Abstract

Spanish (along with English and many other languages) has inferential sentences such as No es que no quiera; Es que no sabe querer, ‘It’s not that she doesn’t love; it’s that she doesn’t know how to love.’ We describe the grammar and pragmatics of these sentences and show how their pragmatic characteristics follow from their grammar and the principles of relevance theory. Inferentials consist of a finite clause embedded as the complement of an expletive copular matrix clause, which may be modified by a focusing particle and/or a negator. Inferentials function as metalinguistic devices which characterize the relevance of the proposition represented by their clause to the processing of an utterance. Negative inferentials characterize that proposition as likely to be considered in the processing but they deny its relevance; positive inferentials characterize the proposition as unlikely to be considered but they assert its relevance. The inferential proposition may be interpreted as an implicated premise or conclusion. If it is taken as an implicated premise then it may be further interpreted as an explanation, reason, or cause; if it is taken as an implicated conclusion then it may be further interpreted as a result, consequence, or conclusion. It may also be taken as a (re)interpretation or reformulation of the target utterance.

Keywords: Grammar and pragmatics, Relevance theory, Inferential sentence, Spanish

1. Introduction

Many, probably all, languages provide devices which allow their speakers to focus entire clauses or sentences. In some languages the device is a focusing particle, for example Swahili ndiyo and sio (Bearth 1997), Akan nà (Bearth 1999), and Wolof moo (Robert 1993, cited in Bearth 1999: 264). In many other languages the focusing is accomplished syntactically by subordinating a clause to an expletive matrix consisting of a dummy or null subject, a form of the copula (in languages that require it), and a subordinating conjunction, as in Es que Florencio está enamorado de una rana ‘It’s that Florencio is in love with a frog.’ This construction has come to be referred to as the “inferential,” a usage we will...
continue here. This paper focuses on the Spanish inferential, with some comparisons with English, and provides an account that invokes the universal pragmatic principles of Relevance Theory (RT) (Sperber and Wilson 1986), and is therefore generalizable to other languages that exhibit the form.

Although the inferential forms and functions are widely distributed among languages, they have been relatively little studied. English inferentials are mentioned but not analyzed in Delahunty (1981), where they are related to cleft sentences. This relation has generally been acknowledged by later analysts both of clefts (e.g., Hedberg 1990) and of inferentials (Declerck 1992: 217); but see Heggie (1988 cited in Hedberg 1990: 183 fn 38) and Collins (1991) who distinguish the two sentence types. Bearth (1999) explores the connections between the constituent and sentential focusing uses of Akan nà, which are analogous to clefts and inferentials, respectively.


2. Issues

A number of issues have emerged from these discussions. The first is the primarily observational one of identifying the contextualized interpretations of inferential sentences. The second is to provide an account for these interpretations which is both general enough not to be merely stipulated and restrictive enough to include only attested interpretations of the form. The third is to explain just why this particular form has just the uses it has.

3. The data

This study is based on a data base of 25 naturally occurring Standard Iberian Spanish examples, collected with as much of their contexts as seemed necessary for their full interpretation and analysis. These were collected from both written and spoken sources. Some of the spoken examples are from TV news bulletins and so may have been scripted. Others occurred in ordinary conversation. Added to this 25 are a further 30 translations of naturally occurring English inferentials, which were translated for this project. These were translated by Laura Gatzkiewicz and checked for grammaticality and appropriateness by Prof. Juan F. Sanz-Cervera. Gatzkiewicz lives and works in Valencia, Spain, as a translator and language teacher. Sanz-Cervera is a native speaker of Spanish and professor of organic
chemistry at the University of Valencia, Spain. Each of these translations is marked as an “English original” in the text, except for the example from Eco (1989) whose original is in Italian, but which we translated from the English version. As neither data set was collected with statistical respectability in mind, we cannot be sure of how well the data represents the distribution of the various subtypes of inferentials in either language. Our discussion and analysis reflect the data we have and native intuitions.

Carefully using appropriately reviewed translations as data allows us to efficiently determine where the two languages overlap in their uses of the form. And we are far from alone in using such methods. Bernini and Ramat (1996) is an important cross-linguistic study of negation that also uses translation data. Parallel corpus work also depends on translated data.

4. The interpretations

The writers cited above have differed in their views of the range of interpretations that the inferential form may be used to communicate. Delahunty’s various articles have provided the most “generous circumscription” (Bearth 1999: 267 fn 28) of this range. These articles claim that the forms may communicate explanations, accounts, reasons, causes, (re)interpretations, reformulations, conclusions, results, and consequences; he also claims that it is not always possible to identify a unique interpretation of the form in context. Declerck (1992: 220) claims that the range of interpretations is limited to explanations and interpretations (whose relationship he does not clarify), which appear to include explanations, causes, interpretations, and reasons, but crucially not conclusions, results, or consequences. Bearth (1999) claims that inferentials communicate only explanations. He believes that Declerck ‘uses “interpretation” as a technical synonym of “explanation”’ (1999: 267 fn 28). Bearth (1999: 264-5) cites Robert’s (1993) claim that Wolof sentential focus constructions communicate explanations.

Besides the interpretations just mentioned, many of those who have written about inferentials have remarked on their discourse repair uses. Horn (1989: 435) claims that the sentential focus cleft, as he characterizes inferentials, is a metalinguistic device whose negative form may be used to reject an utterance and whose positive form may be used to tell “the addressee what aspect of the utterance is being rejected.” Delahunty (1990, 1995) says that inferentials are metadiscourse devices for instructing the audience to reject certain inferences and draw others. Declerck (1992: 206) says that he has “come to the conclusion that it is typical of the inferential it is that-construction that it always expresses what the speaker thinks to be the correct explanation or interpretation of something that has happened or has been said or the fact that a speech act has taken place.” Bearth (1999: 272-3, fn 38) endorses Declerck’s view that inferentials provide correct explanations, and suggests that inferentials are “intrinsically corrective.” The problem, of course, is to explain just how inferentials accomplish these discourse tasks. The literature provides several solutions, which we discuss in due course.

In this paper we demonstrate that the range of interpretations manifested by Spanish inferentials is the same as that manifested by English inferentials. However, Spanish speakers seem to use the form in a broader range of discourse situation types than English speakers. We propose an account, based on the principles of relevance theory, that explains
how the inferential can occur in many unrelated and non-interacting languages with analogous forms and very similar functions and interpretations. We begin by reviewing the range of interpretations of inferential constructions.

