

# Referential Distance, Demonstrative Anaphors and the Current Focus of Attention

Iker Zulaica Hernández ([ikerzulaica@gmail.com](mailto:ikerzulaica@gmail.com))

Department of Spanish and Portuguese, 298 Hagerty Hall  
The Ohio State University  
Columbus, OH 43210 USA

## Abstract

In this paper, I investigate various factors that may affect the referential properties of referring expressions in general and demonstratives in particular. To that aim, factors like referential distance, antecedent type or animacy have been tested in a corpus of Spanish. My initial hypothesis is that demonstrative anaphors, rather than elements that mark a certain degree of accessibility or the cognitive status of their antecedents, should be better conceived as elements used by speakers to mark topic or subtopic shifts in discourse. This is accomplished by focusing the hearer's attentional state on specific discourse referents via demonstrative anaphors. By using this strategy, speakers would make hearers aware of a change in the general or local topic of discourse. The referential properties and distribution of demonstrative anaphors and the personal pronoun 'lo' were compared based on these textual and semantic factors. The corpus data indicate that there are no significant differences between demonstrative anaphors and the personal pronoun with respect to the factors scrutinized. Based on this evidence, I conclude that demonstrative anaphors actively contribute to information structure by encoding a topic marking procedure. Likewise, I believe that notions like accessibility or cognitive focus should be reconsidered to characterize the degree of accessibility or givenness that discourse entities have after the speaker's goal has been achieved.

**Keywords:** accessibility, anaphora, demonstratives, discourse, focus, saliency, reference.

## Introduction

Despite the efforts of many researchers and the progress accomplished over the last decades, pronominal and adnominal demonstrative anaphors (this/that – this/that-NP) still pose interesting questions for those working on referential expressions. Thus, for example, we may ask ourselves whether demonstrative anaphors and deictic demonstratives still have some common features or, conversely, they have become totally separated linguistic elements through the history of language. If, as it has been suggested (Diessel 2006), the demonstrative anaphoric mode has evolved from the more 'primitive' deictic mode then we are left with the task of explaining the manner in which demonstrative anaphors embody this deictic component. Conversely, if we are willing to accept that no trace of deixis is left in demonstrative anaphors we will need to provide a suitable explanation for uses of demonstrative anaphors that closely resemble those of deictic demonstratives (e.g. focusing specific discourse referents, pure discourse deixis, etc.). But even if we accept that no deictic-like component *per se* has been retained in demonstrative anaphors we still may wonder whether the important pragmatic role of

demonstratives in deictic uses (i.e. that of making salient an entity in the speech situation via a pointing gesture, contributing to the joint focus of attention of the interlocutors or even referential disambiguation) has been retained in the anaphoric mode in some way. The relevance of these questions is twofold. On the one hand, succeeding to provide a satisfactory answer to this issue would greatly contribute to the studies of demonstrative expressions in general. On the other hand, it may have an impact on the way we understand referring expressions and the mainstream theories on these elements.

In this paper, I'll argue that demonstrative anaphors still retain a procedural component which is very likely derived from the deictic component of deictic demonstratives (i.e. a by-product of the basic pointing mechanism) whereby discourse entities referred to are rendered into the current focus of attention of the hearer in the utterance situation. Since demonstrative anaphors do not retain an overt pointing device like purely deictic elements do, the way we can characterize this procedure is not at all trivial, but data from Spanish and other languages appear to point in this direction. The procedure, in my opinion, should be conceived as an instruction by part of the speaker for the addressee to focus on a particular discourse entity and with a precise communicative intention, i.e., making the interlocutors aware that a topic shift is taking place or a new local subtopic has been introduced. Evidence in support of my hypothesis comes from various factors possibly affecting the different distribution and referential properties of referring expression. The factors, which have been tested in a corpus of Spanish, are referential distance, semantic and syntactic type of the antecedent, animacy, position of the antecedent within the clause and a language-specific combination consisting of a demonstrative anaphor and a personal pronoun both co-occurring within the same clause, next to one another and being coreferential (*eso lo*). Overall, the results from the corpus show that there are no significant differences between demonstratives and the Spanish personal pronoun *lo* (it) as regards any of the factors examined. Based on this evidence, I'll argue that extensively used notions like accessibility, givenness or cognitive focus should be reconsidered to characterize the degree of accessibility or cognitive status of discourse entities once the speaker's goal has been achieved with the use of a referring expression. In other words, I think it would be convenient to make a difference between the degree of accessibility or the cognitive status a particular discourse referent has prior to the use of a particular referring expression, and the accessibility or cognitive status it will have after the same expression has been used and the

communicative goal has been achieved. This is particularly relevant as regards anaphoric uses of referring expressions that presuppose the existence of a textual antecedent.

### Accessibility and the Current Focus of Attention

Different hierarchical scales have been proposed to account for the different distribution shown by the range of referring expressions across languages. I will primarily deal with the Accessibility Scale (Ariel 1988, 1990) and the Givenness Hierarchy (Gundel et al. 1993) (henceforth AS and GH, respectively) as concepts like cognitive status and degree of accessibility will be profusely used in this paper. In Ariel’s theory, accessibility can be defined as the relative ease with which the addressee can identify the referent of a referring expression or, alternatively, the ease with which the addressee can retrieve the intended referent from memory. Ariel’s AS is shown in (1).

