequent variation in frequency of different communicative actions between the genres, bears importance on the form and distribution that third person references (and other linguistic expressions) take, and the different patterns that are observed. For this reason, it is essential to document and explore genre differences, in order to create a framework within which results from different studies may be interpreted.

This chapter examines differences in linguistic coding of third person references between the two genres under consideration with special attention given to the particular communicative goals of each genre. First, the overall distribution of tokens by referentiality is discussed, and the results indicate that nominal references serve very different functions in each genre. The following section examines the patterns that emerge at the clausal level, or the patterns of Preferred Argument Structure, and how the differences reflect the unique constraints of each genre. The third section in this chapter explores the relationship between transitivity and genre and suggests that the Pear Film narratives are higher in transitivity than the conversations.

8.1 Distribution of referential NPs

The overview of all tokens presented in Table 51 shows that the distribution of tokens according to referentiality in the two genres is strikingly different. Slightly more than one third of the tokens in the narratives are non-referential, whereas nearly two thirds of the NPs in the conversational data are non-referential.

Table 51. Overview of NPs by discourse referentiality (N=4418)

	Narratives		Conversations	
	N	%	N	%
Referents	881	72	1113	35
Non-referential expressions	347	28	2077	65
Total	1228	100	3190	100

Chi-square = 484.795, $p < 0.0001$

Why is there a significantly higher proportion of non-referential tokens in the conversations? The much higher proportion of non-referential tokens in the conversations can be explained in part by the higher overall proportion of non-core arguments (which are largely non-referential, 86% in narratives and 91% in conversations). As seen in Table 52, there is nearly twice the rate of non-core arguments in the conversations (48%) than in the narratives (25%), reflecting the different communicative actions of this genre.

Table 52. Overall distribution of core vs. non-core arguments (N=4418)

	Narratives		Conversations	
	N	%	N	%
Core NPs	923	75	1668	52
Non-core NPs	305	25	1522	48
Total	1228	100	3190	100

Chi-square = 191.298, $p < 0.0001$

Specific non-referential uses were described in Section 6.2. Some of the frequent communicative actions in the conversations (use of vocatives for interactional purposes, expressions of subjectivity found in predicate nominals, for example) are absent or less frequent in the narratives. Non-referential uses of NPs in the narratives are more limited to specific functions such as orienting and situating referential NPs (especially as objects of prepositions). The distribution of NPs by referentiality can then be viewed as stemming from the genre differences – non-referential NPs are more common in the conversations because they are used not only to situate and orient referents, but they also perform a number of interactional functions, such as calling someone by name (vocative) or evaluating a referent (predicate nominal).

Not only are there genre differences with respect to the distribution of referential and non-referential tokens, there are also differences in the distribution of referential tokens by activation status (Table 53).

Table 53. Overview of referents by activation status (N=1861)¹

	Narratives		Conversations	
	N	%	N	%
Given	660	84	748	70
New	130	16	323	30
Total	790	100	1071	100

Chi-square = 45.614, $p < 0.0001$

Note from Table 53 that there is a significantly higher proportion of given referents in the narratives than in the conversations (84% vs. 70%). This suggests a higher rate of persistence in the narratives.² When the rates of persistence of new referents between the two genres are examined, as presented in Table 54, it becomes clear that new referents in the narratives are significantly more likely to be persistent (that is, to be referred to subsequently) than new referents in the conversations (62% vs. 46%). That is, having introduced a new referent, speakers are more likely to refer to it again with the next 10 clauses in the narratives than the conversations. However, given the low number of tokens of new referents (130 in narratives, 323 in conversations), it is not evident that the difference is sufficient to account for the much higher proportion of given referents in the narratives.

Table 54. Persistence of new referents (N = 453)

	Narratives		Conversations	
	N	%	N	%
Persistent	81	62	149	46
Not persistent	49	38	174	54
Total	130	100	323	100

Chi-square = 9.706, $p < 0.01$

1. Note that the total here is smaller than the total number of referential NPs given in Table 51. A subset of tokens were coded for different activation statuses, discussed in 6.2.2.
2. Recall from Chapter 4 that persistence was coded independently from referentiality and information flow. Therefore, persistent expressions, that is, those that were referred to within the previous or following ten clauses, may be new, given or non-referential. Not all new referents are persistent, nor are all given referents persistent (in the event that their previous or subsequent mention occurred at a distance of greater than 10 clauses). Non-referential expressions were coded as persistent or not based upon both previous and subsequent mention.

When the data are examined more closely it is evident that a particular group of referents within the narratives are referred to repeatedly. The 81 persistent new referents in the narratives are subsequently referred to again an average of nearly 10 times (806/81), while within the conversations, the average is about five times (802/149). Therefore it is not just the overall rate of persistent referents that accounts for the higher proportion of given referents in the narratives, but also the “longevity” of their topicality.³ The following example, (117), illustrates the high topicality (or persistence) of the boy who steals the pears in one of the narratives. In this example, the boy is referred to as the subject of 10 out of 11 consecutive main clauses. Compare with (118) (in which all referents are bolded), from the conversational data, where over nine clauses, there are five NPs with different referents.

(117) Topic continuity in narrative

*pasa **un niño** en una bicicleta,
 ... y=,
 ∅ pensó= .. coger .. primero una o dos peras.
 pero ∅ decidió al final,
 coger todo el canasto que estaba lleno.
 ... (2.0) ∅ cogió,
 y ∅ subió a su bicicleta,
 y ∅ se fue por el --
 y ∅ siguió su camino.
 ... (2.0) de ahí,
 bueno,
 pues,
 ... (2.0) ∅ se distrajo al ver una niña ... pasar por a- --
 ... pasar por su --
 ... (2.0) por su delante,
 ∅ se distrajo y ∅ se chocó con una piedra,
 tropezó=.. la bicicleta,
 y **él** se cayó.
 se regó todos los --
 todos los frutos,
 todas las peras,
 .. el canasto,*

3. It is important to keep in mind that this study focuses only on the third person. The rate of persistence (or topicality) of other referents (first and second person) may be much higher for conversation. However, Travis (2007) reports that there are higher rates of topic continuity for 1s in narratives than in conversations. Together with the results presented here, this suggests that there are more topic shifts in conversations than in narratives for all grammatical persons.

'a boy passes by on a bicycle,
 and,
he thought about taking first one or two pears.
 but (**he**) ultimately decided,
 to take the whole basket that was full.
 ... (2.0) he took (it),
 and (**he**) got onto his bicycle,
 and (**he**) went by the --
 and (**he**) went on his way.
 ... (2.0) from there,
 well,
 well,
 ... (2.0) (**he**) got distracted upon seeing a girl ... pass
 by a- --
 ... pass by his --
 ... (2.0) in front of him,
 (**he**) got distracted and (**he**) crashed into a rock,
 the bike.. faltered,
 and **he** fell.
 out spilled all of the --
 all of the fruits,
 all of the pears,
 .. the basket,'

(PS12: 37-58)

(118) Topic continuity in conversational data

- R: **el Diego** juega bien,
 [**el Santi** también].
- J: [**ese Diego** juega bien],
 chucha.
 el= --
 quién más,
 el Guillermo solo.
- R: **el Luis** no.
 el Luis se [da muchas vueltas].
- J: [**el Luis**] --
 <X y viste el X> **Luis**,
 Ø es alegón,
 me da iras,
 porque Ø es alegón.
- R: ah ah.
 .. es bien alegón **el Luis**,
 y **el Roro** también,
 ya estaba igual.