Inferentials are interpreted (somewhat indeterminately) as explanations, accounts, causes, reasons, justifications, (re)formulations, (re)interpretations, results, conclusions, or consequences of some aspect of the local context. (See Delahunty 1995, 1997 for discussion of the ways in which English inferentials are relevant in their contexts.)

Spanish inferentials also manifest these interpretations, and are used most frequently to assert or deny an explanation, account, reason, cause, or justification for something in the context. In the following example, a TV reporter (de la Cruz) and a news anchor (Viza) discuss a review by the Spanish senate of the general council in charge of the judiciary. The inferential that begins Viza’s response suggests why the politicians have the power to solve the council’s problems:

(1) Miguel Ángel de la Cruz: Los presidentes han reconocido la necesidad de renovar urgente (sic) el consejo y le han dicho que será a partir del dos de abril cuando comiencen las conversaciones para comenzar también estas renovaciones del consejo.Olga Viza:  Y es que constituidas las cortes los políticos tienen ahora en su mano la posibilidad de dar solución al problema del consejo. (Las Noticias de Antena 3, 3/28/1996)

‘Miguel Ángel de la Cruz: The leaders have acknowledged the necessity of urgently renovating the council and they’ve notified it [the council] that the talks to begin its renovation will begin as of April the second. Olga Viza: And it’s that the senate is made up the politicians now have in their hands the possibility of solving the problem of the council.’

The ser que sentence above could have been translated as ‘because the senate is now made up.’ However, the causal interpretation is left implicit in the original. In the following, the no ser que sentence proposes a reason (whose relevance it denies) for why the writer should thank Mr. Pedro Schwartz:

(2) Debo agradecer a Don Pedro Schwartz que aludiera repetidamente, en su columna del sábado pasado, a mi libro “El planeta americano.” No es que me tratara bien, sino que, precisamente por su vehemente desacuerdo y sus amañadas interpretaciones, deja su animadversión al aire. (El País, 11/16/1996)

‘I should thank Mr. Pedro Schwartz, who, in his column last Saturday, repeatedly alluded to my book “The American Planet.” It’s not that he treated me well, but rather that precisely for his vehement disagreement and his manipulative interpretations, he allows his animosity to be made public.’

Note here that the negative inferential denies a proposition that has not been asserted but which might readily occur to readers as a contextually plausible inference as to why the
writer should thank Mr. Schwartz.
The *ser que* sentence at the end of the following narrative provides an answer to the question implicitly set up by the prior sentence: What is the explanation for the strange fact that Florencio’s beloved croaks?

(3) Florencio está enamorado; pero no de una princesa. Cada noche, sale de su habitación sin hacer ruido y avanza en silencio por los jardines de palacio. Va a reunirse con su amada. ¡Qué emoción! Ella le espera a la orilla del estanque, bajo la luz de la luna. —Hola—dice Florencio al verla—. Ya estoy aquí. ¡Cuánto te he echado de menos! —¡Croac!—le responde ella. 

_Y es que Florencio está enamorado de una rana._ (Lalana 1996 p.11)

‘Florencio is in love; but not with a princess.

Every night he leaves his room without making a sound and he walks in silence through the palace gardens. He’s going to meet his beloved. How exciting!

She waits for him on the bank of the pond, in the moonlight.

“Hello,” says Florencio when he sees her. “Here I am. How I’ve missed you!”

‘Croak,” she answers him.

And it’s that Florencio is in love with a frog.’

So far our examples have been quite consistent with Declerck’s, Bearth’s, and other writers’ assertion that inferentials communicate explanations and related notions. The examples that follow demonstrate that we must expand that range.

In the following, Pepe Oneto, a TV commentator, remarks on recent political scandals in the institution that oversees Spanish judges. The second sentence is a *ser que* construction which appears to function as a reformulation of, or a reflection on (see Quirk et al 1985: 1392 fn), the idea presented in the first sentence:

(4) Pepe Oneto: No sé si comprendo bien lo que pasa o _es que realmente no quiero comprender_. (Las Noticias de Antena 3, 3/20/1996)

‘Pepe Oneto: I don’t know if I understand what’s happening well, or it is that I really don’t want to understand.’

In the next example the narrator seems to be trying out various descriptions of the character’s state of mind and the inferential is used to communicate a reformulation or reinterpretation of an earlier description:

(5) Había llegado más alla de la razón, incluso su habilidad de burlarse de sí mismo estaba muerta, o _quizá era que parecía que ya no había nada de qué burlarse_. (English original)
'He had got past the stage of reason, even his power of mocking at himself was dead, or perhaps it was that there seemed no longer anything that could be mocked at.' (Somerville and Ross 1977: 209-10)

In the next example, the speaker describes a fascist comic hero, Roberto Alcázar. The *ser que* sentence describes the result or effect of reading the comic book on the speaker, namely, that you die laughing:

(6) Y ni te cuento cuando actúa a dúo con el Pedrín. Entonces es ya el acabóse y la recarba; que te partes con el Pedrín, que tiene unas salidas de pata de banco, que *es que te atornillas de la risa.* ¿Qué pareja! (Sopeña Monsalve 1994: 116)

‘And not to mention when he works as a team with Pedrín. That’s just the be-all and end-all, I mean, you split your sides with Pedrín, who’s got some great lines, I mean, *it's that you die laughing.* What a pair!’

Alternatively, the inferential in this example may be viewed as a reformulation of the speaker’s earlier description of his reaction, *que te partes con el Pedrín,* ‘I mean, you split your sides with Pedrín.’

However, the following example clearly and unambiguously illustrates the use of an inferential to communicate a result, and therefore goes beyond the range allowed by either Declerck or Bearth:

(7) Alguien—quizás Rubenstein—dijo, cuando le preguntaron si creía en dios: “No, yo creo . . . en algo mucho más grande.” Y otro—¿fue Chesterton?—dijo que cuando los hombres dejan de creer en dios, *no es que entonces no crean en nada:* creen en todo. (Translated from English)

‘Someone—Rubenstein, maybe—once said, when asked if he believed in God: “Oh, no, I believe . . . in something much bigger.” And someone else—was it Chesterton?—said that when men stop believing in God, *it isn’t that they then believe in nothing:* they believe in everything.’ (Eco 1989: 620)

We can paraphrase this inferential as “what results when men stop believing in God is not that they then believe in nothing.” Note particularly that the proposition represented by the inferential clause cannot be construed as an explanation of men’s not believing in God because *then* in its clause indicates that believing in nothing follows ceasing to believe in God. Examples such as this demonstrate conclusively that inferentials are not, at least in English and Spanish, grammaticized forms for the communication of explanations alone.