- (1) Extremely High Accessibility Markers (gaps, including pro, PRO and *wh* traces, reflexives and Agreement) > Unstressed pronoun > Cliticized pronoun > Stressed pronoun + gesture > Stressed pronoun > Proximal demonstrative (+ NP) > Distal demonstrative (+NP) > Proximal demonstrative + modifier > Distal demonstrative + modifier > First name > Last name > Short definite description > Long definite description > Full (‘namy’) name .

According to the scale, demonstratives occupy an intermediate position in terms of the degree of accessibility they confer to their referents. The scale indicates that proximal demonstratives mark higher accessibility than distal demonstratives. Note that the amount of extra information that is attached to a demonstrative results in the demonstrative marking lower accessibility. Thus, when a modifier accompanies demonstratives they rank lower in the scale (e.g. that house vs that house next to the bus stop). This is explained for the more accessible a referent the lesser information is needed to for the addressee to identify it. As the scale clearly indicates, the less informative (null) forms (gaps, PRO, etc) occupy the highest position, that is, they are high accessibility markers. Unstressed and cliticized pronouns also occupy a high position in the AS.

Table 1: The Givenness Hierarchy

<b>In focus &gt;</b>	<b>Activated &gt;</b>	<b>Familiar &gt;</b>	<b>Uniquely Identifiable &gt;</b>
<i>it</i>	<i>HE, this, That, this N</i>	<i>That N</i>	<i>The N</i>
<b>Referential &gt;</b>	<b>Type Identifiable</b>		
<i>Indefinite This N</i>	<i>a N</i>		

The six cognitive states proposed (table 1) and their corresponding pronominal or determiner forms are intended to ‘indicate’ that the referent of the nominal expression is assumed by the speaker or writer to have a particular cognitive status (memory and attentional state) for the addressee. Thus, a correlation is proposed between each cognitive state and one or more linguistic elements in such a way that use of a particular anaphor by the speaker would allow the hearer to restrict the set of possible antecedents, hence facilitating anaphora resolution. The GH is an implicational hierarchy, what means that whenever the speaker uses a specific linguistic form he/she would be implying all other states ranking lower in the hierarchy. Another factor playing a role in the Givenness Hierarchy is Grice’s (1981) maxims of quantity:

- **Q1:** Make your contribution as informative as possible
- **Q2:** Do not make your contribution more informative than necessary

The maxims of quantity associated to the hierarchy prevent higher referring expressions from being used to refer to entities with a lower status.

For the purposes of this paper, I’ll only discuss the appropriateness of the IN FOCUS, ACTIVATED and, only occasionally, FAMILIAR statuses as these are the cognitive statuses proposed for demonstrative elements and the unstressed personal pronoun in the GH. Being ACTIVATED for a referent means that, at a given point in the discourse, there must be a representation of the referent in short-term memory. On the other hand, being IN FOCUS means that the referent is not only in short-term memory but also at the current center of attention. As Gundel et al. pointed out, at a given moment in discourse IN FOCUS entities are the partially ordered subset of ACTIVATED entities that are more likely to be the topic in subsequent discourse. The following two examples may serve to illustrate the difference between IN FOCUS and ACTIVATED entities as characterized in Gundel et al.<sup>1</sup>. Note how different contexts favor the use of one anaphor over the other.

- (2) [Mary brings a large packet into the room. Everyone stares at the package as it starts to tick and rock back and forth.]  
John: IT’s going to explode!
- (3) [Mary brings a large packet into the room. Only John notices as the package starts to tick and rock back and forth.]  
John: THAT’s going to explode!

In a subsequent paper, Gundel et al. (2003) analyzed the behavior of English demonstratives ‘this/that’ and the unstressed pronoun ‘it’. Based on evidence from a corpus these authors observed that demonstrative anaphors were

<sup>1</sup> From Wolter (2006).

used to refer to abstract entities in 85% of the cases there scrutinized, whereas only 15% of the cases were anaphorically referred to with the pronoun ‘it’. They explained this fact by assuming that material introduced in clauses (e.g. clausally introduced entities like propositions or events, which are typical antecedents for demonstrative anaphors) is ACTIVATED as compared to material introduced in syntactically prominent noun phrases, which is more likely to be IN FOCUS.

### Comparing Referring Expressions

Spanish belongs in the set of languages with tripartite demonstrative systems. The three paradigmatic demonstratives are *este* (this), *ese* (that) and *aquel* (that further), but these three basic forms inflect for gender and number to show morphological agreement with their nominal heads in adnominal uses. Regarding their referential properties, Spanish adnominal and pronominal demonstratives do not differ much from their counterparts in other languages and they can be used by speakers to refer to a range of entities in the physical or textual context depending on the mode analyzed (deictic or anaphoric). The range of entities referred to vary in their degree of concreteness or abstractness. A group of three pronominal demonstratives (*esto*, *eso* and *aquello*) have been traditionally labeled ‘neuter demonstratives’ based on their referential properties, but the fact is that there is no such neuter grammatical gender in Spanish. Thus, even though they show an invariable, typical masculine gender form (-o marker) they are most commonly used to refer to objects that denote abstract concepts like facts, events, situations, etc. These forms are roughly equivalent to the English demonstrative pronouns *this* (*esto*) and *that* (*eso*). The third form *aquello* would find its closest equivalent in the English expression *that yonder*. All demonstrative forms are shown in table 2.

