ON THE COGNITIVE STATUS OF ANTECEDENTS IN SPANISH DISCOURSE ANAPHORA

Iker Zulaica-Hernández¹
The Ohio State University

Abstract
As anaphoric elements, demonstrative are widely used by speakers to refer to first, second and third order (intensional) entities. In this sense, demonstratives are said to co-refer with individuals, events, situations and propositions thus greatly contributing to discourse cohesion. For some authors, the distinctive behavior of referential expressions like demonstratives, definite, indefinite articles, etc. lies in their capability to ‘mark’ the cognitive status of their antecedents. Thus, for example, demonstratives (this/that) would differ from the neuter pronoun (it) in that the former mark their antecedents as cognitively activated vs the cognitive status in-focus of the latter. This paper focuses on Spanish demonstratives and the theoretical notion of cognitive status and argues in favor of a reconsideration of the status activated for these elements in Spanish based on empirical data from a corpus study that analyzes the factor ‘recency of mention of the antecedent’ as a determining factor.

Keywords: accessibility, anaphora, cognitive status, demonstratives, discourse anaphora, events, focus, information structure, pronouns, reference.

I. INTRODUCTION

Over the last decades, corpus-based research has turned out to be of great importance in helping provide adequate solutions to many theoretical issues in different linguistic fields. Discourse anaphora (reference to discourse entities by means of a range of linguistic devices) still constitutes one of the principal challenges for general anaphora studies partly due to the heterogeneous and sometimes tricky nature of the antecedent in discourse anaphora patterns, partly due to the abundance and eclectic behavior of discourse anaphors within a single and across languages. As a matter of fact, discourse anaphora can be currently considered the Achilles’s heel of anaphora studies. The issue has been tackled from different theoretical perspectives achieving outstanding advances in the comprehension of the mechanisms that govern the class of referential chains that arise in discourse (see, for example, the work by Asher (1993), Byron (2004), Fox (1987), Webber (1979), to name only a few). Nevertheless, there are many questions for which no completely satisfactory answers have been provided yet. Thus, to name only a few: What are the cognitive mechanisms involved in this phenomenon? How does information structure (references) affect the speaker’s choice for one anaphor over another? What are, if any, the differences among anaphors (e.g. the speaker’s preference of a demonstrative pronoun over a demonstrative determiner)?

¹ Address for correspondence: Iker Zulaica-Hernández. The Ohio State University. Department of Spanish and Portuguese 298, Hagerty Hall. Columbus, OH USA. Phone/Fax: 614-292-4958/614-292-7726. ikerzulaica@gmail.com
Natural language corpora have turned out to be particularly relevant tools when it came to scrutinizing pronoun resolution phenomena within the field of anaphora from both a theoretical and purely descriptive computational, semantic or pragmatic perspective. This is indeed the case of pronominal reference in discourse to higher order entities like propositions, facts, and events via demonstrative pronouns. This paper presents a corpus study on the discourse behavior of Spanish demonstrative pronouns as elements which primary anaphoric function is that of referring to a range of discourse entities that still constitute a puzzle to linguists due to their high degree of syntactic and semantic variability. I’m particularly interested in the notions of topic/focus (Givon 1983), given and new information (Prince 1981), and the cognitive status of the antecedents (Borthen et al (1997), Gundel et al (1993), Hegarty et al (2003)) in cases of demonstrative discourse anaphora and the extent to which these phenomena may have an influence on the different behavior shown by Spanish demonstrative pronouns in discourse anaphora patterns. In a series of papers, Borthen et al (1997) and Hegarty et al (2001) defend that the cognitive status of the antecedent of demonstrative pronouns in discourse is ACTIVATED as opposed to the status IN FOCUS that is typical of pronouns like English it or Spanish zero pronoun. The main goal of the present paper is to put to test the cognitive status of the antecedent of demonstrative pronouns in Spanish discourse. With this aim, I have measured the textual distance of the antecedents of demonstrative pronouns in natural discourse as a relevant factor to explain the cognitive status of the antecedent. In view of the data from the corpus study, I’ll argue that the cognitive status of the antecedent of demonstrative pronouns in Spanish can be better explained as being into the focus of attention (IN FOCUS) rather than merely ACTIVATED.

II. SPANISH DEMONSTRATIVE ANAPHORS

In Spanish, demonstrative anaphors like esto, eso and aquello (‘this/that/that yonder’) are commonly used to refer to entities like facts, events, or situations introduced via clauses or sequences of clauses (Asher (1993) for English, Eguren (1999) for Spanish). This point is illustrated in (1), where the medial demonstrative pronoun eso (that) refers back to a piece of textual material at the beginning of the paragraph: tener una relación con un hombre. Semantically, the presence of the predicate of events pasar (happen) appears to force an event reading of the antecedent. Thus, we may say that the antecedent to the pronoun in (1) constitutes an event. On the other hand, the demonstrative in (2) appears to refer back to the statement in the previous clause (underlined): a medida que crecemos nos vamos acostumbrando a la realidad that may be taken to be simply a proposition conveying some factual information.