---

2 The first *que of que es que* is an emphatic particle. It may be translated by ‘I mean’ or ‘like’. In this regard it is perfectly compatible with inferentials used to indicate reformulations or previously asserted or denied propositions. Notice that *que* is used in this way also at the beginning of the sentence immediately before the inferential in this text.
The inferential in the following text also falls outside the range allowed by Declerck and Bearth. It directs the reader not to derive the very reasonable conclusion that Lichardus’ model is wrong from the author’s extensive enumeration of its shortcomings.

(8) El modelo de Lichardus es una variante de una explicación más amplia de los cambios culturales observados en el norte y el centro de Europa en el período neolítico tardío. . . . [Approximately 28 lines of text enumerating the shortcomings of Lichardus’ model.] . . . Parece que el nivel de continuidad que se encuentra entre las culturas “Linear Ware” y Rossen en, por ejemplo, tipos de asentamientos y casas alargadas, es mayoritariamente ignorado y, cuando la discusión se refiere a algún elemento relacionado con la estepa, parece limitarse a categorías muy generales de cambios culturales, los cuales son normalmente considerados como el producto de procesos locales y no de intrusos de fuera. No es que el modelo sea equivocado; simplemente no hay suficiente evidencia propuesta para evaluarlo. (English original)

‘Lichardus’ model is a variant of a broader explanation of the cultural change seen throughout both North and Central Europe in the Late Neolithic. [. . .] The level of continuity that one encounters between the Linear Ware culture and the Rossen culture in, for example, settlement types and long houses, appears to be largely ignored, and where the argument concerns a steppe element, it seems mainly confined to very broad categories of cultural change, regularly regarded as the product of local processes rather than foreign intruders. It is not that the model is wrong; there is just not enough evidence proposed to evaluate it.’ (Mallory 1989: 253)

Here again, the inferential is not to be taken as communicating an explanation (albeit one denied) for the list of shortcomings the author attributes to Lichardus’ model. Rather, it is better viewed as communicating a conclusion that rational readers might derive from the list, but that the author wishes to steer them away from. It is conceivable that an analyst determined to maintain that inferentials communicate only explanations might argue that the inferential might be paraphrased as, “The explanation for my listing the shortcomings of the model is not that the model is wrong.” Or “The interpretation to be made of these shortcomings is not that the model is wrong.” However, neither of these paraphrases seems as natural in the context as something like “Do not conclude from (my saying) all of this that (I think that) the model is wrong.”

Several of the examples we have just discussed have allowed for more than one interpretation. This is consistent with the position we are proposing here. The inferential structure by virtue of the expletive nature of its matrix does not explicitly articulate the relationship of its proposition to its context. Rather, by virtue of choosing the inferential instead of a more explicit form, the speaker leaves it up to the hearer to decide just how to integrate the inferential with its context. However, the hearer’s freedom in this regard is constrained considerably by the form and the search for its relevance, as we describe below.
5. Prior analyses

Declerck (1992) claims that the copula of the inferential matrix is semantically specificational; that is, it specifies a value for a variable. The variable is akin to the heading of a list and the items on the list are its values (Higgins 1976: 95, cited in Declerck 1992: 210). So, from Declerck’s point of view, an inferential sentence assigns a value to a variable. The value is represented by the inferential clause; the variable is a pragmatically identified discourse entity that denotes either an explanation or an interpretation which replaces the matrix subject it at some level of representation. He proposes that the semantic restriction to explanation/interpretation is to be accounted for by claiming that the inferential matrix is of the form “NP is that” and that the NP may be headed by a restricted class of nouns. “Nouns like reason, cause, explanation, interpretation are of this type: they express a notion whose contents can be specified by a that-clause” (p. 220). Via unspecified pragmatic principles the it of the inferential would be replaced by an NP headed by words of the explanation/interpretation type, representing the variable whose value is specified by the clause.

Both of these claims are problematic. As we have seen, the claim that inferential sentences communicate only explanations or interpretations is falsified by examples such as (7) and (8) above. Moreover, note that in his discussion of the Japanese no desu construction, Kuno (1973: 226 example 6a) adds “or evidence” to his gloss:

(9) Byooki desu. Taizyuu ga zyuppono hetta no desu

sick weight ten-pounds lessened

‘I am sick. The explanation (or evidence) for my being sick is that I have lost ten pounds.’

Clearly, the meaning of this example is better paraphrased by The evidence is . . . than by The explanation is . . .

From these examples, it should be clear that the range of interpretations of unimpeachable examples of inferentials is not confined to explanation, cause, interpretation, reason. Rather it must be acknowledged that it also includes at least result, conclusion, and, perhaps, evidence (at least in Japanese).

Declerck’s attempt to explain the range of inferential interpretations by claiming that only NPs of the type explanation/interpretation/reason/cause may occur as heads of the NP in the frame “NP is that” is also inadequate. Many other nouns may occur in this frame with the requisite specificational reading, for example, The result/conclusion/evidence/fact/truth/allegation/ position/prediction/forecast is that inferentials will be underresearched and misunderstood. These nouns go far beyond the range claimed by Declerck. But more importantly they also go far beyond the range of interpretations associated with inferentials. It follows that Declerck’s approach provides no principled means of constraining the set of potential readings. His claim that only nouns such as reason, cause, explanation and interpretation can occur in this NP position is a stipulation with no explanatory value. The impossibility of constraining the range of nouns that can appear in this formula indicates that we should look elsewhere for a principled

Declerck insists that the specificational interpretation of inferentials is a semantic, rather than a pragmatic matter. This claim too is problematic. First, the notion of a separate semantic category of specification is difficult to justify. Hedberg (1990) discusses some of the problems associated with the specification (although she does continue to make use of it). She provides the following list of copular clause types along with the types of their subjects and predicates (p. 56):