Table 2: Spanish demonstratives

Gender	Number	Determiner Pronoun
Masc.	Singular	Este/Ese/Aquel (NP-sing)
Fem.		Esta/Esa/Aquella (NP-sing)
Masc.	Plural	Estos/Esos/Aquellos (NP-pl)
Fem.		Estas/Esas/Aquellas (NP-pl)
Neuter	--	Esto/Eso/Aquello

We want to pay special attention to clausal antecedents and the range of entities that they commonly introduce in the discourse as this kind of antecedent is very common in cases of demonstrative anaphora. Abstract or intensional antecedents show a high degree of syntactic polymorphism as they can be introduced in discourse via noun phrases, subordinate clauses, infinitival clauses, or more complex structures consisting of two or more sentences (see, for example, Asher (1993) for a detailed study on the range of entities that denote abstract objects from a formal semantics perspective and Eguren (1999) for a general overview on the

denotational properties of Spanish demonstratives). This point is illustrated in (4), where the medial demonstrative pronoun *eso* (that) refers back to an infinitival clause in subject position at the beginning of the paragraph: *tener una relación con un hombre*. Semantically, the presence of the predicate of events *pasar* (happen) forces an event reading of the antecedent.

(4) [Tener una relación con un hombre]<sub>j</sub> no está en mis planes por el momento ya que debo esperar a que primero sanen mis heridas. Pero mis amigas que lo conocen me dicen que no sea boba, que me dé la oportunidad ya que **eso**<sub>j</sub> pasa sólo una vez en la vida.

‘[Having a relationship with a man]<sub>j</sub> is not in my plans for the moment as I must first wait until my heart heals. But my friends tell me not to be stupid, that I should give myself a chance as **that**<sub>j</sub> only happens once in a lifetime.’

But, of course, the ability to make reference to abstract objects is not restricted to demonstrative pronouns. Demonstrative determiners also have this ability, which crucially depends on the degree of abstractness of the nominal head, e.g. *este evento* (this event), *aquel hecho* (that fact) vs *este hombre* (this man) or *esa bicicleta* (that bicycle), where the nouns ‘event’ and ‘bicycle’ could be characterized as [+abstract, -concrete][−abstract, +concrete], respectively. Other Spanish discourse anaphors also play an important role in referring to high-order entities, namely, the personal pronoun *lo* and the null subject pronoun (*pro*). Spanish *lo* is the 3<sup>rd</sup> person pronoun accusative singular masculine and neuter form. Pronoun *lo* is a non-actor argument that corresponds to English 3<sup>rd</sup> person singular pronouns ‘it’ and ‘him’. The Spanish pronoun *lo* can be used as an enclitic and proclitic as shown in (5a-b).

(5) a. Lo compré.  
ACC.3 bought-1sing.  
I bought it.  
b. Cómpralo.  
Buy-2sing-ACC.3 (command)  
Buy it!

Regarding its referential properties, Spanish *lo* can also be used to refer to a gamut of discourse entities (individuals, plural objects, events, discourse topics, etc.). Thus, for example, the pronoun *lo* (‘it’) is commonly found to corefer with propositional entities. This point is illustrated in (6), where the two occurrences of the pronoun *lo* in the second sentence refer back in the text to a propositional entity introduced via the first sentence.

(6) [El espacio me obliga a olvidar a muchos de los que asistieron]<sub>j</sub>, lo<sub>j</sub> sé y lo<sub>j</sub> siento.  
‘[Reasons of space oblige me to forget many of those who attended]<sub>j</sub>, I know ?it/that<sub>j</sub> and I’m sorry about ?it/that<sub>j</sub>.’

As a pronoun, Spanish *lo* can only have the grammatical role of direct object (DO) and, consequently, we will never find it filling the syntactic slot of a subject (SUBJ). In this regard, it differs from English pronoun *it* as this pronoun can be indistinctively used referentially as DO or as SUBJ of a sentence. Let us consider the following examples. In (7), the procliticized personal pronoun is a direct object. Compare this use with the role of DO (8a) and SUBJ. (8b) of English *it*.

- (7) A: ¿Partido fácil a priori?  
 B: Bueno, no lo creo.  
 Well, not ACC.3 believe-1sing.  
 A: ‘Would you consider it an easy match beforehand?’  
 B: ‘Well, I don’t think so.’
- (8) a. I know it when I see it.  
 b. It happened once. Could it happen again?