(1) “Tener una relación con un hombre 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 PASA sólo una vez en la vida” manifiesta Maripily mientras posa muy sensualmente en medio de las Calles de Santo Domingo.
‘Having a relationship with a man is not in my plans for the moment as I must first wait until I my heart heals. But my friends tell me not to be stupid, that I should give myself a chance as THAT only HAPPENS once in a lifetime’ Maripily says as she sensually poses in Santo Domingo streets.’

(Vea on line nº1761, 17/05/2003. San Juan de Puerto Rico.)

(2) Creo que la experiencia de la infancia es muy superior a la de la adolescencia. En cada adulto hay un niño, pero no un adulto en cada niño. Cuando llegamos al mundo somos seres curiosos, a medida que crecemos nos vamos acostumbrando a la realidad, por ESO creo que la filosofía es un sistema magnifico para volver a maravillarse con la realidad.

‘I think the experience of childhood is far superior than adolescence. There is a child in every adult, but not an adult in every child. As we get to this world we’re curious beings, we get used to reality as we grow up, and THAT’s why I believe that philosophy is a great system to marvel at reality again.’

(La Vanguardia, 30/11/1995.)

Other Spanish discourse anaphors also play an important role in referring to high-order entities, namely, the clitic propositional lo and the null pronoun (Reig 2008). Thus, for example, the pronoun lo (‘it’) is commonly found to corefer with propositional entities. This point is illustrated in (3), where the two occurrences of the pronoun lo in the second clause refer back in the text to a propositional entity introduced in the previous clause.

(3) El espacio me obliga a olvidar a muchos de los que asistieron, LO sé y LO siento.

‘Reasons of space oblige me to forget many of those who attended, I know ?IT and I’m sorry about ?IT.’

(La Vanguardia, 09/11/1994.)

Spanish pronoun lo can only act as a direct object in Spanish (the internal argument of the verb) and, consequently, we’ll never find it filling the syntactic slot of a subject in a referential use. In this sense, it differs to English pronoun it regarding its syntactic function as the English pronoun it can be indistinctively used referentially as the internal argument of the verb or as the syntactic subject of the sentence. Compare the uses of Spanish propositional lo (4) with the grammatical role of direct object and English it as direct object in (5a) and subject in (5b).

(4) A: ¿Partido fácil a priori?
B: Bueno, no LO creo.
A: ‘Would you consider it an easy match beforehand?’
B: ‘Well, I don’t think so.’
(El Norte de Castilla, 12/01/2001)

(5)  
a. I know IT when I see IT.
b. IT happened once. Could IT happen again?

This is clearly not the case of Spanish, where the pronoun *lo* cannot be found in subject position in a referential use. Certainly, Spanish pronoun *lo* can be found in subject position in certain Spanish constructions (see (6)) but that *lo* is clearly not the propositional *lo* we are dealing with in this paper. In these cases, the pronoun acts as a syntactic placeholder, or a dummy subject in cleft constructions. The clitic *lo* in these range of constructions does not appear to be referential; at least not in the sense that propositional *lo* is.

(6)   
Para nosotros, LO más importante es la calidad de nuestro producto.
‘To us, the quality of our product is the most important thing.’


In (7), the verb *suceder* is a typical predicate of events, that is, it forces a reading in which its argument needs to be some sort of eventuality (action, accomplishment, achievement, etc.), or an event type. Even when embedded within the modal operator *si*, the argument of the verb *suceder* is to be considered a hypothetical event. The syntactic antecedent for the pronoun in (7) can be taken to be the whole discourse span prior to the clause containing the demonstrative anaphor *esto* or the fragment *inclinar la balanza en dirección del presidente Bush*. Compare this typical referential use of the demonstrative with the ungrammatical example of propositional *lo* as subject in (8).

(7)   
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.’

(La Prensa de Nicaragua, 02/11/1994.)

(8)   
*LO sucede todos los días.
‘IT 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 (9).
(9) A: No seas rencoroso. Ya sabes que el hombre cambia.
    B: Sí, ESO sí LO sè.
    A: ‘Don’t be so resentful. You know the man changes.’
    B: ‘Yes, I know THAT.’

(El País, 01/08/1976)

Finally, the null pronoun can also be used referentially in Spanish to refer to propositional entities. This is the case of the null pronoun in (10) (null = ∅), where it appears to be referring back to the proposition antier se ahogó un niño allá (the day before yesterday, a kid drowned over there). This phenomenon though seems to be pretty much restricted to certain varieties of Spanish (Reig 2008)1.

(10) Bueno sí, sí ha habido [accidentes], nada menos antier se ahogó un niño allá, allá, allá ¿sí ∅ sabías?
    ‘Well, yes, yes there have been some [accidents], the day before yesterday a kid drowned there. Did you know ∅?’