R: 'Diego plays well,
[Santi too].

J: [that Diego plays well],
shit.
the= --
who else,
just Guillermo.

R: not Luis.
Luis [runs around a lot].

J: [Luis] --
<X and you saw X> Luis,
(he) is a cheater,
it makes me mad,
because (he) is a cheater.

R: uh huh.
Luis really is a cheater,
and Roro too,
was the same.'

(Jugo:774-791)

Table 55 provides the overall distribution of Full NPs by activation status. It is interesting that a significantly higher proportion of full NPs in the narratives are given, which again reflects the higher information flow pressures of this genre, when speakers must use more information (i.e., longer forms) to indicate which referent they are designating. In the lower information pressure genre, the conversations, Full NPs are more common for first mentions, but with a considerable number (40%) being used for second or subsequent mention. It should also be noted that in a study of American English conversations, Sakita (2003) found that 22% of full NPs were new, and the remaining given and accessible although said study also included non-referential forms in the overall count, skewing the number of given full NPs (recall that even though non-referential NPs may be repeated in the discourse, since they have no associated cognitive file, they are neither given nor new).

Table 55. Overview of full NPs by activation status (N=787)

	Narratives		Conversations	
	N	%	N	%
Given	219	64	177	40
New	123	36	268	60
Total	342	100	445	100

Chi-square = 44.561, $p < 0.0001$

In sum, we see marked differences in the overall distribution of the tokens in the two genres. In the conversations there are more non-referential expressions, partly due to the wider range of discourse functions fulfilled by non-core arguments. In conversations there are more new referents and lower persistence of those referents as speakers switch topics (cf. Travis, 2007). The higher information flow pressures of the narratives are evident in the increase of overt NPs being used for already activated referents.

8.2 Preferred argument structure

Studies of Preferred Argument Structure across many different languages (cf. Du Bois, 1987, 2003b) have shown apparent universal tendencies. Specifically, there tends to be only one new argument per clause, and new arguments are typically found either as subjects of intransitive verbs or objects. In addition, arguments that are subjects of transitive verbs are typically not expressed as full NPs.

As seen in Chapter 6 (Table 27), 87% of new referents in the narratives are in fact found in the S or O role, but the trend is much less strong in the conversations, with only 69% of new referents falling into those two roles. The number of new referents presented in the “Other” syntactic role is surprisingly high in the conversational data (18%) (see (82) and (83)), yet low in the narratives (4%), indicating that there are genre differences with respect to preferred syntactic roles for introducing new referents. However, both genres are equal in the low rate of new A (7%), which is comparable to what other studies of Preferred Argument Structure have found (see Du Bois, 2003b, p. 70 and Ashby & Bentivoglio, 1993, p. 66 for rates of lexical A).

These findings do not invalidate the constraints that have been proposed when describing Preferred Argument Structure, rather they indicate that Preferred Argument Structure is perhaps a little narrow in focus, and by excluding non-core, referential arguments, a large piece of the picture (i.e., the high number of new referents in the “Other” syntactic role) has been missing from our understanding of Preferred Argument Structure, as Herring (1989) also suggested. Additionally, this shows the need to look beyond narrative data and include different genres in the study of PAS.

Another way of looking at PAS is to examine the distribution of the total number of referents in each syntactic role in relation to activation status. Table 56 shows the proportion of given and new referents in each syntactic role for narratives and Table 57 for conversations.

Table 56. Syntactic role and activation status in Narratives (N=786)

	New		Given		Total	
	N	%	N	%	N	%
A	9	5	171	95	180	100
S	68	20	265	80	333	100
Direct object	45	26	129	74	174	100
Indirect object	–	–	75	100	75	100
Other	5	28	13	72	18	100
Continuation of previous syntactic frame	3	50	3	50	6	100
Total	130	16	656	84	786	100

Table 57. Syntactic role and activation status in Conversations (N=1021)

	New		Given		Total	
	N	%	N	%	N	%
A	20	16	108	84	128	100
S	102	22	366	78	468	100
Direct object	105	50	106	50	211	100
Indirect object	9	9	92	92	101	100
Other	55	53	49	47	104	100
Continuation of previous syntactic frame	7	78	2	12	9	100
Total	298	29	723	71	1021	100

Here, the principles of PAS are evident in both the narratives and the conversations. In each genre a low number of arguments in the A role are used to present new information (5% in the narratives and 16% in the conversations; compare to 6% in Ashby & Bentivoglio's (1993, p. 71) study of Venezuelan sociolinguistic interviews). The S role is strikingly similar between the genres (20% new in the narratives and 22% new in the conversations; compare to 29% in Ashby & Bentivoglio's (1993, p. 71) data), yet the O role shows a considerable genre difference, with 50% of direct objects in the conversations being new, yet only 26% in the narratives (note that Ashby and Bentivoglio (1993, p. 71) found 42% new O). This suggests that while both roles are productive for introducing new referents, the larger number of new referents in the conversations is being introduced as O (and "Other"), not S, which may be partially linked to preferred presentational structures, which appear to be genre specific. Recall from Table 29 (Section 6.2.1) that the preferred verbs for introducing new referents in the narratives were

also mostly intransitive – particularly motion verbs (28% of referents introduced with a verb) and copulas (11%), although verbs of perception were also frequent (14%). On the other hand, weaker correlations with presentational structures were observed for the conversational data, but frequent presentational constructions in the conversations included more transitive constructions – although the most frequent verb class in the conversations was copulas, representing 19% of conversational new referents, the “light” transitive verbs *tener*, *hacer*, *dar* and *poner* make up a similar proportion of the new referents introduced with a verb (39/236, 17%). This split O role in the conversations also indicates is that in the conversational data, where information pressures are low, subjects of both transitive and intransitive verbs pattern similarly in their proportion of new and given referents, but in the narratives, with higher information flow pressures, the S and the O roles are more similar. This is consistent with Du Bois’ (1987) prediction that higher information flow pressures result in a stricter adherence to PAS, as evidenced in the more ergative patterning, although O’Dowd’s (1990) findings for a different genre of high information pressure (an orally-delivered professional training for paramedics) indicated that S and A can align in sharing the occurrences of given references contradict this. This highlights the complexity of genre and the importance of separating genre from other factors affecting information flow, such as information pressures.

Regarding the “Avoid lexical A” constraint (Du Bois, 2003a, p. 71) Table 58 shows that nearly a quarter of subjects of transitive verbs in each genre are lexical. While it is true that overall, most subjects of transitive verbs are not lexical, there do appear to be frequent contexts in which lexical A is permitted, which are further examined.

Table 58. Form of expression for subjects of transitive verbs (N=308)⁴

	Narratives		Conversations	
	N	%	N	%
Full NP	40	22	31	24
Pronominal or unexpressed	140	78	97	76
Total	180	100	128	100

Chi-square = 0.074, *p* = 0.7851 (not significant)

4. Table 58 shows only given and new referents in the A role; non-referential A are excluded (although these NPs may occur with non-referential O).