There is, nevertheless, a *lo* neuter definite article that can be found in sentence SUBJ position. An example of the neuter definite article is given in (9). As we won’t be dealing with definite descriptions, we won’t take this element into consideration. Suffice to say that the English counterpart to this element it is the construction featuring a definite article ‘the ADJECTIVE thing’.

- (9) Para nosotros, lo más importante es la calidad de nuestro producto.  
 ‘To us, the quality of our product is the most important thing.’

Demonstratives and personal pronoun *lo* differ in that the former can be the SUBJ of a sentence whereas the latter cannot. The verb *suced*er (happen) in (10) is a typical predicate of events, that is, it forces a reading in which its argument needs to be an eventuality (action, accomplishment, achievement, etc.), or an event type. Even when embedded within the modal operator *si* (if), the argument of the verb *suced*er is to be considered a hypothetical event. The antecedent of the demonstrative anaphor *esto* (this) in (10) is ambiguous, as it may well be the whole discourse span prior to the clause containing the anaphor or the smaller fragment *inclin*ar la balanza en dirección del presidente Bush. Either way, compare this typical referential use of the demonstrative in subject position with the ungrammatical example of propositional *lo* in (11).

- (10) Tengo que añadir el pequeño pero importante “factor Nader“, cuyo 1% en el país puede ser hasta un 3% en algún estado, e incluso, si se trata de unos escasos centenares de votos pudiera inclinar la balanza en dirección del presidente Bush. O sea no se asusten, si **esto** sucede.

‘I have to add the small but important “Nader factor“, whose 1% over the country may become up to a 3% in

some states, and even if it only amounts to a few hundred votes it could tip the balance in favor of President Bush. That is to say, do not scare if **this** happens.’

- (11) \*Lo sucede todos los días.  
 ‘?It/that happens everyday.’

The pronoun *lo* and the demonstrative pronoun can both co-occur in the same sentence and functioning as the direct object of the verb. This point is illustrated in (12).

- (12) A: No seas rencoroso. Ya sabes que el hombre cambia.  
 B: Sí, eso sí lo sé.  
 Yes, that yes ACC.3 know-1sing  
 A: ‘Don’t be so resentful. You know the man changes.’  
 B: ‘Yes, I know that.’

Finally, a null pronoun can also be used referentially in Spanish to refer to propositional entities. This is the case of the null pronoun in (13) (null =  $\emptyset$ ), where it appears to be referring back to the proposition *antier se ahogó un niño allá*. This phenomenon, though, is very uncommon and apparently restricted to certain varieties of the Spanish-speaking world (Reig 2008)<sup>2</sup>.

- (13) Bueno sí, sí ha habido [accidentes], nada menos antier se ahogó un niño allá, allá, allá ¿sí  $\emptyset$  sabías?  
 ‘Well, yes, yes there have been some [accidents], the day before yesterday a body drowned there. Did you know ? $\emptyset$ /it/that?’

## The Role of Information Structure

The primary goal was to test our hypothesis in a corpus of Spanish texts. The corpus used for this study (CREA)<sup>3</sup> has clear advantages as a source of data, but disadvantages too. As advantages we may list a huge language database comprising all language varieties, text types and genres, the possibility to retrieve a piece of text long enough to find the intended antecedent in almost 100% of cases and the ability to narrow down the initial results to reduce the samples to the researcher’s needs. The main disadvantage is that the CREA corpus is a non-annotated corpus so all cases must be carefully scrutinized to find out the intended antecedent of the referring expression. The undesired consequence of that is the potential subjective bias of the researcher towards the cases analyzed.

Regarding the methodology followed to gather the data I carried out several corpus searches to test various factors that may have an effect on the referential behavior of the referring expressions under consideration. One of these

<sup>2</sup> This example is from Reig (2008).

<sup>3</sup> The CREA corpus of Spanish is a very large collection of texts. A dedicated search interface allows the user to search the corpus for words and phrases and display the search result as a concordance with limited context (the sufficient amount of context for the purposes of this paper.)

factors is referential distance, that is, the textual distance of the antecedent relative to the position of the referring expression in the discourse span. Referential distance has already been considered as a factor possibly influencing the degree of accessibility of different referring expressions (see Maes & Noordman 1995 and Ariel 2001 for discussion on this topic).

For the purposes of this paper, the unit of measure to test referential distance will be the clause so all the cases analyzed were segmented into clauses as in example (14). Each clause was given a clause number from  $CL_0$  (the one containing the referring expression) to  $CL_{\geq 4}$ . The number assigned to each clause increases as distance from the referring expression increases. Thus, the clause immediately preceding the referring expression was numbered  $CL_1$ , and so forth. In (x), the most likely antecedent is highlighted in bold characters within  $CL_1$ . Likewise, the demonstrative anaphor is given in bold characters in  $CL_0$ .

(14) P: ¿Hay muchos mitos en esta historia?  
 R: [Sí, los hay] $_{CL4}$ ; [como que Watson y Crick trabajaban en una barraca o en un cobertizo para bicicletas] $_{CL3}$ . [No es verdad] $_{CL2}$ . [Es cierto que **este grupo o cierta parte de este grupo ocupó la barraca que se convirtió en laboratorio**] $_{CL1}$ , [pero **esto ocurrió** cuatro años más tarde, en el 57] $_{CL0}$ . Es como la historia de Newton y la manzana: hace falta la barraca y hace falta la manzana.