(Corpus de habla de Monterrey.)

III. ACCESSIBILITY OR THE COGNITIVE STATUS OF THE ANTECEDENT

Prince (1981b) was the first to propose a hierarchy for discourse entities called the Scale of Familiarity, which is based on three main factors: predictability, saliency and the common knowledge shared by speaker and hearer. Prince conceives the text as a discourse model and discourse entities as objects in that model. She postulates a taxonomy that defines entities on the basis of the type of information they convey. Thus, we find three basic types of entities that convey new, inferable or evoked information. The scale of familiarity is strictly ordered and is represented graphically as in (11).

(11) Prince’s Scale of Familiarity
    \{E/E^{S}\} > U > I > I^{C} > BN^{A} > BN

Use of the scale of familiarity is made in relation with the hypothesis that the speaker has about the hearer’s set of beliefs. In other words, if the speaker can use a certain linguistic form on the basis of her hypothesis about the hearer’s knowledge, but instead she chooses a lower position in the scale for reference to that entity the speaker wouldn’t be cooperative or she simply would be consciously deviating from the conversational norm. For the purposes of this paper, it is interesting the treatment given by this author to propositional anaphora. She treats propositional anaphora –anaphoric reference to propositional objects in discourse- as a case of inferable (I/I^{F}). An inferable discourse entity can be defined as an entity, which require of a supplementary inferential effort by part of the hearer/reader to be fully processed. Inferables
are normally inferred from other linguistic constituent, which has been explicitly mentioned or simply evoked in previous discourse. This is illustrated in (12).

(12) [Talking about how the kids across the street threw paint in their yard]
    Those kids are just – And she’s pregnant with another one.

The referent of the pronoun she is not explicitly mentioned in this utterance. Consequently, the hearer has to infer, via the NP the kids, that the referent of the pronoun is the mother of the kids. In my view, most cases of reference to propositions and events should not be considered inferables. On the one hand, the special relations that can be observed between the trigger and the inferable pronoun (specific entity to generic kind, generic kind to set of specific entities, individual to couple, individual to group, etc.) do not arise in most cases of abstract anaphora.

(13) Se inició una obra sin licencia. Por ESO, es necesario que se suspendan los trabajos.
    ‘The works began without a permit. For THAT reason the works must be suspended.’
    (El Tiempo, 11/11/1996. Colombia, prensa.)

In (13), the demonstrative eso unequivocally refers to the propositional content introduced in the first clause. It may be argued that the pronoun’s reference is ambiguous between a reading in which it refers to the whole proposition or to the proposition denoted by the clause headed by que but, either way, the propositional referent can be characterized as a discrete propositional discourse referent where no particular inferable relation occurs between antecedent and pronoun. Therefore, a subordinate clause or an entire clause can be semantically represented as an abstract discourse entity that can be old or new regarding their informational contribution to the discourse but it constitutes a discrete discourse entity similarly to the discrete semantic entity that the common noun denoting NP el coche (the car) uncontroversially introduces in discourse. Thus, the anaphoric relation between an anaphor and its abstract antecedent must be considered a direct relation contrary to what Prince proposes. As we will see in the following sections, the different denotations of common nouns, propositions or events do not have real implications regarding their potential availability as antecedents to anaphors or can be used to explain the different referential behavior of discourse anaphors.

Gundel, Hedberg and Zacharski’s (1993) Givenness Hierarchy is an implicational hierarchy of cognitive states and linguistic forms that aims at resolving the differential anaphoric behavior of pronominal and non-pronominal anaphors. The hierarchy for English and Spanish as proposed in Gundel et al. is given in Figure 1.
The six cognitive states (IN FOCUS, ACTIVATED, FAMILIAR, UNIQUELY IDENTIFIABLE, REFERENTIAL and TYPE IDENTIFIABLE) and their corresponding pronominal or determiner forms are intended to signal 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, there exists a correlation 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. It is an implicational scale, what means that whenever the speaker uses a specific linguistic form he/she would be implicating 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. The Givenness Hierarchy is intended to be universal. As we can see in Figure 1, the referent of a demonstrative pronoun in Spanish must be at least ACTIVATED. Being ACTIVATED means that there should be a representation of the referent in short-term memory. The referent of a zero pronoun or the English weak pronoun *it* must be cognitively IN FOCUS, meaning that the referent not only is in short-term memory but it is also at the current center of attention. 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. Thus, IN FOCUS entities commonly include the topic of the previous clause as well as other still relevant topics. The following examples from Wolter (2006) illustrate the difference between IN FOCUS and ACTIVATED entities.