<table>
<thead>
<tr>
<th>Type</th>
<th>Subject</th>
<th>Predicate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicational</td>
<td>Referential</td>
<td>Predicational</td>
</tr>
<tr>
<td>Specificational</td>
<td>Superscriptional</td>
<td>Specificational</td>
</tr>
<tr>
<td>Identificational</td>
<td>Referential</td>
<td>Identificational</td>
</tr>
<tr>
<td>Identity</td>
<td>Referential</td>
<td>Referential</td>
</tr>
</tbody>
</table>

Clearly, the types of the subject and predicate of a copular clause are sufficient to determine the type of the entire clause. To require also that the copula be many ways ambiguous is redundant. Much work in the 1980s demonstrated the superiority of pragmatic over semantic accounts of ambiguity of this sort, most notably of and (Carston 1988), but also including numbers, quantifiers, modal expressions, and gradable expressions (Horn 1988). (See also R. Kempson 1986.) We note here too that while many languages, including Spanish, distinguish lexically between a copula and a verb to be, none (that we know of) lexically distinguishes amongst the various readings of the copula.

As we demonstrate below, the matrix clause of inferentials is expletive. Neither the subject, whether null or an expletive such as it, nor the copula, nor the subordinating conjunction, have any conventional meaning in this construction. It follows that the interpretation of the matrix is a pragmatic matter. Significantly, Declerck’s discussion itself undermines his claim for the semantic status for specification. He demonstrates that a “specificational” reading is characteristic of clefts and can be imposed on any sentence by contrastively stressing a constituent, a manifestly pragmatic operation. So, rather than assuming that there are two sources for specification, semantic in copular clauses, and pragmatic in focusing constructions and utterances with contrastive stress, the assumption that there is but one, pragmatic, source is more general, more convincing, and more explanatory.

Bearth (1997, 1999) argues that inferentials are a specialized device for repairing “inferential gaps.” Inferential gaps arise when the “inferability maxim” is violated. Inferability is a requirement for discourse coherence. According to Bearth (1999: 269),

A claim made by a dialogue participant is coherent to the extent to which its validity is inferable for the hearer independently of the statement on which the claim is based. Failure to satisfy the inferability requirement results in an inferential gap, which in turn calls for the activation of an appropriate repair mechanism.
Inferential processes may be analyzed as consisting of three phases: (1) the emergence of a verbal or non-verbal trigger (the source of the inference) as part of ongoing discourse activity, (2) the mediating phase (the overt verbal utterance or the verbal exchange carrying the inferential operation), and (3) the target (the inferential proposition which the speaker intends to be accepted by the addressee).

What this amounts to is the claim that a discourse participant may make an assertion which contradicts a proposition which they had earlier committed themselves to as long as they subsequently provide a premise from which the assertion follows. That the premise is after the fact is indicated by the inferential marking. By virtue of providing a premise from which the target proposition follows, the inferential is explanatory:

A manifest failure to realize r before s or as a constituent part of s (s (s’ <- r)) results in an inferential gap $\lambda x[(x -> s)]$. The inferential gap produces a salient discourse variable which calls for a subsequent repair action through which a value comes to be assigned to the pre-constructed inferential variable $(\lambda x[(x -> s)](x = r'))$. The surface result of the preferred order of inferential processing is the marked sequence (s <- r’)—where r’ is liable to be described as an “explanation.” (Bearth 1999: 270)$^4$

So the interpretation of an inferential as an explanation of a locally relevant assumption derives from the fact that it encodes a proposition from which another, locally salient proposition is inferable.

However, as we have seen, inferentials can be used to communicate propositions that are interpreted as conclusions or results. That is, in Bearth’s terms, we must allow not just (s <- r’) but also (s -> r’), i.e., the possibility that the proposition encoded by the inferential follows from s. Bearth does not account for his exclusion of this order of derivation. As he acknowledges (1999: 275 fn 41) in response to Delahunty’s (1997) analysis, the inferential gap theory is “unteachable insofar as it crucially depends on the notion of marked vs. unmarked directionality of inference processing.” Clearly, the marked direction of processing is not a necessary prerequisite for the occurrence of an inferential form.

His formulation of the inferability maxim commits Bearth to the position that inferential gaps (and consequently inferential sentences) can occur only after a speaker has made a claim in a discourse. Inferentials, therefore, should never occur as discourse initial utterances. Our observations of Spanish inferentials, however, show that they can occur

$^4$ We will attempt here to clarify Bearth’s notation. Bearth uses s to refer to a discourse “claim”, (1999: 269) which must be inferable from other elements of the discourse besides just its utterance by the speaker if the discourse is to be coherent. r represents a discourse element (a “trigger”) from which s can be inferred. In the normal course of discourse events s should precede r, i.e. (s - r). However, should s not be inferable from some prior trigger, then the discourse processor creates a trigger variable, x, from which s might be inferred, represented as (x - s). This variable is bound by the lambda operator, $\lambda$, which turns the expression of which it is a part into a predicate, represented as ($\lambda x[(x - s)]$). Subsequently, another discourse element may be presented from which s can be inferred, r’. This element, r’, is then processed as an argument of the lambda predicate and the value of the variable x, which results in the expression (r’ - s).
without any prior verbal context (see example 18 below). Declerck (1992: 221) allows for this possibility. He claims that “[w]e automatically interpret [inferentials] as offering an explanation of an interpretation of a situation or proposition mentioned in the context or inferable from it.”

Indeed, Bearth’s inferability maxim and his “law of sequential ordering,” which requires logical antecedents to precede logical consequents (1999: 269 fn. 31), seem to be unmotivated accretions to theories of discourse coherence. The law of sequential ordering is just a special case of Grice’s maxim “Be orderly” and both it and inferability follow from Sperber and Wilson’s relevance principle, in that the opposite order would require greater processing and would therefore be dispreferred, except where the extra processing costs are either eliminated or recompensed by greater contextual effects (see Blass 1993).

While Bearth does attempt to provide an account for how inferential gaps may be repaired, he does not explain why inferentials are “specialized” inferential gap fillers. This is not surprising, given the lexical nature of Swahili, Akan, and Wolof inferential marking. However, he does not raise the issue at all, even for French inferentials which are lexically and syntactically parallel to English and Spanish inferentials. We account below for both the hows and the wherefores of inferentials.