Q: Are there too many myths in this story?

A: [Yes, there are] $_{CL4}$ ; [like that one that says that Watson and Crick used to work together in a hut or a bicycle shed] $_{CL3}$ . [That's not true] $_{CL2}$ . [It is true that **this group, or a certain part of this group, occupied the hut that eventually became a lab**] $_{CL1}$ , [but **this happened** four years later, in 1957] $_{CL0}$ . It's like the story about Newton and the apple: we need the hut as much as we need the apple.

### The Spanish personal pronoun *lo*

Two typical corpus cases of pronoun *lo* are given in (15) and (16). Note the antecedent of the pronoun is the entire preceding subscripted clause in (15). Note also the two subscripted little pros both referring to the same individual (the speaker). Conversely, the two pronouns in the last sentence of (16) are coreferential with the NP *un departamento*.

(15) [En la vida real  $pro_j$  no soy nada mentiroso] $_k$ ,  $pro_j$  no  $lo_k$  necesito.  
 'In real life I'm not a liar, I don't need ?it/that.'

(16) Tengo un pequeño terreno y una casa en 'ballenita' que la conseguí con un préstamo del seguro social; un departamento $_j$  que  $lo_j$  convertí en oficina y  $lo_j$  tengo alquilado.

'I own a little piece of land and a house in 'ballenita' that I got with a loan from the social security; an apartment $_j$  that I made (it $_j$ ) into an office and I'm leasing (it $_j$ ).'

All the cases of personal pronouns analyzed were divided into three groups corresponding to three different corpus searches: *lo entiendo* (I understand it), *lo necesito* (I need it) and *lo tengo* (I have it). 40 occurrences per group were scrutinized (120 cases in total). The three groups would allow us to test not only referential distance but also additional factors that may possibly affect the cognitive status and different accessibility marking shown by referring expressions. One of these factors is referential type, that is, the semantic type of the antecedent. Another factor is antecedent type or the syntactic type of the antecedent. By using three different predicates *entender* (understand), *necesitar* (need) and *tener* (have/own) we have tried to force different semantic readings for the antecedent insofar as this can be ascertained. Thus, the predicate *entender* (understand) would show a preference for higher order antecedents, as we normally understand concepts, ideas, hypotheses rather than concrete, physical objects. Conversely, the verb *tener* as in *lo tengo* (I have/own it) would rather show a preference for physical objects as, under normal conditions, people own physical objects rather than abstract entities. Of course, the semantic selection properties of these predicates cannot be taken categorically for it may well be the case that we don't understand (the behavior of) a particular individual or that we have, for example, some particular trait of personality. The verb *necesitar* (need) is intended to occupy an intermediate position in between the former two predicates. The aim was to obtain a sample ample enough to be able to draw some initial conclusions regarding the possible influence of the semantic nature of the antecedent.

The example in (17) shows a clear case of topic continuity. Remember that Spanish is a subject pro-drop language. There are two little pros in (17). One of them ( $pro_j$ ) serves as antecedent to the personal pronoun. The second ( $pro_k$ ) corresponds to the subject of the apodosis in the conditional construction. The referent of ( $pro_j/lo_j$ ) is an individual (very likely an NP like, for instance, 'the player' or a proper name) but its first discourse mention does not appear in our sample text fragment. The second pro has the speaker as its referent. We have found that the type of configuration whereby the antecedent of a personal pronoun is a pro is actually quite common in Spanish. As regards antecedent type, we decided to include pros in the antecedent category named 'NP'. As regards antecedent distance, we decided to split conditional constructions into its two constituent clauses (protasis and apodosis) for the purposes of clause count. In this particular example, the antecedent of the two personal pronouns is located in clause  $CL_1$  (the protasis).

(17) Si en Olimpia  $pro_j$  estuviera jugando como central izquierdo  $pro_k$  sí  $lo_j$  llamo porque es en esa posición donde  $lo_j$  necesito.

‘If he<sub>j</sub> were playing as a left mid in Olimpia I would call him<sub>j</sub> because that is the position where I need him<sub>j</sub>.’

The results of our first corpus sample are given in tables 4 and 5.