(14) [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!
(15) [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) analyze the behavior of referential pronouns without explicit antecedents. In that paper, pronouns without NP antecedents were classified as pleonastic (lacking a referent), possibly pleonastic, or referential. Referential pronouns
were further classified by virtue of the type of referent (fact, proposition, activity, reason) that was introduced by a non-NP or as an inferable. They observed that, something already found in Hegarty et al (2003), only around a 15% of pronominal reference to clausally-introduced material in English was made using the pronoun *it* as opposed to demonstrative pronouns *this/that*. These were the results of an empirical study based on the Santa Barbara corpus of spoken American English. 330 third-person personal pronouns (16.1%) out of 2,046 were found to lack NP antecedents. In 110 out of these 330 cases, the pronoun *it* was found to refer to some type of abstract object (fact, proposition, event, etc.)

| Table 1: Referential Pronouns without NP Antecedents (Gundel et al, 2003) |
|-----------------|-----------------|
|                 | Number | Percentage |
| Fact            | 6      | 5.45       |
| Proposition     | 10     | 9.09       |
| Activity        | 27     | 24.55      |
| Event           | 3      | 2.73       |
| Situation       | 63     | 57.27      |
| Reason          | 1      | .91        |
| Total           | 110    | 100.00     |

As Table 1 indicates, the number of eventualities (activities, events and situations) found in their corpus study is comparatively much higher than that of propositional entities such as facts or propositions. They come to the conclusion, backed by Hegarty et al’s suggestions, that the Givenness Hierarchy can explain the fact that most reference to abstract objects is made via demonstrative pronouns if it is assumed that material introduced in clauses or sequences of clauses is ACTIVATED as compared to material introduced in syntactically prominent noun phrases which is more likely to be IN FOCUS. The figures in table 1, which are only based on the third personal pronoun *it*, can be explained on the basis of the higher or lower degree of immanence of the entities referred to. Thus, the fact that pronoun *it* (which correlates with IN FOCUS entities according to the hierarchy) is the preferred element of the hierarchy for reference to eventualities would indicate that entities with a high degree of world immanence, i.e., showing causal efficacy and spatiotemporal location, would be more likely to be brought into focus than propositions or facts.

In a series of papers, Hegarty (2003, 2006) and Hegarty et al (2001, 2003) also studied abstract object anaphora from a purely semantic perspective. Briefly, he analyzed the anaphoric possibilities to entities such as events, situations, propositions and facts by taking into account different parameters like the cognitive status, the ontological nature of the antecedent and their position along the spectrum of immanence. One more dimension is crucially included in Hegarty’s work, i.e., the syntactic category of the abstract-entity denoting expression: clausal (IP, CP) or nominal (NP, DP) for as he points out “it will be observed that clausal and nominal expressions referring to what is, ostensibly, the same kind of entity can have divergent semantic behavior, implicating different semantic types associated with the different syntactic categories, and correspondingly, different semantic
ontologies, specifically, denotation domains with different mathematical structures.” (Hegarty, 2000)

For Hegarty, there is a clearly different referential behavior between events and nominals denoting propositions, facts, reasons or situations on the one hand and clausally introduced propositions, facts, reasons and situations, on the other. This differential semantic behavior would be due to the former group being first-order entities of type $e$ as opposed to the latter being of entities of the higher semantic type $\langle\langle s, t \rangle, t \rangle$. Thus, events and proposition-denoting nominals would referentially behave as concrete entities and clauses denoting propositions, facts, situations and reasons share a common semantic behavior with respect to pronominal anaphoric reference, coordination and quantification by quantity adverbs over their denotation domains. One of the most interesting outcomes is that events are clearly separated from the rest of abstract entities. Furthermore, Hegarty’s investigation is complemented with the accessibility status of the entities under study. The following two examples from Hegarty (2003) illustrate this point.

(16)  
\begin{itemize}
  \item a. John broke a priceless vase. That/this is intolerable to the embassy.
  \item b. John broke a priceless vase. ??It is intolerable to the embassy.
\end{itemize}

In (16), reference is made to a situation introduced in the first sentence. This should be compared with (17) where reference is made to an event. The two different predicates force this ontological distinction: the predicate *intolerable* apply most naturally to situations, whereas the predicate *happen* is a typical predicate of events.

(17)  
\begin{itemize}
  \item a. John broke a priceless vase. That happened at noon.
  \item b. John broke a priceless vase. It happened at noon.
\end{itemize}

Clausally-introduced situations pattern with propositions, facts and reasons in that they are more commonly anaphorically referred to with a demonstrative pronoun as in (16), hence indicating that the cognitive status of these entities is ACTIVATED, and further accessibility to pronominal reference is as postulated in the *Givenness Hierarchy* (Figure 1). On the other hand, clausally introduced events are immediately rendered IN FOCUS and pronominal reference is possible via the third person pronoun *it* and with the demonstrative pronoun *that*. The fact that clausally introduced propositions and facts can in some cases be referred to with a personal pronoun in discourse can be explained in terms of a change in the semantic type of the entity and a positive epistemic gradient. According to this gradient, some verbs would act as facilitators that confer a higher epistemic status to the propositional content of the referent.

