7th International Conference on Corpus Linguistics: Current Work in Corpus Linguistics: Working with Traditionally-conceived Corpora and Beyond (CILC 2015)

Contextual analysis on the expression of uncertainty: certainly versus maybe. Study in a computerized corpus

María José Barrios Sabador*
Nebrija University, Pirineos St., Madrid, 28040, Spain

Abstract

In this study there is analyzed the functioning of two probability markers, certainly and maybe, elements of very different degree of certainty, in a speakers’ oral corpus of Spain. This analysis reveals their presence in causal environments, as well as their recurrence together with diverse probability markers, cognitive verbs or expressions of doubt, along with other phenomena. The behavior of these two exponents reveals notable similarities, though with minor frequency of certainly in some of the analyzed contexts, probably due to its higher level of security.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/). Peer-review under responsibility of Universidad de Valladolid, Facultad de Comercio.

Keywords: certainly; maybe; orality; corpus; Spanish

1. Introduction

The study of certainly (seguro que) and maybe (tal vez) presented here is part of a wider research¹ which aims to analyze the performance of probability markers in oral texts. This wider research is part, in turn, of an academic work that examines the properties of these elements in their semantic, syntactic and pragmatic aspects. The linguistic

* Corresponding author. Tel.: +34-91-452-1100; fax: +34-91-452-1110. E-mail address: mjbarrio@nebrija.es

¹ Barrios Sabador, M.J. (2012)
expression of probability, which could be defined as the “manifestation of an oriented uncertainty”, shows a great
certainty in the descriptive level, given the high number of lexical and morphological resources involved. Unlike
the doubt, which reflects speaker’s perplexity, the expression of probability provides an interpretation of a given
state of affairs, but not affirmed as true. The expression of probability in the Spanish language uses a wide range of
both lexical and grammatical elements with different semantic and pragmatic values regarding the level of
commitment that the speaker is willing to assume, on a scale distributed configuration. Such elements are fuzzy -
following the designation of Lakoff (1973) and Zadeh (1975) - and show vague boundaries due to their semantics
and their dependence on other members that form a scale that goes of the total uncertainty up to the absolute
certainty.

In order to elucidate the behavior of probability components in authentic contexts of communication, we turned
to oral texts by the Reference Corpus of Current Spanish (Corpus de Referencia del Español Actual – from now on
CREA), focusing on the area of Spain, where twenty three markers were investigated with a total of five thousand
paragraphs. We restricted ourselves to the analysis of certainly and maybe, both highly different in their scalar
nature. The review of certainly yields 196 cases, representing 3.9% of the corpus investigated; while maybe has 228
matches, equivalent to 4.6% of the corpus. Our research detected the presence of the probability marker in a series
of environments belonging to the semantic area of the causality, such as causal, concessive, adversative, illative and
conditional constructions. Other properties in the markers’ context are the presence of cognitive verbs, expressions
doing doubt and disjunctions. Less frequently, but not because of it less meaningful, the modal marker is inserted into
an assertive clause. Finally, it can be said that the basic illocutionary aim that prevails in most of the contexts is the
attenuation of the assertive act, apart from those which show the speaker’s true insecurity.

2. Certainly and maybe in causality environments

As we said above, our research detected the presence of the probability marker in causal, concessive, adversative
and illative clauses. It is also found in conditional constructions, if to a lesser extent. All of them share, from
different perspectives, a relation of cause - effect. The causal and concessive constructions are based on the reason,
whereas illative sentences rely on the effect. Given the properties linking probability statement (qualified as weak
assertion) to assertive statement, it’s not surprising that causation is an enabling semantic field to modalization of
uncertainty.

| Nomenclature |
|--------------|
| CREA         | Reference Corpus of Current Spanish (Corpus de Referencia del Español Actual) |
| NGRALE       | New Grammar of Spanish Language (Nueva Gramática de la Lengua Española) |

2.1. ‘Certainly’ and ‘maybe’ in causal constructions

The presence of a probability marker in a causal construction is in agreement with the semantics of these
constructions, whose role is to supplement the meaning of the main verb, adding reasons or justifications of the state
of affairs described (NGRALE 2009: 3461). The insertion of the probability marker in these structures is examined
in two ways. On the one hand, we study the causal sentences that justify the declaration of uncertainty - declaration
that constitutes the effect member -. On the other hand, we look at the probability sentences that endorse an
assertion, that is to say, those with the marker in the cause member.