Table 3: Referential distance for accusative personal pronoun (cases)

	CL <sub>0</sub>	CL <sub>1</sub>	CL <sub>2</sub>	CL <sub>3</sub>	CL <sub>≥4</sub>
Pronoun <i>lo</i> <sub>1</sub> *	3	30	4	--	3
Pronoun <i>lo</i> <sub>2</sub> **	2	37	--	1	--
Pronoun <i>lo</i> <sub>3</sub> ***	11	26	2	1	--
TOTAL	16	93	6	2	3

\* I need it \*\* I understand it \*\*\* I have/own it

Table 4: Referential distance for accusative personal Pronoun (frequencies)

	CL <sub>0</sub>	CL <sub>1</sub>	CL <sub>2</sub>	CL <sub>3</sub>	CL <sub>≥4</sub>
Pronoun <i>lo</i> <sub>1</sub> *	7.5%	75%	10%	--	7.5%
Pronoun <i>lo</i> <sub>2</sub> **	5%	92%	--	2.5%	--
Pronoun <i>lo</i> <sub>3</sub> ***	27%	65%	5%	2.5%	--
Abs. Freq.	13%	77%	5%	1.7%	2.5%

\* I need it \*\* I understand it \*\*\* I have/own it

## Demonstrative determiners and pronouns

Regarding demonstratives we carried out three separate corpus searches. On the one hand, we retrieved a sample of 50 adnominal demonstratives divided into two groups of 25 cases each: *este hecho* (this fact) and *ese hombre* (that man). The results of this sample are shown in tables 5 and 6. An additional corpus sample consisted of 157 cases of demonstrative pronouns (63 cases of demonstrative pronoun *esto* (this), 69 of *eso* (that) and 25 of *aquello* (that further). The results of this sample are shown in tables 7 and 8. The disparity of the cases analyzed per demonstrative pronoun, in particular the low number (25) of cases for pronoun *aquello*, is due to the actual frequency of use of demonstratives in modern Spanish. Overall figures found in the corpus show that pronominal demonstrative *aquello* has a low frequency of use (6%) as compared to *esto* and *eso*. Even between these two pronouns the differences are quite relevant (*eso* 60%) and (*esto* 34%). This is a clear indication that pronominal demonstrative *eso* (that) is preferred for speakers when it comes to referring to higher order entities in discourse like events, propositions or discourse topics. Nevertheless, overall figures change when the frequency of use of demonstrative adjectives is calculated. *Aquel* (NP) still shows a very low frequency of use (9%) when compared to *este* (NP) and *ese* (NP). But the total figures for these two demonstratives point in exactly the opposite direction as 30% of all corpus occurrences belong to demonstrative *ese* and a percentage as high as 61% to proximal *este*.

Our sample of demonstrative pronouns was restricted to events as type of referents of demonstrative anaphors. In

order to restrict the referential potential of demonstratives, we searched the corpus for expressions consisting of a combination of a demonstrative pronoun plus a typical predicate of events like *suced*er (happen), *ocurrir* (occur) or *pasar* (happen), e.g. *eso sucedió ...* (that happened...), which forces an event reading. The principal advantage of this strategy was to restrict the high number of demonstratives found in the corpus as well as facilitating the annotation part of the study by narrowing down the type of referent to be scrutinized.

Table 5: Referential distance for demonstrative determiners (cases)

	CL <sub>0</sub>	CL <sub>1</sub>	CL <sub>2</sub>	CL <sub>3</sub>	CL <sub>≥4</sub>
Dem. + NP <sub>1</sub> *	1	21	3	--	--
Dem. + NP <sub>2</sub> **	2	19	2	--	2
TOTAL	3	40	5	--	2

\* *Este hecho* (this fact) \*\* *Ese hombre* (that man)

Table 6: Referential distance for demonstrative determiners (frequencies)

	CL <sub>0</sub>	CL <sub>1</sub>	CL <sub>2</sub>	CL <sub>3</sub>	CL <sub>≥4</sub>
Dem. + NP <sub>1</sub> *	4%	84%	12%	--	--
Dem. + NP <sub>2</sub> **	8%	76%	8%	--	8%
Abs. Freq.	6%	80%	10%	--	4%

\* *Este hecho* (this fact) \*\* *Este hombre* (this man)

Table 7: Referential distance for demonstrative pronouns (cases)

	CL <sub>0</sub>	CL <sub>1</sub>	CL <sub>2</sub>	CL <sub>3</sub>	CL <sub>≥4</sub>
<i>Esto</i>	--	50	11	1	1
<i>Eso</i>	--	54	5	7	3
<i>Aquello</i>	--	19	4	-	2
TOTAL	--	123	20	8	6

Table 8: Referential distance for demonstrative pronouns (frequencies)

	CL <sub>0</sub>	CL <sub>1</sub>	CL <sub>2</sub>	CL <sub>3</sub>	CL <sub>≥4</sub>
<i>Esto</i>	--	79%	17%	1.5%	1.5%
<i>Eso</i>	--	78%	7%	10%	4%
<i>Aquello</i>	--	76%	16%	-	8%
Abs. Freq.	--	79%	12%	5%	4%

## Discussion

From a careful observation of the data in tables 3-8, we can draw important conclusions regarding the referential properties of the expressions involved in this study. As regards textual distance of the antecedent, there are no significant differences between demonstrative expressions and the accusative personal pronoun *lo*. In fact, there are not even significant differences between demonstrative determiners and pronouns. In all cases, the antecedent shows

a strong preference to be found in the clause immediately preceding the anaphor in question (CL<sub>1</sub>). Absolute frequencies are very similar for the three types of referring expressions: 77% (pronoun *lo*), 80% (demonstrative determiners) and 79% (demonstrative pronouns). With only minor exceptions, a general tendency is observed that can be worded as follows “the higher the textual distance of the antecedent with respect to its anaphor the lower the frequency of occurrence” or, in other words, the frequency of occurrence is inversely proportional to referential distance. As we said, CL<sub>1</sub> is, by far, the preferred location of the antecedent for all expression types but personal pronoun *lo* shows a somewhat significant rate of occurrence of antecedents in CL<sub>0</sub> (the clause containing the pronoun itself). This is primarily due to a frequent Spanish construction that combines a demonstrative with a *lo* pronoun. This particular combination is shown in example (18).