Delahunty (1990, 1995, 1997) demonstrates that the range of interpretations of inferentials includes explanations, causes, reasons, interpretations, reformulations, results, consequences, as well as evidence (at least in Japanese). He attempts to provide a pragmatic account for these interpretations. In the 1990 and 1995 papers the account is framed in Gricean terms; in the 1997 paper it is framed in relevance theory terms. Neither account is successful in limiting the interpretations to just those attested in the data. The 1995 account does however point to the metalinguistic inference monitoring and repair functions of inferentials. It claims that the metalinguistic or metadiscoursal function of inferentials is a generalized conversational implicature, while the specific interpretations, such as explanation, reinterpretation, and result, are particularized conversational implicatures.

The 1997 account proposes that inferentials impose a processing constraint on the relevance of the proposition represented by the inferential clause, specifically, that the proposition represented by the clause and assumptions derived from the local context are related as antecedent and consequent in an implication: If P then Q. Crucially, each class of assumptions can play either of the roles in the implication. That is, the inferential proposition may be either P or Q, while contextual assumptions play the complementary role. When the inferential proposition plays the antecedent role it is interpreted as a cause, reason, and the like; when it plays the consequent role, it is interpreted as a consequence, result, and similar roles. Missing from this account is a place for inferentials which are interpreted as (re)interpretations, or reformulations of elements of their context, and a full integration of the account into relevance theory. What is valuable about the 1997 analysis is that it provides a deductive basis for the interpretations of inferentials. The 1997 presentation is consistent with, but not fully derived from, relevance theory. It is this approach that we develop here in such a way as to be able to account for all of the observed interpretations of inferentials and to better integrate it into RT.
6. A relevance theory analysis of inferentials

In this section we provide an account of how inferentials function in discourse, of how they have the interpretations that we have observed, and of why the inferential form has just such functions and interpretations. Readers will recall that the inferential consists minimally of a matrix clause comprising a null or expletive subject, a form of the copula, and a subordinating conjunction introducing a complete tensed sentence. Our analysis makes crucial use of the fact that the matrix is a semantically expletive device that marks an entire sentence.

We begin with two uncontroversial claims. First, by virtue of the expletive matrix the inferential form is a marked construction. It marks a complete tensed sentence-level constituent. Hence the claim that it is a sentential focus construction. Second, because the marked constituent is a complete tensed sentence, it represents a propositional level unit in the semantics.

Somewhat more controversially, we claim that the expletive structure is semantically null. In support of this claim, consider the following points.

First, the subject of the matrix clause cannot be questioned:

(10) *¿Qué/cuál no es solo que sea cotilla?
*What is not only that I am nosey?

In having a null subject, the inferential is like Spanish cleft sentences such as the following from Butt and Benjamin (1988: 390):

(11) Es esa chica la que/quien lo hizo.
‘It is that girl who did it.’
(12) Fue usted el que/quien lo dijo.
‘It was you who said it.’

And in spite of the different verbs, the inferential is also similar in this respect to Spanish meteorological and existential sentences such as:

(13) Estaba lloviendo.
‘It was raining.’
(14) Hay mucha gente en la calle.
‘There are lots of people in the street.’

Our claim is that the it subject of English inferentials and the null subject of Spanish inferentials are both referentially and semantically null. Other analysts, notably Hedberg (1990) have argued that the it of it-clefts (e.g. It’s the woman who buys the house from Studs Terkel “Working” cited in Hedberg 1990:1) and of inferentials is referential. In support of this position Hedberg claims that the it of clefts is parallel to this or that in the subject of th-clefts (e.g. This was Jeanette who told me, [Hedberg 1990: 4]) which are

---

5 The inferential clause corresponds to the focus of a cleft sentence, not its clause, which accounts for the difference in complementizer systems that introduce the two clauses in Spanish.
clearly referential. She also points out that there are instances of clefts and inferentials in which the matrix subject has a discourse antecedent. She does, however, concede that expletive *it* does occur in English in meteorological sentences, and we assume that she would agree that the null subject of Spanish meteorological sentences is also expletive. There is no question in our minds that *th*-clefts exist and that their matrix subjects are referential. There is also no doubt in our minds that “cleft” sentences with a referential matrix *it* subject also exist. Given the nature of the copula, these and many other sentence types are possible. However, Hedberg offers no direct demonstration that the subject of a cleft matrix clause cannot be expletive, that it must be referential. She presents the parallel with *th*-clefts and assumes that readers will draw the same conclusions that she drew. However, it needs to be directly demonstrated that the *it* of inferentials is referential.

We detect an ambiguity in some inferentials that follow a suitable discourse antecedent, such as the following:

> (15) Si el incremento observado en la temperatura es realmente un calentamiento de tipo ‘invernadero’ y no solo “ruído”—una fluctuación al azar—se podría justificar la diferencia de varias formas. Quizás los modelos sean simplemente el doble de sensible a incrementos pequeños en los gases que generan el efecto invernadero, o quizás la red de termómetros, que no es ni completa ni homogénea, haya subestimado el calentamiento global. Posiblemente, otro factor, que no está bien explicado en el modelo, está retrasando o contrarretrasando el calentamiento. *Puede ser que la capacidad de calor de los mares sea más grande que la calculada mediante los modelos actuales, que la producción de calor del sol haya desminuido ligeramente, o que los volcanes hayan inyectado más polvo en la estratosfera que lo que se sabe actualmente, reduciendo así la energía solar que llega a la tierra.* (English original)

‘If the observed temperature increase really is a greenhouse warming and not just “noise”—a random fluctuation—one might account for the disparity in various ways. Perhaps the models are simply twice too sensitive to small increases in greenhouse gasses, or perhaps the incomplete and inhomogeneous network of thermometers has underestimated the global warming. Conceivably some other factor, not well accounted for in the models, is delaying or counteracting the warming. *It might be that the heat capacity of the oceans is larger than current models calculate, that the sun’s output has declined slightly or that volcanoes have injected more dust into the stratosphere than is currently known, thereby reducing the solar energy reaching the ground.*’ (Schneider 1989)

The Spanish highlighted sentence may be interpreted as having an anaphoric null subject whose antecedent is *otro factor* ‘some other factor.’ However, this sentence may also be paraphrased, a la Declerck, as “The explanation/reason for the delay in global warming might be . . .”, that is, as a true inferential. On this interpretation, the Spanish null subject and English *it* are expletive.