After verifying the existence of probability markers in causal environments, one might expect the greatest
abundance of probability sentences supported by an assertion, at the expense of the opposite case. However, the data
reveal a parity of both formations, with 49.95% of justification for the assertion and 50.05% of explanation of an
assumption. Furthermore, we paid attention to the nature of the causal construction. The causal sentences, as is well
known, meet a double typology that distinguishes between internal (they explain the state of affairs specified by the main verb) and external (they argue why it is said or inferred something). In the cases analyzed in the Corpus, external causals prevail, both explicative and speech act modifier, although examples of real world causals are also found. In this paragraph of maybe, the cause and the effect are asserted. It's the link between them which is called into question:

- Although they have also told me that the floor of the highway tire wears much more than an ordinary floor. Yes. Because it's harder, maybe, or for the stretch marks that it has, in order that the cars hold tight. Yes, yes.

In the next fragment of certainly, we find the explanation of an assumption:

- Well, one that if this, that if the other that ... "get off and sit here! Get out! No! Go out to sunbathe!". Now ¿¿I called her??? by phone, she certainly was sitting there at the door because she has not taken me the phone, so to tell her that such this was ... ???. ...

Epistemic modalization of uncertainty in causal sentences is a clearly outstanding phenomenon in the oral productions of Spanish speakers, with 16.3 % of matches in certainly and 14 % in maybe. The high certainty of certainly can be the reason for which a 11% of the sentences with the adverb are justified by assertions, in contrast with 5 % of assertions that explain the sentences containing maybe. On the contrary, maybe endorses 9.2 % of assertions, in contrast with 5.6% of assertions relying on certainly.

2.2. ‘Certainly’ and ‘maybe’ in illative sentences

In the illative sentences, the marker offers a major number of coincidences in the effect member. That is to say, after making a statement, the speaker extracts an epistemically modalized deduction. Though illative sentences share with causal constructions the same argumentative purpose, they are less common in the corpus (5.1% in certainly; 2.2 % in maybe).

In these constructions, one could assume a major number of texts with the probability marker being part of the effect member than the opposite case, since it seems to be more defensible to draw a conclusion from a statement than inferring something from a conjecture, not committed to the truth of the state of affairs described. Nevertheless, both markers show a 60% of instances where the cause member is epistemically modalized and it’s followed by an asserted conclusion. That can be seen in the next example or certainly, where the effect member is left unfinished:

- Not them if you order more than one, the following are pink They are pink... of ten. They are pink. It had never happened to me. It certainly has happened to Carmen and because of it what? What? Buying more of one? Yes. I there are times that I have bought three at one stroke.

2.3. ‘Certainly’ and ‘maybe’ in hypothetical sentences

The marker of probability in the hypothetical sentence causes, in our view, a weakening of the dependency relationship that links the protasis to the apodosis. Hypothesis reveal points of contact with the assertion, as the

---

2 We follow the tripartite distinction by Sweetser, E. (1990).

3 It should be remembered that we are looking at oral texts, and that at all time we have faithfully respected the original transcription.
speaker contemplates the propositional content as slightly feasibly or clearly counterfactual, for which he has to anticipate or know, respectively, their mismatch with a state of affairs. This character does not meet modified, in essence, by the presence of the marker in the apodosis (10.5% of cases in maybe; 2% in certainly):

- Young people cannot be accused of this lack of interest in something that remains essential since the dawn of time, such as literature. What happens is you have to win them, perhaps differently. Maybe we should win them otherwise.

2.4. ‘Certainly’ and ‘maybe’ in adversative and concessive constructions

In the semantic area of the causality, two constructions linked with some frequency to the probability sentence are concessive and adversative clauses. Both concession and adversativeness are resources of argumentative character. With the adversative strategy, the speaker objects a state of affairs or the linguistic act of the listener, whereas with the concession he assumes it to all appearances, being opposed to it at once (Flamenco García, 1999: 3810-3811).

Olmos and Ahern (2009: 63 - 65) indicate the different intensity of ‘but’ with regard to ‘though’: in the adversative relation, new information is inserted, whereas in the concessive one that information is of discursive nature. A pretty good number of occurrences of the studied markers show the adversative strategy.

![Figure 1. Concessive and adversative constructions with certainly and maybe](image)

Probability markers can also be detected, if to a lesser extent, in “concurrent concessive – adversative formulae” - following the denomination coined by Flamenco García (1999) -, these consisting of an epistemically modalized declaration which opposes an inference. The intended purpose here can be to safeguard the positive image of the speaker or the listener:

- And the public ... when the public picks on you, they call you Indian. Are we a racist people, the Spanish? Maybe... maybe it is possible that there is something of influence, but I think that more or less in all the countries adjectives are commented, as when somebody wants to be a nuisance or to insult someone.