One recurrent context (10/31) for lexical A in the conversations is with a pronominal, given object (appearing as a clitic), as in (119). This is also seen in the narratives (15/40 lexical As follow this pattern).

(119) Lexical A with pronominal object

J: **el Roddy** me estaba entrevistando porque soy
<VOX famoso VOX>.

J: **Roddy** was interviewing me because I am
<VOX famous VOX>. (Jugo, 390)

Another common context in which a lexical A is likely to be found is when the object is non-referential, as in (120), where the NP *laptop* is non-referential, and therefore may require less attention or “work” on the part of the interlocutor to process, as no cognitive file is associated with this mention.

(120) Lexical A with non-referential object

A: **el Paco** también quiere una laptop,

A: ‘**Paco** also wants a laptop,’ (Clases, 526)

Lexical A is also frequently found in V-O Compounds (cf. Thompson and Hopper, 2001, p. 46), as in (121). This context is more common in the conversations (21/31 lexical As) than in the narratives (8/40). This reinforces the importance of distinguishing between discourse referential and non-referential NPs, and perhaps the constraint could be modified to indicate that speakers tend to avoid “more than one lexical *referential* core argument”.

(121) Lexical A with V-O Compound

... (2.0) y el niño **sigue su camino**.

‘... (2.0) and the boy **goes on his way**.’ (PS13:56)

Lastly, a closer look at the IUs containing lexical As in the narratives reveals that a significant proportion of these IUs (17/40, 43%) are truncated (see (122), the truncation is indicated by two dashes “--”). It is hypothesized that truncation, as a type of disfluency, may be another measure of high information flow pressures that occur when speakers manage the introduction and tracking of several referents in a short timespan. An additional eight of these lexical As appear in a different IU than the direct object, indicating that it is difficult for speakers to produce a lexical A and direct object in the same IU. When the proportion of lexical As that are found in truncated IUs (17/40, 43%) is compared to the proportion of other arguments in the narrative data that occur in truncated IUs (214/1188, 18%) it shows that the number of lexical As occurring in truncated IUs is significantly higher (Chi-square = 7.656, $p < 0.01$). This supports the notion that lexical As are a manifestation of higher information flow pressures.

- (122) Lexical A in truncated IU
- (H) *un niño* lleva sus frutos en --
.. en una bicicleta,
'(H) *a boy* carries his fruit on --
.. on a bike,'

(PS14: 34–35)

In sum, this section has shown that, in general, the constraints of Preferred Argument Structure are evident in both genres under consideration here, but that violations of or deviations from the proposed constraints are different between the genres. Specifically, a higher proportion of new arguments are introduced in and O and “Other” roles in the conversations than in the narratives, O and S pattern similarly in the narratives (ergative patterning) and S and A in the conversations (nominative-accusative), and the clauses that do not adhere to the “Avoid lexical A” constraint demonstrate different characteristics between the genres.

8.3 Transitivity across genres

In Chapters 5, 6 and 7 we have seen how the linguistic coding of referential and non-referential core arguments relates to information flow and other discourse goals and actions. In this section, direct comparisons between referential and non-referential expressions are made in order to examine the relationships between transitivity and genre, and transitivity and referentiality. The first two tables in this section, Table 59 and Table 60, provide different ways to model the transitivity in the two genres. The difference in transitivity among third person subjects (both referential and non-referential) between the genres can be seen in the proportion of subjects which are subjects of transitive versus intransitive verbs between the genres; this is an indication of the higher degree of agentivity of 3rd person subjects in the narratives. As Table 59 shows, a significantly higher number of subjects in the narratives are subjects of transitive verbs than in the conversations (32% vs. 26%).

Table 59. Third person referential and non-referential subjects and verb transitivity in narratives and conversations (N=1596)

	Narratives		Conversations	
	N	%	N	%
Transitive	185	32	263	26
Intransitive	385	68	763	74
Total	570	100	1026	100

Chi-square = 265.144, *p* <0.0001

It is possible that conversations and narratives are equally high in transitivity, but that transitivity is more common among first and second person subjects in the conversations than in third person. For this reason, all third person objects, regardless of the person of the subject, were examined. As Table 60 shows, the NPs and pronouns (referential and non-referential alike) in the narratives are more likely to be a direct object, which is taken to be another measure of higher transitivity (22% of all NPs and pronouns in narratives are direct objects, versus 16% in conversations). These results combined then support a higher degree of transitivity in the narratives.

Table 60. Proportion of direct objects by genre (N=4418)

	Narratives		Conversations	
	N	%	N	%
DO	275	22	506	16
Other syntactic role	953	78	2684	84
Total	1228	100	3190	100

Chi-square = 25.550, $p < 0.0001$

As Hopper and Thompson (1980) and Thompson and Hopper (2001) note, transitivity should be viewed as scalar, and the transitivity of a clause depends on a number of different factors. One of these factors is the referentiality of the arguments within the clause. Hopper and Thompson state (1980, p. 253) that referential objects are more highly affected (i.e., an indicator of higher transitivity) than non-referential objects. It follows that the referentiality of the subject of the clause may also have some bearing on the degree of the transitivity of the clause; specifically, a referent in the agent role likely to be an indicator of a higher degree of transitivity. Table 61 and Table 62 show that the arguments in transitive clauses (both the As and the Os) in the narratives are much more likely to be referents than in the conversations, but this is in part because the narratives are overall higher in referents than the conversations (71% of tokens in narratives are referential, 33% of all tokens in conversations are referents). However, the conversations are clearly lower in transitivity, with less referents in A and O (50% and 43% referential) than S (64% of S are referents, see Table 62), again supporting the hypothesis that the narratives are higher in transitivity, and by extension, information flow pressures.

8.4 Summary

Throughout the analyses presented in this book, attention has been drawn to several genre-specific differences. Many of the observed genre discrepancies are

Table 61. Referentiality in transitive clauses in narratives (N=923)

	Referents		Non-referential expressions		Total	
	N	%	N	%	N	%
Subject of transitive verb	184	99	1	1	185	100
Subject of intransitive verb	350	91	35	9	385	100
Direct object	220	80	55	20	275	100
Indirect object	78	100	–	–	78	100

Table 62. Referentiality in transitive clauses in conversations (N=1668)

	Referents		Non-referential expressions		Total	
	N	%	N	%	N	%
Subject of transitive verb	132	50	131	50	263	100
Subject of intransitive verb	487	64	276	36	763	100
Direct object	219	43	287	57	506	100
Indirect object	104	76	32	24	136	100

linked to varying degrees of information flow pressures and transitivity. Other observed differences are more accurately reflections of rates of types of communicative actions between the genres (i.e., more proper nouns in the conversations, more referents in the narratives), than genre differences themselves. The accurate interpretation of the patterns borne out in the data relies on the principle that speakers create grammar online, while attending to linguistic and extralinguistic cues that affect their choice of form and packaging of third person references.