On Hedberg’s analysis, the matrix subject and the clause of clefts are coindexed,
indicating that the clause, at some level of representation, modifies the pronoun. Because inferentials, as we have defined them, and as we have found them, do not include a clause corresponding to the clause of a cleft, there is nothing for the inferential *it* to be coindexed with.

Second, besides having an expletive subject, the form of *ser* in the matrix must always be third person singular:

(16) *Soy/eres/somos/sois/son que ha fracasado en todo.

‘1sg/2sg/1pl/2pl/3pl that has failed in everything.’

If the verb never changes and its subject is always expletive, then we have good evidence that we’re dealing with an expletive or lexically meaningless element.

Third, as we discussed earlier in regard to the specificalational reading of copular clauses, *ser* contributes no lexical meaning to the sentences it occurs in. In fact, many languages have no copulas and others optionally omit them. According to Lyons (1977: 471), the copula “is a meaningless lexeme whose syntactic function it is to convert whatever it combines with into a verbal (i.e. predicative) expression.” The subject and predicate contribute all the lexical meaning there is in copular sentences. *Ser* contributes certain categories of grammatical meaning, specifically tense, mood, person, and number, but it does not otherwise semantically constrain its subject and complement—they constrain each other. It should be remembered also that the ability to convey meaning depends upon the possibility of contrast. As no other verb can occur as the matrix main verb in inferentials, no lexical contrast can occur, and hence no lexical meaning can be communicated. As the only contrasts that can occur are grammatical, only grammatical meanings may be communicated by *ser*. It is possible that in contexts in which *ser* can contrast with other verbs it does communicate lexical meaning. This possibility, however, does not affect the validity of our claim. In fact, English *it*, *there*, and *that* provide examples of elements which are meaningless under some circumstances and meaningful under others, namely those in which they can be replaced by contrasting elements.

*Ser*’s semantic emptiness is consistent with the expletive nature of its subject. Because *ser* has no lexical meaning it cannot semantically require a subject, and because the material in the predicate of inferentials is syntactically and semantically complete, it creates no syntactic or semantic need for a subject.

Finally, the subordinating conjunction, *que*, is also invariable and therefore meaningless, just like the English subordinating conjunction *that*. These two merely indicate that the construction that follows is a finite subordinate clause. *Ser* and *que* function as grammatical devices for subordinating a clause. The matrix, therefore, is an expletive structure, which, while it serves as a structure for the placement of a negator

---

6 A reviewer takes issue with this claim and mentions languages which have a single preposition that contrasts with its absence and which has a “very abstract relational meaning.” Note that in Spanish inferentials *ser* does not contrast with its absence - it’s an obligatory element of the matrix, so the analogy with one preposition languages fails. We also think that the burden is on those who would claim that the copula has a lexical meaning to characterize that meaning, more specifically and precisely than “very abstract relational meaning.” At the very least, our claim that copulas represent no lexical meaning is falsifiable, whereas the alternative suggestions that we have heard have been too vague to test.
and/or a focusing adverb, primarily serves merely to indicate that the complement clause is marked and therefore requires extra processing. We explain why this is so in our discussion of relevance theory below.

Inferentials are, as Horn (1989) and Delahunty (1995) observe, metalinguistic or metadiscoursal devices. As such they address a range of issues pertaining to the utterance or situation at which they are directed, “including the conventional or conversational implicata it potentially induces, its morphology, its style or register, or its phonetic realization,” as Horn (1989: 363) says of metalinguistic negation. The repair and corrective functions of inferentials observed by Declerck (1992) and Bearth (1999) are just such metalinguistic functions. This, of course, raises the obvious question of how inferentials get to be metalinguistic or metadiscoursal.

The metadiscoursal functions of inferentials derive straightforwardly from the fact that they are marked structures. Marking is, by definition, relational so it implies alternatives and imposes a ranking on them. Because an inferential marks a propositional level unit, the marking implies a ranking among units of that type. The alternatives to the proposition represented by the inferential clause are pragmatically determined. By virtue of its markedness, a positive inferential asserts the relevance of a proposition that is less likely (at least from the point of view of the producer of the inferential) to be considered in the processing of the target utterance or situation; a negative inferential denies the relevance of a proposition that is more likely to be considered in the processing of the target utterance or situation (again, from the point of view of the producer of the inferential). The opposite makes no sense. Why mark a proposition that is already readily available? Why mark and negate a proposition that the audience is unlikely to consider any way?7

We turn now to explain why inferentials have the interpretations we have observed, namely causes, reasons, explanations, (re)interpretations, reformulations, results, conclusions, and consequences. We demonstrate that these interpretations follow straightforwardly from elements of RT.

RT's basic claim is that ostensive communication guarantees optimal relevance, i.e., adequate contextual effects for no gratuitous processing effort. These implications are derived by decoding the linguistic form, enriching the result in a manner consistent with the search for relevance, and then deriving further inferences. The first interpretation consistent with optimal relevance in a way that the speaker could manifestly have foreseen is taken by the hearer to be the one the speaker intended.

It is in principle possible to imagine a very large range of potential interpretive relations between a pair of contiguous utterances, including cause, explanation, evidence, justification, motivation, implication, contrast, problem, solution, evaluation, or irrelevance, and many more that have been mooted in the pragmatics and discourse analysis literature. A hearer will posit a relationship consistent with the principle of relevance. Speakers can, however, constrain hearers’ assumptions by the addition of various

---

7 Ordinary cleft sentences may be understood in pretty much the same way. Positive clefts assert the greater relevance of the denotation of the focused constituent to the cleft clause, in relation to other pragmatically determined alternatives of the same type; negative clefts deny the relevance of the denotation of the focused constituent to the cleft clause, in relation to other pragmatically determined alternatives of the same type. Clearly, where no clause is present, the denotatum of the focal constituent must be relevant in the local non-linguistic context.
expressions such as after all, so, and the like, which according to Blakemore (1987) impose semantic constraints on relevance. Alternatively a speaker may choose a specialized syntactic form, such as the inferential. RT allows for a range of possible relevance relations between linguistic form and meaning (Wilson and Sperber 1993), several of which are potentially applicable to inferentials.