- (18) Considero que [todo acto o acción en que participe un miembro del partido afecta al partido]<sub>k</sub>, y *eso<sub>k</sub> lo<sub>k</sub>* tengo muy claro.  
 ‘I think that [every act or action in which a member of the party takes part affects the party]<sub>k</sub>, and that<sub>k</sub> (it)<sub>k</sub> is very clear to me.’

The last clause of (18) features this particular combination of demonstrative and personal pronoun *eso lo* (that it). The English gloss for this clause is: *and that ACC.3 have-Ising. very clear*. In our corpus study, these cases of *lo* were accounted for as CL<sub>0</sub>. Note that the original (first mention of the) antecedent occurs in the previous clause. The demonstrative anaphor *eso* refers back to the original antecedent in the first clause and the pronoun *lo*, in turn, should refer to the demonstrative. All three elements are coreferential as manifested by the subscript. The relevance of this kind of construction lies in the ability for both referring expressions to co-appear within the same clause while referring to the same antecedent. In my opinion, this may be a clear indication that there must be ‘something else’ in addition to accessibility or cognitive status’ marking that distinguishes between these two referring expressions. It may be argued that the combination *eso lo* in (18) is to be explained on syntactic rather than on semantic grounds (i.e. that use of the demonstrative necessarily triggers the use of the clitic, etc.). I won’t deny here the effect that syntax may have on this particular configuration, but the fact is that both expressions can appear next to one another while being coreferential. That co-occurrence is not syntactically necessary is manifested by the ability of pronoun *lo* to appear without the demonstrative in the same type of construction. This is shown in (19).

- (19) [Es algo incómodo revisar tu trabajo]<sub>k</sub>, pero (*eso<sub>k</sub>*) ya *lo<sub>k</sub>* tengo asumido.  
 ‘[It is somewhat uncomfortable to revise your own work]<sub>k</sub>, but (that<sub>k</sub>) I have already accepted it<sub>k</sub>.’

A few observations have to be made regarding the syntactic and semantic types of the antecedents involved. The results of the study involving antecedent type are shown in table 9.

Table 9: Syntactic type of antecedent

	NP	Other (clausal)
<i>Pronoun lo<sub>1</sub></i>	20	20
<i>Pronoun lo<sub>2</sub></i>	13	27
<i>Pronoun lo<sub>3</sub></i>	33	7
	55%	45%
<i>Dem.+ NP<sub>1</sub></i>	6	19
<i>Dem.+ NP<sub>2</sub></i>	24	1
	60%	40%
<i>Dem. Pron. Esto</i>	6	57
<i>Dem. Pron. Eso</i>	16	53
<i>Dem. Pron. Aquello</i>	5	20
	17%	83%
TOTAL	123	204
Abs. Frequency	37.6	62.3

Within the category *Clausal*, we have not only included strictly clausal antecedents (see, for example, the antecedent in (15) that spans over the entire subordinate clause that follows the SVO pattern, which is common to Spanish) but also other not so clearly delimited antecedents, though quite frequent in Spanish, like infinitival clauses, e.g. *tener una relación no está en mis planes* (having a relationship is not in my plans). Other antecedents in this group include non-NP antecedents that expand beyond the clause limits, i.e. complex objects consisting of two (or more) clauses. On the other hand, the category NP is clear enough to deserve any further explanation. Generally speaking, clausal antecedents are widely preferred (62.3%) over NP antecedents (37.6%). Nevertheless, taken in isolation, the personal pronoun *lo* shows a slight preference for non-clausal antecedents (55%) over clausal ones (45%). Thus, there are individual differences that clearly depend on the type of predicate that accompanies the personal pronoun or demonstrative determiner scrutinized. For example, *lo<sub>2</sub>* (understand it) shows a preference for clausal antecedents over NPs (27 vs 13 cases, respectively). Conversely, *lo<sub>3</sub>* (have/own it) shows the opposite preferences. Demonstrative pronouns (*esto*, *eso* and *aquello*) do actually show a strong preference for clausal antecedents (83%), whereas demonstrative determiners show opposite preferences depending on the NP involved in the expression (NP1: *hecho* (fact), NP2: *hombre* (man)).

In view of our data, it appears that the syntactic type of the antecedent does not contribute to discriminate the referential properties of the personal pronoun and the demonstratives. We’ve seen that these two types of referring expressions can have antecedents of assorted syntactic configurations. Furthermore, the semantic type (or denotation) of the referring expression does not seem to contribute either to accessibility marking or the cognitive status of the expression. Thus, we have found numerous cases of *lo* referring to a range of entities like individuals, propositions,

factual information or events. The same applies to demonstrative determiners.