A number of formal differences were observed that appear to reflect the higher information flow pressures and how speakers attend to this genre difference. The much higher proportion of referents (to non-referential expressions) in the narratives (see Table 51) reflects the speaker’s discourse goals in the narratives – to introduce and track entities through the text, and the lower rate of referents in the conversations reflects the lower information pressures of this genre. In the conversational data speakers use third person expressions for many purposes that are not related to the introduction and tracking of discourse entities. This fundamental genre difference is also seen in the rates of persistence of referents across the genres. New referents are significantly more likely to be persistent in the narratives than in the conversations (Table 54), although there is a lower overall proportion of new referents in the narratives than in the conversations (Table 53),

and more given referents in the narratives. This is linked to the greater topic continuity of narratives than conversational data (cf. Travis, 2007). It was also seen that resources for tracking given referents, such as renaming them and referring to members of a set are more productive in the narratives than in the conversations (Table 32), and that the need for more condensed “packaging” of the information in the narratives is evidenced in the higher proportion of third person referents that are modified by relative clauses in the narratives (see Table 23).

Argument structure and the distribution of arguments across different syntactic roles are also different between the genres. The preferred syntactic role for new referents is different – nearly one out of five new referents in the conversational data is introduced in the “Other” syntactic role, whereas the subject of intransitive verb (52%) and direct object roles (35%) are preferred in the narratives. The “Other” role proves more productive for given referents in the conversations (7%) than in the narratives (2%; see Table 28) as well. Section 8.2 showed that the narrative data demonstrate a much stricter adherence to PAS in terms of lower proportions of both New and Lexical A. In addition, ergative patterning (similarities between S and O roles) is evident in the narratives, but nominative/accusative patterning (similar S and A) is evident in the conversations. This suggests that under higher information pressures, the S role is more productive for introducing new information than under lower information pressures.

As we saw in Chapter 6, the presentative verbal lexicon is different between the genres. Copulas are more frequently used as presentatives in the conversations; whereas motion verbs are the most common presentative verbs in the narratives (see Table 29). When compared with the overall frequency of the verb classes in those genres, the data showed that copulas and perception verbs present significantly more discourse new referents in the narratives (as opposed to occurring with already mentioned referents), and *haber* and *tener* in the conversations, although these are not the most common verbs for introducing new referents.

The data also reveal differences in the frequency at which different types of referents are found in each genre. The increased shared knowledge and shared frames or schema among participants in the conversations is seen in the higher proportion of discourse new referents modified by definite articles in the conversations (63% versus 33% in the narratives); this reflects the type of referents present in the genre – specific, known, people (first names are often modified by the definite article in this variety of Spanish, and mentions of relatives are also often accompanied by the definite article), and the higher rate of the indefinite article with new mentions (43% of new referents in the narratives are introduced with the indefinite article, vs. 13% in the conversations, see Table 21).

The patterns uncovered in the data are also indicative of differences in transitivity. Given the higher number of given third person referents in the A and O

roles in the narratives (see Table 28), the narrative data exhibit a higher degree of transitivity than the conversations. In sum, the different discourse needs and pressures that result from the two different genres lead to a number of measurable differences in the rates at which different grammatical resources are used (i.e., relative clauses, different argument structures, definite marking, etc.) and the rate at which third person referents occur and persist; however there is no evidence that the linguistic conditioning of forms (i.e., the uses of full NPs vs. pronouns, use of articles) differs from genre to genre.

Conclusions

In seeking the answers to the research questions proposed in Chapter 2, a number of interesting findings were reported in addition to the findings that address the original research questions. This comprehensive account of linguistic expressions used to designate third person in two genres of spoken Spanish highlights the relationships between form and function, particularly as they relate to cognitive (the management of information flow and referent tracking) and interactional (turn-taking, indicators of stance, etc.) concerns. I will first summarize the major findings of this investigation and secondly will touch on areas of future research.

9.1 Summary of major findings

9.1.1 Grammar and competing motivations

The suggestion that speakers may at times struggle with competing motivations was first referred to here in Chapter 2. Throughout this investigation we saw that a variety of factors can correlate with the linguistic coding of a third person expression. There were cognitive concerns related to information status, referent introduction and tracking, and high information flow pressures; interactional concerns such as managing turn-taking and expressing stance; and there was even evidence of discourse structural concerns, such as opening and closing topics.

It is only natural that over the course of a conversation there are times when two or more of these factors come into play together. For example, a speaker might introduce a new referent and then proceed to continue to talk about the referent, but shift topics. If the factors do not coincide in their linguistic coding of a reference (in the previous example the general pattern for second mentions is not a full NP, but a full NP is often used in topic shifts), then a competition results between the different ways of coding a particular reference. Speakers then must then choose which factor exhibits stronger influence in the context and code accordingly.

Information status is always a factor relevant to the coding of referents, and identifiability or shared knowledge also proves to be an important notion, as evidenced in the differences in definite and indefinite articles with new referents

between the two genres under consideration. Theories of anaphoric coding that link referent realization with activation status (e.g., Givón, 1981) do account for a large portion of the data – 90% of new referents are coded as full NPs in the conversations and 73% of given referents are coded in lesser forms (pronoun and unexpressed mention). From this we can gather that information status is a very influential factor in the coding of the majority of referents, and therefore results in the least marked pattern of coding. There is, however, a significant number of referential tokens that are coded in ways not predicted by theories of information flow, and must be accounted for under different frameworks. Many times it is evident that local interactional concerns do override the effects of information flow on the linguistic coding of a referent. It was also shown that there are times in which the coding of the referent reflects both the interactional and information flow context of the discourse (e.g., listing new referents), so while there are two or more motivations for coding the referent in a particular way, they are not in competition.

What we, as researchers, should interpret from this is the knowledge that speakers are incredibly adept at dealing with a number of different influential forces on grammar at any one time. They rely on their past conversational experience in order to choose the grammar that best reflects what they are trying to say. Grammar is an incredibly rich and flexible resource that speakers can adapt to their immediate cognitive and interactional needs.

9.1.2 Referentiality

One of the methodological aims of this study was to operationalize coding of third person expressions for referentiality. In doing so it was found that referential uses of third person are best evidenced in the co-occurrence of several linguistic features and is not exclusively correlated with just one trait. It was shown that a number of characteristics that had previously been suggested to be indicative of non-referentiality are indeed highly correlated with non-referentiality (bare NPs, full NPs, and oblique syntactic role), but that none of these can be considered characteristic of all non-referential NPs. As Chapter 7 showed, one of the reasons that non-referentiality is not evidenced in a single criterion is because non-referential uses are not a single uniform class, rather there are many different ways speakers use expressions non-referentially. Let us consider the case of morphological marking. Recall from Chapter 5 that all bare NPs are non-referential, but only 19% (narratives) and 43% (conversations) of non-referential NPs are bare. So while it is true that if a NP is bare it is non-referential, it is not true that if a NP is non-referential, it is very likely to be bare. Non-referential expressions were shown to be a distinct category from both given and new referents in many ways – realization of non-referential

expressions is in no way constrained by previous mention, morphological marking (definite, indefinite or bare) rates are different from those seen for referents, and distribution by syntactic role is very different.