First it is conceivable that inferentials license pragmatic assignments to the explicature, like those described in Carston's (1988) analysis of and. However if this were so the matrix should contribute to the truth conditions of the overall proposition, which should then fall within the scope of logical operators. The non-contradictoriness of the following example shows that we are not dealing with aspects of the proposition expressed.\(^8\)

\[(17)\] Normalmente, por principio evito que mis amigos y conocidos se conozcan. 
\(No es que uno tema la traición, aunque por supuesto lo haga. \)Qué miedo humano es más profundo? Sin embargo, surgen numerosos e innecesarios pequeños problemas de tales relaciones.

\(8\) Normalmente, por principio evito que mis amigos y conocidos se conozcan. 
\textit{No es que uno tema la traición, aunque por supuesto lo haga. } ¿Qué miedo humano es más profundo? Sin embargo, surgen numerosos e innecesarios pequeños problemas de tales relaciones.

\(\text{(English original)}\)

\(\)‘On principle I usually avoid introducing my friends and acquaintances to each other. \textit{It is not that one fears treachery, although, of course, one does.} What human fear is deeper? But endless little unnecessary troubles usually follow from such introductions.’ (Murdoch 1975: 43)

Alternatively, inferential sentences might grammatically encode procedural, non-truth conditional, and non-conceptual constraints on relevance like those described in Blakemore's (1987) analysis of discourse particles. However these expressions seem to have quite determinate interpretations; one use of \(so\), for example, indicates that the expression it introduces is relevant as a conclusion. Such implications are "stipulated in the grammar" (Wilson and Sperber 1993: 6). The inferential is relevant in much less determinate ways, so its implicatures appear not to be grammatically encoded.

What we seem to be dealing with is a processing constraint on relevance: the expletive matrix forces the hearer to do extra processing, which must, according to RT, communicate extra contextual effects (see Blass 1990; Moeschler 1993; Wilson and Sperber 1993). These extra contextual effects are the inferential interpretations we have observed. They follow from the elements of RT that we describe next.

According to RT, interpreting an utterance in context involves a deductive process. Sperber and Wilson (1987: 702) argue that “the spontaneous and essentially unconscious formation of assumptions by deduction is a key process in non-demonstrative inference. More generally, the ability to perform deductions provides the mind with a uniquely adapted means of extracting more information from information it already possesses, of deriving the maximum cognitive benefit from new information, and of checking the mutual

---

\(^8\) The non-contradictoriness of this example is noteworthy in comparison with the clear contradictoriness of its non-inferential counterpart: \textit{Uno no teme la traición, aunque por supuesto lo haga.} ‘One does not fear treachery, though of course one does.’ If the inferential form licensed assignments to the explicature then the inferential should be as contradictory as the non-inferential.
consistency of its assumptions.”

There are three elements in the interpretive process:

a. the utterance to be processed;
   i. implicated premises
   ii. implicated conclusions

(Sperber and Wilson 1986: 193-202)

According to Sperber and Wilson (1987: 705), “implicated conclusions are deduced from the explicit content and its context,” and “implicated premises are added to the context by the hearer, who either retrieves them from memory or constructs them ad hoc.” Regarding the status of such premises as implicatures, Sperber and Wilson say (1987: 705):

[T]he speaker must have expected the hearer to supply them, or some of them, in order to be able to deduce the implicated conclusions and thereby arrive at an interpretation consistent with the principle of relevance.

To illustrate, consider dialogue (6):

(6) (a) Peter: Would you drive a Mercedes?
     (b) Mary: I wouldn’t drive ANY expensive car.

The explicit content of Mary’s reply does not directly answer Peter’s question. However, processed in a context containing (7), (6b) yields the contextual implication (8):

(7) A Mercedes is an expensive car.
(8) Mary would not drive a Mercedes.

We have a situation, then, in which Mary, in producing (6b), has not directly and explicitly answered Peter’s question, but has made manifest a contextually implied answer, that is, (8). Given that she could not expect her utterance to be relevant unless it made manifest such an answer, this implied answer is manifestly intentional: It is an implicated conclusion of her utterance. Since it is manifest that Peter would not have deduced this conclusion without adding (7) to the context, (7) is an implicated premise of Mary’s utterance. (Numbers as in original.)

As metadiscoursal or metalinguistic devices, inferentials may be used to address implicated premises, implicated conclusions, or the target utterance itself. If they are directed at implicated premises, then they will be interpreted as explanations, reasons, causes for the (situation described by the) utterance. If they are directed at implicated conclusions, then they will be interpreted as results, conclusions, or consequences of the utterance proposition. If they are directed at the utterance to be processed, they will be interpreted as interpretations, reinterpretations, or reformulations.

Spanish inferentials may be the first utterances in a situation, contrary to Bearth’s analysis, which requires a prior claim to license an inferential, and to Declerck’s (1992:
claim that in the unlikely event that an inferential is used discourse initially, “the missing variable must be inferable from the context that follows” (our emphasis). In the following example it should be clear that Spanish inferentials may be the first utterances in a situation and that they deal with mental representations of context that are created prior to (not following) their utterance. The example in (18) was said by a husband to his wife when she walked into their house and found that he had made a mess, before she said anything (though after she looked exasperatedly at the scene):

(18)  Es que estaba buscando las ceras para la niña.
     ‘It’s that I was looking for the crayons for the child.’
     (Observed/collected by Laura Gatzkiewicz)

Because Spanish inferentials may be used discourse initially, our theory must allow them to be directed at the mental representation of a situation. Interpreting a situation requires us to formulate a mental representation of it, and to select/construct contextual assumptions (premises and conclusions), aided perhaps by non-verbal or para-linguistic activities of the participants. Just as with utterances, inferentials may be used to address the premises or conclusions constructed in interpreting the situation, as well as the mental representation of the situation itself. Both uses involve the construction of mental representations, and both involve the adjustment of those representations. It is as if the inferencing in the interpretation of (18) went something like this:

I think that you think I have done something wrong. But it’s not what you might think. I can explain and justify the situation. It’s that . . .