(18)  
\begin{itemize}
  \item A: Susan’s boyfriend has graduated.
  \item B1: I doubt that/#it; he doesn’t have very many credits.
  \item B2: I don’t doubt it; he has a lot of credits.
\end{itemize}
In (18), the fact newly introduced by clause A can be referred to with a demonstrative pronoun in B₁—thereby conferring it the common cognitive status ACTIVATED— or with a personal pronoun in B₂. In this second case, the expression don’t doubt acts as a facilitator, hence facilitating the accommodation of the entity to a discourse-old status with an IN-FOCUS cognitive status. For Hegarty, a proposition whose epistemic status increases is promoted to the cognitive status IN FOCUS over ones whose epistemic status diminishes or remains constant. This could be explained by presuming that the higher the likelihood of being true of a proposition the higher its degree of salience.

There is an important point to be made regarding the theoretical appropriateness of the cognitive statuses as reflected in the givenness hierarchy. Hegarty states this point very clearly “An entity will be IN FOCUS only if it has been mentioned by a nominal expression in a prominent syntactic argument position earlier in the utterance or in the previous utterance. This supposition is compatible with Centering Theory (Grosz et al., 1995) and results in the experimental psycholinguistic literature (Bock & Warren, 1985; Gordon et al. 1993; Almor, 1999). More peripherally introduced entities, including those introduced by less prominent nominal expressions and by clauses, will be ACTIVATED upon their introduction, placed in working memory within the field of attention, but not at the center of attention.” (Hegarty, 2000)

Poesio and Modjeska (2005) try to make more precise the cognitive notions of IN FOCUS, ACTIVATED, and SHORT TERM MEMORY as they appear in Gundel et al. They primarily adopt the computational approach to anaphora resolution of Centering Theory (Grosz, Joshi & Weinstein (1995))² and follow previous findings on that field to better define these notions. It is worth noting that Poesio and Modjeska focus on constructions of the type THIS-NPs and, as we already know from the Givenness Hierarchy, the common activation/cognitive status of the referent for this type of expressions is not IN FOCUS but rather ACTIVATED. Leaving quite a few details of their work apart for reasons of space, Poesio and Modjeska worked on an annotated corpus (GNOME) and tested the following hypothesis regarding the speaker’s non-preference to use THIS-NPs to refer to IN-FOCUS entities:

- THIS-NPs are preferentially used to refer to entities other than the CB(Uᵢ), the CB containing the utterance of the THIS-NP.
- They are used to refer to entities other than the CB(Uᵢ₋₁), the CB of the previous utterance.
- They are used to refer to entities other than CP(Uᵢ₋₁), the most highly ranked entity of the previous utterance.

And the following hypothesis regarding the speaker’s preference to use THIS-NPs for reference to ACTIVATED (‘ACTIVE’ in their own terminology) discourse entities. Thus, an entity is ACTIVE if:

- It is in the visual situation; or
- It is a CF of the previous utterance; or
- It is part of the implicit linguistic focus. They only considered as part of the implicit focus those entities that can be CONSTRUCTED out of the previous utterance. And an entity can be constructed out of an utterance if:
(a) It is a plural object whose elements or subsets have been explicitly mentioned in
that utterance; or
(b) It is an abstract entity introduced by that utterance. They consider two types of
abstract entities:
   i. Propositions
   ii. Types

They got the following results in terms of distribution:

<table>
<thead>
<tr>
<th>Class</th>
<th>Number (Percentage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anaphora</td>
<td>45 (40%)</td>
</tr>
<tr>
<td>Visual Deixis</td>
<td>28 (25%)</td>
</tr>
<tr>
<td>Discourse Deixis</td>
<td>19 (17%)</td>
</tr>
<tr>
<td>Type</td>
<td>9 (8%)</td>
</tr>
<tr>
<td>Plurals</td>
<td>1</td>
</tr>
<tr>
<td>Ellipsis</td>
<td>1</td>
</tr>
<tr>
<td>Time</td>
<td>1</td>
</tr>
<tr>
<td>Unsure</td>
<td>5</td>
</tr>
<tr>
<td>Disagreement</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>112</strong></td>
</tr>
</tbody>
</table>

With respect to the correlation between focus and THIS-NPs, they found the following
principal results:

- 8-11 violations to the hypothesis that a THIS-NP is used to refer to entities other than
  the CB(U_{i-1}) were found, which is therefore verified by 90%-93% of THIS-NPs.
- The hypothesis that THIS-NPs are used to refer to entities other than CP(U_{i-1}) is
  verified by 75-80% of THIS-NPs.
- The hypothesis that THIS-NPs are used to refer to entities other than CB(U) is verified
  by 61-65% of THIS-NPs.

Based on these results and an in-depth study of the violation cases they proposed the
version that leads to the fewest number of violations to Grice’s maxim of Quantity:

- **The THIS-NP Hypothesis**: THIS-NPs are used to refer to entities, which are
  ACTIVE in the sense specified above. However, pronouns should be preferred to
  THIS-NPs for entities other than CB(U_{i-1}).