The findings of this study attest to the importance of coding tokens for referentiality in the quantitative study of language. In this study we saw how referentiality comes into play in the realization of PAS, as well as how different discourse functions are met by referential and non-referential tokens. Referentiality has been linked to transitivity (cf. Hopper & Thompson, 1980) – clauses with referential arguments are higher in transitivity than those with non-referential arguments. This is important when considering PAS, as these data revealed. In conversation especially, the constraint of “one lexical core argument per clause” could be modified to “one lexical *referential* core argument per clause”, as seen in the relative frequency of clauses with lexical A and O, where the O proved to be non-referential. This also comes into play when measuring the overall transitivity of a genre – the lower transitivity of conversations is evidenced in the lower occurrence of referential NPs, especially A and O, and the high frequency of V–O Compounds, which Hopper and Thompson classify as intransitive predicates.

We also saw how speakers make use of the patterns of referentiality for more marked discourse functions. Consider the analysis of the syntactic role “Other”, itself a marked grammatical pattern. It was shown that referents in this marked role perform functions more related to information flow needs (e.g., referent negotiation) and non-referential NPs in this more marked role are more likely to perform interactional functions ((dis)agreement, stance-taking, attitude, etc.).

9.1.3 Genre

This investigation has much to contribute to our understanding of genre differences. The task of narrating the Pear Film narratives shows evidence of higher in information flow pressures and a higher usage of transitive constructions than the communicative actions carried out in the conversations by several measures – a higher concentration of referential expressions, proportionately more arguments in the A and O roles, and the greater rate of persistence of referents.

The higher information flow pressures of such narratives have grammatical consequences, including a higher rate of referents modified by relative clauses, more objects of prepositions, and a stricter adherence to the constraints of PAS. The narratives show a more ergative structure, as reflected in the alignment of the S and O roles in introducing new information, whereas S and A are more closely aligned in the conversations.

Different discourse needs between the genres are also reflected in different rates of use of grammatical markers. For example, this is very clear in the

distribution of articles. The notion of identifiability proved to be important; the lower confidence in the amount of shared knowledge among the interlocutors in the narratives led to more indefinite first mentions in this genre. More shared experiences between the participants in the conversation appears to be linked to identifiability in the conversations, as reflected in the higher rates of definite first mentions. One example of this is seen in first names, which are common in the conversations and do not appear in the Pear Story narratives, leading to a significant number of referents being introduced with a definite article in the conversations (as first names are nearly always marked with the definite article when referential). More proper nouns (excluding first names) in conversations lead to more bare referential nouns than in the narratives.

The differential distribution of referents and core arguments between the genres can partially be understood by the wider range of functions performed by non-referential expressions in the conversations, including vocatives, interrogative pronouns, non-specific 3rd person plural subjects, indicating closed topics and regulating turn-taking. There were also differences in the presentative lexicons of the two genres, with copulas and perception verbs showing a strong presentative function in the narratives, and the verbs *haber* and *tener* were found to be strongly correlated with a presentative function in the conversations.

It was also seen that the syntactic role “Other” is much more productive in conversations than in the narratives. It plays a significant role in the introduction of new referents in the conversations in addition to fulfilling a wide range of interactional functions.

I would like to point out the important methodological and theoretical implications that arise from these findings. These data show that if “grammar” is interpreted as rates of usage, then “grammar” across genres can look very different. For instance, if one were to compare the rates of NPs that occur with and without determiners in these data, one could arrive at the conclusion that articles are not used as often in conversations as they are in narratives. Close examination of the *functions* of determiners explains why we observe the different rates: specifically the higher rate of non-referential expressions (which are much more likely to be bare than referential expressions) produces a higher rate of bare NPs in the conversational data. The researcher must be very careful not to confound rates of occurrence of grammatical forms with their linguistic conditioning; and it should be noted that no differences in the latter were observed in these data. Additionally, the genres themselves must be carefully defined. Categories such as “narrative” and “conversation” are helpful tools, but even within these categories, there are subtypes. The narratives used in this research are a particular kind of narrative and future work should compare Pear Film narratives to other types of narratives, such as narratives of personal experience, in order to draw broader conclusions.

Similarly, conversation encompasses a large number of communicative acts and the type of conversations included in this study (informal conversations between friends and family members) may not reflect the same discourse needs present in other types of conversations. As more and more corpus-based research is conducted, researchers must be aware of the ways in which genre differences are manifested in differential linguistic coding and must take care when comparing results of linguistic investigations conducted with different corpora and different genres.

9.2 Avenues of future research

This study has also drawn attention to areas of investigation that call for further research. The findings in Chapter 6 are suggestive of a possible priming effect in 3rd person subject expression, yet a much more comprehensive study which focuses solely on subjects would be necessary to either prove or disprove the existence of priming among third person subject realization.

This work has also paved the way for more studies of interactional functions of third person expressions. A larger sample, including tokens which do not display a more marked grammar could be examined in order come to a greater understanding of the factors that correlate with form and distribution of third person expressions. The study of narratives could also be broadened to include the study of how factors beyond those related to information flow affect the form and distribution of third person expressions in narratives. In addition, more attention to the role that third person expressions play in discourse structure could also prove to be fruitful in both the conversations and narratives.

References

- Ariel, M. (1990). *Accessing noun-phrase antecedents*. New York, NY: Routledge, Chapman & Hall.
- Ashby, W. J. (1995). French presentational structures. In J. Amastae, G. Goodall, M. Montalbetti, & M. Phinney (Eds.), *Contemporary research in Romance linguistics*, (pp. 91–104). Amsterdam: Benjamins. doi:10.1075/cilt.123.09ash
- Ashby, W. J. & Bentivoglio, P. (1993). Preferred argument structure in spoken French and Spanish. *Language Variation and Change*, 5, 61–76. doi:10.1017/S095439450000140X
- Ashby, W. J. & Bentivoglio, (1997). Strategies for introducing new referents into discourse: A comparative analysis of French and Spanish presentational structures. In R. M. Hammond, & M. G. MacDonald (Eds.). *Linguistic studies in honor of Bohdan Saciuk* (pp. 9–25). West Lafayette, IN: Learning Systems Inc.
- Balasch, S. (2011). Factors determining Spanish Differential Object Marking within its domain of variation. In J. Michnowicz and R. Dodsworth (Eds.), *Selected Proceedings of the 5th International Workshop on Spanish Sociolinguistics*. (pp. 113–124). Somerville, MA: Cascadilla Proceedings Project.
- Bauman, R. (1999). Genre. *Journal of Linguistic Anthropology*, 9(1–2), 84–87.
doi:10.1525/jlin.1999.9.1-2.84
- Bentivoglio, P. 1993. Full NPs in spoken Spanish: A discourse profile. In W. J. Ashby, Marianne Mithun, Giorgio Perissinotto, & Eduardo Raposo (Eds.), *Linguistic perspectives on the Romance languages* (pp. 212–224). Amsterdam/Philadelphia: John Benjamins.
doi:10.1075/cilt.103.22ben
- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511621024
- Biber, D. 1995. *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511519871
- Biber, D., Davies, M., Jones, J. K., & Tracy-Ventura, N. (2006). Spoken and written register variation in Spanish: A multi-dimensional analysis. *Corpora* 1(1): 1–37.
doi:10.3366/cor.2006.1.1.1
- Biber, D. & Finegan, E. (Eds.) (1994). *Sociolinguistic perspective on register*. New York: Oxford University Press.
- Biber, D. & Conrad, S. (2009). *Register, Genre, and Style*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511814358
- Biber, Douglas, Conrad, S. & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511804489
- Blackwell, S. E. (2003). *Implicatures in Discourse: The Case of Spanish NP Anaphora*. Amsterdam/Philadelphia: John Benjamins. doi:10.1075/pbns.105
- Bosque, I. (1989). *Las categorías gramaticales*. Madrid: Síntesis.
- Butt, J. & Benjamin, C. (2004). *A new reference grammar of modern Spanish*. 4th edition. Chicago: McGraw Hill.

- Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780195301571.001.0001
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University. doi:10.1017/CBO9780511750526
- Bybee, J. & Hopper, P. (Eds.) 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins. doi:10.1075/tsl.45
- Cameron, R. (1993). Ambiguous agreement, functional compensation, and nonspecific *tú* in the Spanish of San Juan, Puerto Rico and Madrid, Spain. *Language Variation and Change*. 5(3): 305–334. doi:10.1017/S0954394500001526
- Cameron, R. (1994). Switch reference, verb class and priming in a variable syntax. *Papers from the Regional Meeting of the Chicago Linguistic Society: Parasession on variation in linguistic theory*, 30(2), 27–45.
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects and topics. In C. Li (Ed.), *Symposium on subject and topic* (pp. 25–55) New York: Academic Press.
- Chafe, W. (1980). (Ed.) *The pear stories: Cognitive, cultural and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
- Chafe, W. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed), *Spoken and written language: Exploring orality and literacy [Advances in discourse processes]* (pp. 35–53). Norwood, NJ: Ablex Publishing Corporation.
- Chafe, W. (1985). Linguistic differences produced by differences between speaking and writing. In D. R. Olson, N. Torrance, & A. Hildyard (Eds), *Literacy, language and learning: The nature and consequences of reading and writing*, (pp. 105–123). Cambridge: Cambridge University Press.
- Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 21–51). Amsterdam/Philadelphia: John Benjamins. doi:10.1075/tsl.11.03cha
- Chafe, W. (1994). *Discourse, consciousness and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- Clark, H. & Edward, S. (1989). Contributing to Discourse. *Cognitive Science* 13, 259–294. doi:10.1207/s15516709cog1302_7
- Croft, W. (1991). *Syntactic categories and grammatical relations: The cognitive organization of information*. Chicago: University of Chicago Press.
- Croft, W. (1991). (1995). Intonation units and grammatical structure. *Linguistics* 33, 839–882.
- Croft, W. (1991). (2007). Intonation units and grammatical structure in Wardaman and English. *Australian Journal of Linguistics*. 27, 1–39. doi:10.1017/S0022226700012391
- De Mello, G. (1992). El artículo definido con nombre propio de persona en el español contemporáneo. *Studia Neophilologica* 64, 221–34. doi:10.1080/00393279208588100
- Dixon, R. (1979). Ergativity. *Language*, 55, 59–138. doi:10.2307/412519
- Donati, M. (2010). Per una teoria del vocativo. Sistema, asimmetria e persona. *Linguistica e Filologia*, 30, 11–47.
- Du Bois, J. (1980). Beyond definiteness: The trace of identity in discourse. In W. Chafe (Ed.). *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. (pp. 203–274). Norwood, NJ: Ablex.
- Du Bois, J. (1985). Competing motivations. In J. Haiman (Ed.). *Iconicity in syntax*. (pp. 343–366). Amsterdam: Benjamins. doi:10.1075/tsl.6.17dub
- Du Bois, J. (1987). The discourse basis of ergativity. *Language* 63(4), 805–855. doi:10.2307/415719

- Du Bois, J. (2003a). Discourse and grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure*, vol. 2, (pp. 47–88). Mahwah, NJ: Lawrence Erlbaum Associates.
- Du Bois, J. (2003b). Argument structure. In J. Du Bois, L. Kumpf and W. Ashby (Eds.), *Preferred Argument Structure. Grammar as architecture for function*. (pp. 11–60) Amsterdam: Benjamins. doi:10.1075/sidag.14.04dub
- Du Bois, J., Scheutze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. Edwards & M. Lampert (Eds.), *Talking data: Transcription and coding in discourse*, (pp.45–89). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Du Bois, J. & Thompson, S. (1991). *Dimensions of a theory of information flow*. Manuscript.
- Dumont, J. (2006). Full NPs as Subjects. In N. Sagarra & A. J. Toribio (Eds.), *Selected Proceedings of the 9th Hispanic Linguistics Symposium*, (pp. 286–296). Somerville, MA: Cascadilla Proceedings Project.
- Dumont, J (2013). Another Look at the Present Perfect in an Andean Variety of Spanish: Grammaticalization and Evidentiality in Quiteño Spanish. In J. Cabrelli Amaro et al. (Eds), *Selected Proceedings of the 16th Hispanic Linguistics Symposium*, (pp. 279–291). Somerville, MA: Cascadilla Proceedings Project.
- Durie, M. (2003). New light on information pressure: Information conduits, “escape valves”, and role alignment stretching. In J. Du Bois, L. Kumpf, & W. J. Ashby (Eds.), *Preferred Argument Structure*, (pp. 159–196). Amsterdam: Benjamins. doi:10.1075/sidag.14.09dur
- Dutra, R. (1987). The hybrid S-category in Brazilian Portuguese: Some implications for word order. *Studies in Language*, 11(1),163–180. doi:10.1075/sl.11.1.07dut
- Ewing, M. (2005). *Grammar and inference in conversation: Identifying clause structure in spoken Javanese*. Amsterdam: John Benjamins. doi:10.1075/sidag.18
- Flores-Ferrán, N. (2002). *Subject personal pronouns in Spanish narratives of Puerto Ricans in New York City: A sociolinguistic perspective*. Munich: Lincom Europa.
- Ford, C. Fox, B., & Thompson, S. (2002). Constituency and the grammar of turn increments. In C. Ford, B. Fox, & S. Thompson (Eds.) *The language of turn and sequence*, (pp. 14–38). Oxford: Oxford University Press.
- Ford, C. & Thompson, S. (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for turn management. In E. Ochs, E. Schegloff, & S. Thompson (Eds.) *Interaction and Grammar*, (pp. 134–184). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620874.003
- Fox, B. (1987). *Discourse structure and anaphora: Written and conversational English*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511627767
- Fox, B. & Thompson, S. (1990). A discourse explanation of the grammar of relative clauses in English Conversation. *Language*, 66(2), 297–316. doi:10.2307/414888
- Garachana Camarero, M. (2009). La creación y generalización del artículo indefinido. In C. Company Company (ed.), *Sintaxis histórica de la lengua española. Segunda parte: La frase nominal* (2 vols.). México: UNAM/FCE.
- Givón, T. (1978). Definiteness and referentiality. In J. Greenberg (ed.), *Universals of human language: Volume 4, Syntax*, (pp. 291–330). Stanford: Stanford University Press.
- Givón, T. (1981). Logic vs. pragmatics, with natural language as the referee. *Journal of Pragmatics*, 6, 81–133. doi:10.1016/0378-2166(82)90026-1
- Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.) *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam: Benjamins. doi:10.1075/tsl.3