Although not formally included in the present study, I would like to briefly mention two additional factors that arise from the corpus cases scrutinized. These factors are animacy and prominence of the antecedent. Animacy refers to the [+animate] or [+inanimate] semantic feature of the referent. In Spanish, animacy does not appear to be a factor affecting the cognitive status or accessibility marking of the referring expressions under study. We have seen that both demonstrative determiners and the pronoun *lo* can both equally refer to animate and inanimate entities. Most clausal antecedents denote genderless, non-concrete, inanimate entities like events, propositions, factual information, etc. but many NP antecedents denote animate entities. Let us recall that the accusative personal pronoun *lo* can be indistinctively used to refer pronominally to third person masculine singular (him) or to 'propositional' or higher order entities. Likewise, demonstrative determiners can equally refer to [+animate] (*ese hombre* (that man)) or [+inanimate] entities (*este hecho* (this fact)) -our dem.+ NP<sub>1</sub> and NP<sub>2</sub> constructions in tables 5 and 6-. Regarding animacy, neuter demonstrative pronouns cannot be taken into account as we forced an event (hence inanimate) referential reading for the purposes of this study.

Finally, a couple of observations have to be made regarding the position of the antecedent within the clause that contains it. Through a close scrutiny of all our corpus cases, we didn't find any particular differences between referring expressions or antecedent types as regards the position of the latter within the clause. Thus, our NP antecedents could be found in subject position as well as in less topical positions (e.g. direct object) and irrespective of the referring expression involved. On the other hand, clausal antecedents, which are commonly assumed to have a lower accessibility than, say, NPs were also found occupying typical prominent positions within the sentence (e.g. the infinitival clause in subject position of example (4)) as well as informationally less prominent positions (e.g. subordinate clauses headed by the complementizer *que* and acting as the internal argument of the verb).

## Conclusions

In this paper, I have presented data in favor of the hypothesis that one important role of demonstrative anaphors involves marking (sub)-topic shifts in the discourse and, more importantly, inform the hearer about this new situation by focusing the referent that is intended to be the new (sub)-topic. I defend that this focusing property closely resembles that of deictic demonstratives when used in physical context situations or nuclear pitch accent in phonological focus marking. The effect would be identical, namely, making the intended referent salient enough so as to override potential referential ambiguity due to intervening textual material or other specific factors. In this respect, demonstratives may be understood as lexical focus markers. The referents of demonstratives are commonly assumed to occupy a mid position in terms of degree of accessibility. On the contrary,

the personal pronoun *lo* should mark its referents as highly accessible. Our data from Spanish seem to contradict this view though.

By comparing demonstrative determiners and pronouns with the Spanish personal pronoun *lo*, we have showed that there are no relevant differences between these referring expressions as regards the various textual factors investigated (referential distance, type of antecedent, semantic kind of the referent involved, animacy and prominence). The different nature of these two referring expressions must then lie somewhere else. Recall that these expressions can even co-appear in discourse and anaphorically refer to the same discourse entity in combinations in which the demonstrative immediately precedes the personal pronoun. The referential puzzle presented in this paper is still far from being solved but, in my opinion, the path related to the focal capabilities of demonstrative anaphors is worth investigating.

## References

- Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, 24(1), 65-87.
- Ariel, M. (1990). *Accessing Noun Phrase Antecedents*. London/New York: Routledge.
- Ariel, M. (2001). Accessibility theory: an overview. In T. Sanders, J. Schilperoord & W. Spooren, *Text Representation: Linguistic and Psycholinguistic Aspects*. John Benjamins.
- Asher, N. (1993). *Reference to Abstract Objects in Discourse*. Kluwer.
- Diessel, H. (2006). Demonstratives, joint attention and the emergence of grammar. *Cognitive Linguistics* 17(4), 463-489.
- Grice, H.P. (1989). Presupposition and conversational implicature. In H.P. Grice (Ed.), *Studies in the Ways of Words*. Cambridge, MA: Harvard University Press.
- Gundel, J. K., Hedberg, N. & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language* (69), 274-307.
- Gundel, J. K., Hedberg, N. & Zacharski, R. (2003). Pronouns without NP antecedents: how do we know when a pronoun is referential?, In A. Branco, T. McEnery & R. Mitkov (Eds.), *Anaphora Processing: Linguistic, Cognitive and Computational Modelling*. John Benjamins.
- Maes, A. & Noordman, L.G.M. (1995). Demonstrative nominal anaphors: a case of nonidentificational markedness. *Linguistics* (33), 255-282.
- Reig-Alamillo, A. (2008). *Cross-dialectal Variability in Propositional Anaphora: A Quantitative and Pragmatic Study of Null Objects in Mexican and Peninsular Spanish*. Doctoral dissertation, Department of Spanish and Portuguese, The Ohio State University, Columbus, USA.
- Wolter, L.K. (2006). *That's That: The Semantics and Pragmatics of Demonstrative Noun Phrases*. Doctoral dissertation. The University of California, Santa Cruz, USA.