Clearly the inferencing associated with discourse initial and non-initial uses of inferentials are identical. In both instances the inferential seeks to repair a mental representation, that is, an interpretation, of a situation. They differ only in that for the non-initial uses, the interpretation is based on the processing of an utterance; in the initial uses it is based on the processing of a non-verbal situation. Crucially the same pragmatic principles can account for both situations. However, as we have not found a single English inferential that is the first utterance in its discourse, we tentatively conclude that English speakers do not use inferentials as discourse initial utterances. To account for this apparent difference in the distribution of English and Spanish inferentials, we must appeal to a distinction between pragmatics and discourse patterns. We suggest that English and Spanish speakers operate with the same universal pragmatic principles but differ in their assumptions about the discourse conditions under which the inferential may be used. At this point in our research, it must be said that we do not have a more principled account of this difference between Spanish and English (and apparently Italian) uses of the inferential; nor, indeed, are we certain that English speakers never use inferentials discourse initially. If it turns out that they do, then this apparent difference between Spanish and English will (conveniently) disappear.

To see how our account works, let’s examine the following relatively complex example. This example contains three *ser que* sentences in a row, the first of which is negative and the other two are positive. It is not unusual to have two inferentials in a row, the first negative, the second positive. The negative sentence denies an inference or
interpretation that is quite reasonable in the discussion, in this instance, that Manuela’s mother doesn’t love. The two positive inferentials assert alternative, but less likely, interpretations of the woman’s character: that she doesn’t know how to love and that she has failed at everything.

(19) Ángeles García: Puede ser una comprensión interesada en el sentido de que también es muy difícil asumir que no nos quieran. Especialmente difícil de asimilar si quien no te quiere es tu padre o tu madre.
Maruja Torres: Claro que es interesada, de lo contrario sería imposible salir adelante. Pero quiero matizar. La madre de Manuela, que está inspirada en la mía, no es que no quiera: es que no sabe querer, es que ha fracasado en todo; y la hija es para ella la prueba vivente de su fracaso. Y eso, a mí, me da un punto de partida para la compasión. Otra cosa es el caso de la tía, que ni quiere a nadie ni nunca sabrá lo que es el cariño. (El País 1/19/1997)

‘Ángeles García: It can be an understanding motivated by self-interest in the sense that it is also very difficult to accept that we are not loved. It’s especially difficult to assimilate if the person who doesn’t love you is your father or your mother.
Maruja Torres: Of course it’s self-interested, if it weren’t it would be impossible to go on. But I want to make some distinctions. Manuela’s mother, whose character is based on my mother, it’s not that she doesn’t love; it’s that she doesn’t know how to love, it’s that she has failed at everything, and the daughter is living proof of her failure. And this, for me, gives us a starting point for compassion. Something else is the case of the aunt, who neither loves nor will ever know what affection is.’

According to our account a negative inferential denies the relevance of a proposition that is readily accessible in the context. The proposition in question is that Manuela’s mother does not love, a proposition clearly implicated by García. And Torres uses the negative inferential to deny that proposition.

Our account says of positive inferentials that they assert the relevance of propositions less plausible in the context. Clearly, as the proposition La madre de Manuela no quiere ‘Manuela’s mother doesn’t love’ is the most salient in the context, the alternatives must be less plausible.

Our account also says that the inferential may address any aspect of the processing of the prior utterance, including its representation of a situation or premises or conclusions implicated in its interpretation. Just which of these three will be addressed is determined in the hearer’s search for an interpretation consistent with the principle of relevance. In the current instance, the most relevant target would be the formulation of an understanding of Manuela’s mother’s character (and given the interviewer’s interest in finding models for the novel’s characters in the author’s life, also for an understanding of Torres’ mother’s character). Given the prominence of the issue of the effects of not being loved by a parent, on both Manuela and perhaps on Maruja Torres herself, and particularly given the interviewer’s explicit formulation of the issue as being that of a parent not loving a child, and the fact that the negative inferential clause repeats the interviewer’s formulation almost verbatim (no te quiere ‘doesn’t love you’ and no quiera ‘doesn’t love’) the most accessible
assumption is that the negative inferential is addressing the formulation of the issue. The negative inferential provides adequate contextual effects as it denies the relevance of a proposition proposed by a speaker in the context.

The smallest context in which the second, positive, inferential is processed is that which is created in the processing of the first inferential, namely that the representation of Manuela’s mother as not loving is irrelevant. It follows that the least costly interpretation of this second inferential is as the assertion of a proposition which is less salient in the context, and as an alternative representation of Manuela’s mother’s character, given that with the denial of the proposition that she doesn’t love, we are left without a relevant representation of her character.

The least costly context in which the third inferential can be processed is that created in the processing of the second one. As we now have a representation of Manuela’s mother’s character, and as the third inferential is syntactically parallel with the second, the least costly interpretation with adequate contextual effects is that the third is either an alternative or an additional representation of Manuela’s mother’s character. As the speaker gives no clue as to which of these options is the intended one, for example adding ‘or’ or ‘and,’ we are left with an indeterminacy. This is of course quite compatible with RT, and is just how we have observed inferentials to behave.

7. A final issue

As we noted earlier, inferentials occur in a wide array of unrelated, non-contiguous languages with analogous forms and very similar meanings and discourse uses. It seems most unlikely, given the global distribution of these languages, that inferentials are an areal phenomenon. The most likely assumption is that the form may be created independently in any language that has the requisite lexical and syntactic characteristics required to take advantage of the interaction of those characteristics with universal pragmatic principles. We hypothesize that any language with the resources to create an expletive matrix that can subordinate a clause or sentence will have the inferential construction with its characteristic interpretations. These resources include a null or expletive subject, actual or zero copula (or some other semantically neutral verb such as Mandarin shi . . . de [Hedberg 1999]), and perhaps a subordinating conjunction. The universality of the uses and interpretations follows from the universality of pragmatic principles such as those articulated in RT. One factor that is likely to block the development of an expletive matrix inferential in a language is the existence of an inferential particle in that language.

8. Conclusion

To conclude we return to the question of where inferentials may occur in discourse. Bearth’s (1999) answer to this question is whenever an inferential gap needs to be filled. Our answer is that inferentials occur when a speaker believes (or at least wishes to act as if s/he believed) that an addressee might process the mental representation of an utterance or of a situation in a context of assumptions other than those intended by the speaker. That is, whenever the potential implicata or the representation itself are other than the speaker
wishes to endorse.

References