- Givón, T. (1995). *Functionalism and grammar*. 1–42. Amsterdam: Benjamins.
doi:10.1075/z.74.02pro
- Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.
- Goodwin, C. (2007). Interactive footing. In *Reporting talk: Reported speech in interaction*, E. Holt & R. Clift (Eds.). (pp. 16–46) Cambridge: Cambridge University Press.
- Green, C. (2012). Noun phrase cohesion in English discourse: A corpus-based analysis of patterns and influences. *International Journal of English Linguistics*, 2(4), 44–57.
doi:10.5539/ijel.v2n4p44
- Gundel, J., Hedberg, N. & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274–307. doi:10.2307/416535
- Gutiérrez, M. (1995). On the future of the future tense in the Spanish of the Southwest. In C. Silva Corvalán (Ed.) *Spanish in Four Continents: Studies in Language Contact and Bilingualism*, (pp. 214–226). Washington DC: Georgetown University Press.
- Haboud, M. (1998). *Quichua y castellano en los Andes ecuatorianos: Los efectos de un contacto prolongado [Quichua and Spanish in the Ecuadorian Andes: The effects of a prolonged contact]*. Quito: Abya-Yala.
- Helasvuo, M. (1993). *Noun phrases and clauses as syntacticized discourse units in Finnish conversational discourse*. Ms. University of Santa Barbara, California.
- Helasvuo, M. (2001a.) Emerging syntax for interaction: Noun phrases and clauses as a syntactic resource for interaction. In E. Couper-Kuhlen & M. Selting (Eds.) *Studies in interactional linguistics*. (pp. 25–50). Amsterdam/Philadelphia: John Benjamins.
doi:10.1075/sidag.10.04hel
- Helasvuo, M. (2001b). *Syntax in the making: The emergence of syntactic units in Finnish conversational discourse*. *Studies in Discourse and Grammar*, John Benjamins, Amsterdam.
doi:10.1075/sidag.9
- Herring, S. (1989). Verbless presentation and the discourse basis for ergativity. In B. Music, R. Graczyk, & C. Wiltshire (Eds.) *Papers from the 25th annual regional meeting of the Chicago Linguistics Society. Part two: Parasession on language in context*. 123–37.
- Hochberg, J. (1986). Functional compensation for /s/ deletion in Puerto Rican Spanish. *Language* 62(3):609–621. doi:10.1353/lan.1986.0041
- Hopper, P. (1987). Emergent grammar. *BLS* 13, 139–157.
- Hopper, P., & S. Thompson. (1980). Transitivity in grammar and discourse. *Language* 56(2), 251–299. doi:10.1353/lan.1980.0017
- Hopper, P., & S. Thompson. (1984). The discourse basis for lexical categories in Universal Grammar. *Language*, 60(4), 703–752. doi:10.1353/lan.1984.0020
- Jefferson, G. (1990). List-construction as a task and a resource. In G. Psathas, (Ed.). *Interaction Competence*. (pp. 63–92) Washington, D.C.: University Press of America.
- Kärkkäinen, E. (1996). Preferred argument structure and subject role in American English conversational discourse. *Journal of Pragmatics*, 25(5):675–701.
doi:10.1016/0378-2166(95)00010-0
- Klein, F. (1976). “Same vs. different” cross-linguistically: The “articles” in English and Spanish. *Papers from the regional meeting of the Chicago linguistic Society*. 413–424.
- Kumpf, L. (1992). Preferred Argument Structure and second language acquisition. *Studies in Language*, 16(2):369–403. doi:10.1075/sl.16.2.06kum

- Kumpf, L. (2003). Genre and preferred argument structure: Sources of argument structure in classroom discourse. In J. Du Bois, L. Kumpf, & W. Ashby (Eds.) *Preferred argument structure: Grammar as architecture for function*, (pp. 109–130) Amsterdam: Benjamins. doi:10.1075/sidag.14.07kum
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (2014). Oral narratives of personal experience. In P. Hogan (Ed.), *Cambridge Encyclopedia of the Language Sciences*. Cambridge: Cambridge University Press.
- Lambrecht, K. (1988). There was a farmer had a dog: Syntactic amalgams revisited. *Proceedings of the Berkeley Linguistics Society*, 14, 319–39.
- Lambrecht, K. (1994). *Information structure and sentence form: topoic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620607
- Lapidus, N., & Otheguy, R. (2005). Contact induced change? Over nonspecific *ellos* in Spanish in New York. In L. Sayahi and M. Westmoreland (Eds.), *Selected proceedings of the second workshop on Spanish sociolinguistics*. (pp. 67–75) Somerville, MA: Cascadilla Proceedings Project.
- Lerner, G. (1991). On the syntax of sentences-in-progress. *Language in Society*, 20, 441–458. doi:10.1017/S0047404500016572
- Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Lipski, J. (1994). *Latin American Spanish*. London: Longmans.
- Matsumoto, K. (2003). *Intonation Units in Japanese conversation: Syntactic, informational, and functional structures*. Amsterdam: Benjamins. doi:10.1075/slcs.65
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63:81–97. doi:10.1037/h0043158
- O'Dowd, E. (1990). Discourse pressure, register, and grammatical alignment – after Du Bois. *Studies in Language*, 14, 365–403. doi:10.1075/sl.14.2.05odo
- Ono, T. & Thompson, S. (1994). Unattached NPs in English Conversation. In *Proceedings of the 20th annual meeting of the Berkeley Linguistics Society* (pp. 402–419). Berkeley: Berkeley Linguistics Society.
- Ortiz Ciscomani, R.. (2009). La creación y generalización del artículo definido. In Concepción Company Company (Ed.), *Sintaxis histórica de la lengua española. Segunda parte: La frase nominal* [Historic syntax of the Spanish language. Second part: The noun phrase] (2 vols.). México: UNAM/FCE.
- Paredes Silva, V. (1993). Subject omission and functional compensation: Evidence from written Brazilian Portuguese. *Language Variation and Change*, 5(1): 35–49. doi:10.1017/S0954394500001381
- Prince, E. (1981). Toward a typology of given-new information. In P. Cole (Ed.) *Radical pragmatics*. (pp. 223–55). New York: Academic Press.
- Prince, E. (1992). The ZPG letter: Subjects, definiteness and information-status. In W. Mann & S. Thompson (eds.), *Discourse description: Diverse linguistic analyses of a fund-raising text*. (pp. 295–325). Amsterdam: John Benjamins. OI: 10.1075/pbns.16.12pri
- Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735. doi:10.1353/lan.1974.0010
- Sakita, T. (2003). Preferred argument structure for conversational reporting. *Doshisha studies in language and culture*, 6(2):275–297.