In this section, I have presented a brief review on the notion of cognitive status of
antecedents in cases of discourse anaphora and a summary of the literature on the subject. As
proposed in Gundel et al’s Givenness Hierarchy, the cognitive status of the referent
(antecedent) to demonstrative anaphors must be ACTIVATED. As we have seen, data from
diverse studies seem to confirm this hypothesis for English. There is, nevertheless, an
important objection to be made concerning the statuses IN FOCUS and ACTIVATED presented in this section, which is central to the main thesis presented in this paper. My objection relates to the insufficient explanation as to how IN FOCUS vs ACTIVATED entities can be distinguished in discourse. But this will be made clear in the next sections.

IV. CORPUS STUDY: METHODOLOGY AND RESULTS

The corpus used for this study was the CREA corpus of Spanish language (Corpus de Referencia del Español Actual). For the purposes of this paper, I carried out corpus queries consisting of combinations of demonstrative pronoun plus verbs, which are commonly considered to be typical predicats of events (Vendler's event containers, 1967). The purpose of this was twofold. On the one hand, a purely procedural purpose was that of constraining the high number of occurrences of demonstrative pronouns in the corpus (very high in the case of demonstratives ESTO and ESO). The number of occurrences thus obtained was tractable for a manual count and exploration of every case. The sample obtained is, we believe, representative enough to be able to draw any potential conclusions that may be derived from this study. On the other hand and more importantly, the main goal of this study is to explore the behavior of events as semantic referents of demonstrative pronouns in cases of discourse anaphora. In consequence we needed a method to constrain the range of potential referents that may serve as semantic antecedents for demonstratives. The only possible way to constrain potential semantic referents to events was the use of event containers along with demonstratives in varied combinations of the type ESTO/ESO/AQUELLO + PREDICATE OF EVENT (THIS/THAT/YONDER + PREDICATE OF EVENT). Nine different event containers were used in our study, all of them combined with all three demonstratives. The list of predicates are shown in (19):

(19)  
\begin{align*}
Suceder & \quad \text{to happen} \\
Ocurrir & \quad \text{to occur/happen} \\
Terminar & \quad \text{to finish/end/come to an end/terminate} \\
Terminarse & \quad \text{same as above (reflexive meaning)} \\
Durar & \quad \text{to last} \\
Tener lugar & \quad \text{to take place} \\
Pasar & \quad \text{to occur/happen} \\
Acabar & \quad \text{to end/come to an end} \\
Acabarse & \quad \text{same as above (reflexive meaning)}
\end{align*}

I have restricted this study to only one tense (preterite) for it might well serve as a valid sample to draw conclusions from. Once I obtained my corpus sample, the next step was to measure the referential distance of antecedent and anaphor in order to test whether this factor may have an influence in the behavior of Spanish demonstrative anaphors as vehicles of anaphoric reference. For this purpose, I segmented all cases under scrutiny into clauses and quantified the textual distance of the propositional antecedent relative to the demonstrative anaphor. The segmentation procedure is illustrated in (20). Note only the relevant discourse was segmented (the second paragraph in the case of (20)), and all clauses prior to the one
containing the anaphor numbered with a subscript in a decreasing order. Thus, in (20), the referent of the demonstrative anaphor was found in clause 1 (\([\text{CLAUSE}_1]\)). This means that the referent (underlined) is found in the clause immediately preceding the anaphor *esto*.

(20) P: ¿Hay muchos mitos en esta historia?  
R: [Sí, los hay]\(_4\); [como que Watson y Crick trabajaban en una barraca o en un cobertizo para bicicletas]\(_3\). [¡No es verdad!]\(_2\) [Es cierto que este grupo, o cierta parte de este grupo, ocupó la barraca que se convirtió en laboratorio]\(_1\), pero *esto ocurrió* cuatro años más tarde, en el 57. Es como la historia de Newton y la manzana: hace falta la barraca, y hace falta la manzana.

P: ‘Are there too many myths in this story?’  
R: ‘[Yes, there are]\(_4\); [like the one that says that Watson and Crick used to work together in a hut or in a bicycle shed]\(_3\). [That’s not true]\(_2\), [It is true that this group, or a certain part of this group, occupied the hut that eventually became a lab]\(_1\), but *this happened* four years later, in 1957. It is like the story about Newton and the apple: we need the hut as much as we need the apple.