- Sánchez-Ayala, I. (2003). Constructions as resources for interaction: Lists in English and Spanish conversation. *Discourse Studies*, 5, 323–349. doi:10.1177/14614456030053003
- Sánchez-Ayala, I. (2005). *The Spanish presentative verbal lexicon*. Unpublished manuscript.
- Schegloff, E. (1982). Discourse as an interactional achievement: Some uses of “uh huh” and other things that come between sentences. In D. Tannen (Ed.), *Analyzing Discourse: Text and Talk*, (pp. 71–93). Georgetown: Georgetown University Press.
- Schegloff, E. (1987). Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organisation. In G. Button & J. Lee (Eds.) *Talk and social organization*, (pp. 70–85). Clevedon, England: Multilingual Matters.
- Schegloff, E. (1996). Turn organization: one intersection of grammar and interaction. In E. Ochs, E. Schegloff, & S. Thompson (Eds.) *Interaction and grammar*, (pp. 52–133). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620874.002
- Scheibman, J. (2001). Local patterns of subjectivity in person and verb type in American English conversation In J. Bybee & P. Hopper *Frequency and the emergence of linguistic structure*, (pp. 61–89). Amsterdam: John Benjamins. doi:10.1075/tsl.45.04sch
- Silva-Corvalán, C. (1994). *Language contact and change: Spanish in Los Angeles*. Oxford: Clarendon Press.
- Silva-Corvalán, C. (2001). *Sociolingüística y pragmática del español*. Washington DC: Georgetown University Press.
- Tannen, D. (2007). *Talking voices: Repetition, dialogue, and imagery in conversational discourse*. 2nd edition. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511618987
- Tao, Hongyin. 1992. NP intonation units and referent identification. In Proceedings of the 18th annual meeting of the Berkeley Linguistics Society, 237–247. Berkeley: Berkeley Linguistics Society.
- Thompson, S. (1984). Subordination in formal and informal discourse. *Georgetown University Round Table on Languages and Linguistics*, 85–94.
- Thompson, S. (1997). Discourse motivations for the core-oblique distinction as a language universal. In A. Kamio (Ed.), *Directions in functional linguistics*, (pp.59–82). Amsterdam. doi:10.1075/slcs.36.06tho
- Thompson, S. (2002). “Object complements” and conversation: Towards a realistic account. *Studies in Language*, 26(1), 125–164. doi:10.1075/sl.26.1.05tho
- Thompson, S. & Hopper, P. (2001). Transitivity, clause structure and argument structure: Evidence from conversation. In J. Bybee & P. Hopper (Eds.) *Frequency and the emergence of linguistic structure*, (pp. 27–60) Amsterdam: Benjamins. doi:10.1075/tsl.45.03tho
- Tomlin, R. (1987). Linguistic reflections of cognitive events. In R. Tomlin (Ed.) *Coherence and grounding in discourse*, (pp. 455–480). Benjamins: Amsterdam. doi:10.1075/tsl.11.20tom
- Toribio, A. (1993). *Parametric variation in the licensing of nominals*. Ph.D. Dissertation, Cornell University, Ithaca, New York.
- Toribio, A. (1996). Dialectal variation in the licensing of null referential and expletive subjects. In C. Parodi, C. Quicoli, M. Saltarelli & M. Zubizarreta (Eds.) *Aspects of Romance linguistics: Selected papers from the Linguistic Symposium on Romance Languages XXIV March 10–13, 1994*, (pp. 409–432). Washington D.C.: Georgetown University Press.
- Toribio, A. (2000). Setting parametric limits on dialectal variation in Spanish. *Lingua*, 110(5), 315–341. doi:10.1016/S0024-3841(99)00044-3
- Torres Cacoullos, R. & Aaron, J. (2003). Bare English-origin nouns in Spanish: Rates, constraints, and discourse functions. *Language Variation and Change*, 15, 289–238. doi:10.1017/S0954394503153021

- Torres Cacoullos, R. & Travis, C. (2015). Foundations for the study of subject pronoun expression in Spanish in contact with English: Assessing interlinguistic (dis)similarity via intra-linguistic variability. In A. Carvalho, R. Orozco & N. Lapidus Shin (Eds.) *Subject pronoun expression in Spanish: A cross-dialectal perspective*, (pp. 169–188). Georgetown University Press.
- Travis, C. (2005). The yo-yo effect: Priming in subject expression in Colombian Spanish. In R. Gess & E. Rubin (Eds.), *Theoretical and experimental approaches to Romance linguistics: Selected papers from the 34th Linguistic Symposium on Romance Languages, 2004*, (pp. 329–349). Amsterdam/Philadelphia: John Benjamins. doi:10.1075/cilt.272.2otra
- Travis, C. (2007). Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change*, 19(2), 101–135. doi:10.1017/S0954394507070081
- Trudgill, Peter. (1974). *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Appendix

Transcription symbols (Du Bois et al. 1993)

(H)	inhalation
(Hx)	exhalation
@	laughter (each “@” represents one “pulse” of laughter)
--	truncated IU
-	truncated word
.	final intonation
,	continuing intonation
?	rising intonation
..	short pause
...	long pause
...()	measured pause
[]	speech overlap
=	lengthened syllable
<X X>	unclear speech

Index

A

Activation status 15–19, 51–52, 149
Adjacency pair 32
Anaphora 31–32

B

Bare NPs 72

C

Classifying NP 58–59
Competing motivations 14, 92–101, 130, 163–64
Core syntactic roles 74–77, 132, 136

D

Definite article 101–102, 107, 113
Definiteness 47, 101
Demonstrative pronoun 66–70
Discourse analysis 1–2
Discourse referentiality 19–23, 56–61, 81, 164–165

F

Free NPs 11, 49, 78, 115–117, 153
Full NPs 66

G

Genre 3–5, 29, 166–167
Genre effects 4, 30, 122, 133, 135, 140, 147–161, 165–167

I

Identifiability 23–26, 101–102, 106
Increments 32
Indefinite article 102, 107
Indefiniteness 47, 101
Information flow 15
Information flow pressures 6, 29, 35, 112–114, 138, 152, 165
Intonation unit 38–39

L

List construction 32–33, 99

N

Non-referential expressions 10–11, 148

O

Obliques 74–77
Orienting NP 58–59, 139

P

Pear Film 36
Persistence 63–66, 149
Predicating NP 58–59, 140–141
Preferred argument structure 14, 29, 47, 115, 153–157
Priming 90–92
Pronouns 66, 70–71

Proper names 66, 103

Q

Quiteño Spanish 35

R

Referential distance 55
Referential expressions 10–11
Relative clauses 47
Repetition 92–93

S

Semantic class of verb 50
Specificity 27–28, 52–53, 73–74

T

Transcription method 38–39
Transitivity 157–158
Turn-taking 2

U

Unexpressed mentions 66, 70, 135

V

V-O compounds 134
Vocatives 66, 76, 143–144

This volume, a case study on the grammar of third person references in two genres of spoken Ecuadorian Spanish, examines from a discourse-analytic perspective how genre affects linguistic patterns and how researchers can look for and interpret genre effects. This marks a timely contribution to corpus linguistics, as many linguists are choosing to work with empirical data. Corpus based approaches have many advantages and are useful in the comparison of different languages as well as varieties of the same language, but what is often overlooked in such comparisons is the genre of language under examination. As this case study shows, genre is an important factor in interpreting patterns and distributions of forms.

The book also contributes toward theories of anaphora, referentiality and Preferred Argument Structure. It is relevant for scholars who work with referentiality, genre differences, third person references, and interactional linguistics, as well as those interested in Spanish morphosyntax.

“In this very thorough and detailed analysis, Dumont shows that third-person linguistic expressions are motivated by cognitive, interactional and discourse structural factors which at times act in tandem while at others in competition. This book is not only an important contribution to our knowledge of Spanish but more generally to the field of usage-based grammar.”

Javier Rivas, *University of Colorado Boulder*

ISBN 978 90 272 1581 9



John Benjamins Publishing Company