The results of the study are shown in Tables 3-6, but before we go on to analyze them a note on notational convention is needed. In total 193 cases were scrutinized: 82 cases of demonstrative *esto*, 80 cases of *eso* and 31 cases of *aquello*. We looked for seven different possibilities as shown in every table: (1) stands for the antecedent being found on the clause immediately prior to the occurrence of the demonstrative; (2) stands for the antecedent being found on the second clause or, in other words, on the second clause prior to the one containing the demonstrative; (3) for the antecedent being found on the third clause, etc. The category labeled (+) stands for cases where the antecedent was found in the fifth clause prior to the one containing the demonstrative and higher (6, 7, etc.) Category N/A stands for all those cases that, for some reason, could not be determined (i.e. the antecedent could not be found in the surrounding text, the antecedent could not be clearly delimited, etc.) In those cases where the antecedent expands into more than one single clause, we accounted for the last clause of the multi-clausal antecedent. Finally, the category labeled (-1) stands for cases of cataphora, that is, the antecedent in these cases was found in the clause immediately following the anaphor. In fact, only one case of cataphora was found and the syntactic antecedent was found in the clause immediately following the one containing the demonstrative anaphor.
Table 3. Relative textual distance antecedent-anaphor for pronoun ‘esto’ (this)

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(+)</th>
<th>N/A</th>
<th>(-1)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Esto sucedió</td>
<td>12</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>19</td>
</tr>
<tr>
<td>Esto ocurrió</td>
<td>24</td>
<td>4</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>30</td>
</tr>
<tr>
<td>Esto (se) terminó</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Esto duró</td>
<td>4</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>Esto tuvo lugar</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Esto pasó</td>
<td>5</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>8</td>
</tr>
<tr>
<td>Esto (se) acabó</td>
<td>5</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>11</td>
<td>19</td>
</tr>
</tbody>
</table>

Total (per segment) 50 11 1 - 1 18 1 82

Percentage 61 13.4 1.2 - 1.2 22 1.2 100.00

Table 4. Relative textual distance antecedent-anaphor for pronoun ‘eso’ (that)

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(+)</th>
<th>N/A</th>
<th>(-1)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eso sucedió</td>
<td>9</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>12</td>
</tr>
<tr>
<td>Eso ocurrió</td>
<td>23</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>27</td>
</tr>
<tr>
<td>Eso (se) terminó</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>Eso duró</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>Eso tuvo lugar</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Eso pasó</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>Eso (se) acabó</td>
<td>12</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>17</td>
</tr>
</tbody>
</table>

Total (per segment) 54 5 7 3 - 11 - 80

Percentage 67.5 6.2 8.7 3.7 - 13.7 - 100.00

Table 5. Relative textual distance antecedent-anaphor for pronoun ‘aquello’ (that yonder)

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(+)</th>
<th>N/A</th>
<th>(-1)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aquello sucedió</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>Aquello ocurrió</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>6</td>
</tr>
<tr>
<td>Aquello (se) terminó</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>Aquello duró</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>4</td>
</tr>
<tr>
<td>Aquello tuvo lugar</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Aquello pasó</td>
<td>3</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>Aquello (se) acabó</td>
<td>6</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>10</td>
</tr>
</tbody>
</table>

Total (per segment) 19 4 - 1 1 6 - 31

Percentage 61.2 13 - 3.2 3.2 19.3 - 100.00

Individual data per demonstrative are given in Tables 3-5 including N/A cases. Table 6 summarizes the results in Tables 3-5 excluding those cases that could not be determined for
whatever reason (the N/A cases shown in tables 3-5). The results are very straightforward.

There are no major differences among demonstratives and referential distance: 78.0 % for demonstrative *esto*, 78.2 for *eso*, and 76.0 for *aquello*. The initial conclusion we may draw from these data is that all three demonstratives show a strong anaphoric preference for antecedents in the immediately prior clause (a frequency ranging from 76% to 78.2%). If we compare the results for proximal *esto* and distal *aquello* we find no significant differences regarding the distance of the antecedent. Demonstrative *eso* only shows a not too significant difference as compared to the other two demonstratives and this different behavior shows up in categories 1 and 2.

Table 6. Summary of textual distance (excluding N/A cases)

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(+)</th>
<th>(−1)</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Esto (this)</td>
<td>78.0</td>
<td>17.2</td>
<td>1.55</td>
<td>-</td>
<td>1.55</td>
<td>1.5</td>
<td>100.00</td>
</tr>
<tr>
<td>Eso (that)</td>
<td>78.2</td>
<td>7.2</td>
<td>10</td>
<td>4.3</td>
<td>-</td>
<td>-</td>
<td>100.00</td>
</tr>
<tr>
<td>Aquello (that yonder)</td>
<td>76.0</td>
<td>16.0</td>
<td>-</td>
<td>4.0</td>
<td>4.0</td>
<td>-</td>
<td>100.00</td>
</tr>
</tbody>
</table>

If, according to the Givenness Hierarchy, the propositional referents of demonstrative anaphors are only cognitively ACTIVATED and not IN FOCUS then we may initially conclude from our data that an 80% of cases of referents to demonstrative anaphors in Spanish are, at least, cognitively ACTIVATED. Now, it is difficult for me to agree with the idea that entities introduced in the clause immediately prior to the anaphor are not rendered into the focus of attention. In my opinion, these propositional entities (call them events, propositions or facts) are actually the current focus of attention and thus should be treated as cognitively IN FOCUS instead of merely ACTIVATED. The entities under study are salient enough to the participants in the conversation (they have been mentioned very recently in discourse). Regarding their syntactic prominence, most of the times they are entire clauses hence being difficult to demote them to a secondary, non-prominent position within the clause. In addition, they seem to commonly constitute the current topic of the conversation or, at least, one of the current topics of the discourse. Taken all these factors into consideration, I’ll argue that, in Spanish, the referents to demonstrative anaphors should be better viewed as entities IN FOCUS on a par with the referents of null pronouns.

V. CONCLUSIONS

In this paper, I have investigated the discourse anaphoric properties of Spanish demonstrative pronouns with the aim of determining the cognitive status of their referents following the hypothesis of the *Givenness Hierarchy*. According to this theory, the cognitive status of these propositional referents is ACTIVATED as opposed to the status IN FOCUS of null pronouns. A second but equally important point was to test if data from Spanish would confirm an observation already made for English, namely, that antecedents to demonstrative anaphors are commonly found in the clause immediately preceding the one containing the anaphor. In order to accomplish this, I quantified the *referential distance* as a main textual factor that contributes to promoting a discourse entity to a certain cognitive status. The data
from the empirical study presented in this paper appear to confirm this observation: in Spanish (as in English), the antecedents of demonstrative pronouns show a strong preference to be found in the clause immediately preceding the anaphor and irrespective of the demonstrative element under scrutiny. Based on the data from this study regarding recency of mention of the referent, I have argued in favor of propositional referents in Spanish as being into the current focus of attention and not ACTIVATED as it is proposed in the Givenness Hierarchy. My hypothesis is that demonstrative pronouns are devices used by speakers to actually mark merely ACTIVATED entities into the current focus of attention hence promoting their propositional referents to the cognitive status IN FOCUS. The way demonstrative anaphors achieve this it is unclear to me. Perhaps, demonstrative anaphors have not entirely lost their deictic component and contribute with some sort of ‘pointing’ device thereby making entities salient in discourse similarly as purely deictic uses of demonstratives do in ‘real context’ situations.

Demonstrative expressions have been extensively studied in Linguistics over decades. Today, there is a total consensus among linguists in recognizing the deictic component of demonstratives as a basic distinguishing factor that differentiates these from other referential non-deictic expressions. Thus, unlike other referential expressions, demonstratives need to be interpreted relative to at least one of the basic three dimensions of the context of utterance (the speaker/addressee, the location or the utterance time: the I, here and now of every utterance situation) in order to be fully processed. This feature is common to all deictic expressions, be it demonstrative pronouns and determiners (this/that, this/that-NP), temporal (now, then, tomorrow), locative adverbs (here, there), etc. Another common feature of deictic demonstrative expressions is that they can be accompanied by a deictic gesture (a pointing finger, a nod of the head, etc.) by the speaker. In some languages, deictic demonstratives appear to have developed non-deictic, anaphoric uses like the ones presented for Spanish throughout this paper. At least apparently, in these anaphoric uses, demonstratives have evolved into mere pronouns without a deictic component. When used in discourse, these demonstrative anaphors are not accompanied by a pointing gesture and, unlike their deictic counterparts, do not need to resort to the basic dimensions of the utterance situation (speaker/addressee, location and time) to be fully licensed and interpreted by language users. What these anaphors appear to need to be fully licensed in discourse is merely another linguistic expression previously uttered in the preceding (anaphora) or in subsequent (cataphora) discourse and with which they corefer. If my hypothesis about Spanish demonstratives herein presented is right, even in anaphoric uses these expressions would still retain some deictic content that would help the speaker to mark discourse entities as cognitively IN FOCUS thus clearly imitating the use of deictic gestual demonstratives in the spatial, three-dimensional world. Let’s recall that the primary role of (gestual) deictic demonstratives is to bring the hearer’s attention upon an entity that is not (or may not be) in his/her current focus of attention thereby making it salient enough for the hearer to recognize it. All this would amount to saying that Spanish demonstratives have not totally evolved into purely anaphoric expressions and still retain some deictic component. The way we should
formally characterize the proposed deictic content of the so-called ‘anaphoric’ demonstratives is far beyond the limits of this paper. Hopefully, future research will help us in this task.

NOTES

1 This example was taken from Reig (2008).
2 In Centering Theory it is assumed that new discourse entities (Forward-Looking Centers) introduced by each utterance are ranked based on information status. The CB is Centering’s equivalent of the notion of topic or focus. The Backward-Looking Center of Uₙ connects with one of the Forward-Looking Centers of Uₙ₋₁. CB(Uₙ), the Backward-Looking Center of Utterance Uₙ is the highest ranked element of CF(Uₙ₋₁) that is realized in Uₙ.
3 For Poesio and Modjeska (2005), types are those cases of generic reference that have concrete objects as instances.

REFERENCES