3. Certainly and maybe in other contexts

As we said above, these two probability markers appear alongside cognitive verbs, other probability exponents and expressions of doubt. They are also found in disjunctions and in assertive contexts. Finally, the attenuating function is noticeable in many of the phenomena studied in this article.
3.1. ‘Certainly’ and ‘maybe’ with cognitive predicates

Under this name we refer to predicates such as to believe, to think, to suppose, it seems to (me), etc. In the CREA we have confirmed their presence in 15% of the instances of maybe and in 9% of cases of certainly. With them, the already weak assertive character of the probability sentence, derivative of its non-factive nature - it does not presuppose the truth of its complement- is emphasized. The sensation of incertitude of the probability sentence will be highlighted by such a predicate forming a part of it, but also its presence in the surrounding sentences propitiates a 'contextual environment' of uncertainty that interacts with the probability sentence. Such insecurity can be real or a strategy of lessening of the illocutionary force, eventually destined to support self-image or the face of the addressee.

A property shared by both resources is the creation of a mental space (Fauconnier, 1985/1994: 2007) regarding speaker's beliefs. Like probability markers, cognitive predicates are not part of the declarative content, but they act as signals that guide the listener into the correct insight of exposed content, consisting that of a “read with care” (Urmson, 1952: 495). In these predicates, the individual is configured as the center of the cognitive process, across which the conception of the situation represented by the clause that follows is filtered (Schack Rasmussen, 2000: 316). In spite of the semantic similarities that link probability markers to cognitive predicates, both with a double function of allegation of uncertainty and of mitigation of the illocutionary force of the speech, these two procedures differ in the role of the conceptualizer and in the perspective adopted in the construction of the entity or situation (Langacker, 1990, 1991/2002). With the cognitive verbs, the grounding takes place by means of the presence of the speaker across the verbal morpheme of the first person, sometimes reinforced by the pronoun. The conceptualizer is in the scene, for what these elements are minimally subjective, with the conceptualizer being part of the conceived object. In contrast, probability markers leave the conceptualizer out of the predication, since the predication constitutes the focused entity, which implies a subjective role of the conceptualizer and the objectification of the scene.

Given the uneven configuration implied by these means, it is legitimate to wonder about the effect of their conjunction in a statement, a relevant question given the relatively high frequency of this combination in the analyzed paragraphs. The markers of probability profile a situation sketched by the conceptualizer as uncertain and place it in the foreground of the communicative act. In clear contrast, the cognitive verb implements a different framework inasmuch as the conceptualizer is in his enunciation. However, such a conciliation of asymmetric resources must not be considered to be unmotivated, since the speakers are linguistically competent and we presume their choices as pertinent.

The cases of cognitive predicates close to the probability sentence are arranged in two groups, according to their meaning: modal or expression of an opinion. This organization is based on the classification of De Saeger (2008), who distinguishes, following Nuyts (2001) two uses: representational and qualificational. In the first one, the verb is a component of the object of conceptualization, that is to say, it is part of the propositional content of the utterance. De Saeger includes here the reported speech (At the time I thought, “This is not the way to the shelter”) and the description of mental acts or states (He supposed that he would have no choice but to take a drastic action). In its second mission, the verb of cognition does not have a propositional meaning: it qualifies the content of the statement rather than being part of it, therefore implying the argumentative attitude of the speaker. The representational – qualificational axis combines with a second coordinate incorporating the perspective of the subject and that of the speaker. Such a connection allows to perceive two qualifying employments of these verbs: expression of opinion (from the perspective of the subject) and of doubt (from the perspective of the speaker - modal-). In the analysis of the confluence of probability markers and cognitive verbs, uses of probability are predominant over those of opinion.
The percentage analysis shows little differences in certainly (none in maybe) between the probability sentence and the previous one. Their frequency is slightly lower for both markers in the next sentence.

The presence of cognitive verbs is in line with the establishment of an epistemic mental space where the probability marker is fully justified. In spite of the unlike conception of the scene imposed by both means, their coexistence reflects the need of the speaker to recreate a discursive framework where the probability sentence arises with support of these verbs - so much if they transmit uncertainty as opinion - as semantically and pragmatically motivated. Let’s look at an example where believe express an opinion after the probability sentence with certainly:

• But I have to tell you also that to affirm that a process of regeneration is necessary can give the impression, that it’s certainly not what you want to say to me, of that the public life in Spain is degenerating, and I would not agree with a statement of that kind. I believe that the public Spanish life, here, as in all the countries, the management of the public power sometimes demands an explanation to citizens.

3.2. ‘Certainly’ and ‘maybe’ with other probability markers in the sentence and discursive environment

Two of the most recurrent phenomena in the performance of the probability markers analyzed here are, firstly, the resort to more than one probability exponent in the sentence and, secondly, the emergence of several probability
exponents in the surrounding context. In order to examine this phenomenon we started from a quantitative criterion (number of markers in the sentence or discursive environment), which led us to a description of qualitative nature (which these markers are and their relationship to that to which they are related).

Concerning the quantitative analysis, maybe has a 14% of other markers in its sentence and 16% in the discursive context, while certainly gives a lower percentage in both cases: 8% of markers in the same sentence and 13% in its immediate context. As for the qualitative analysis, it would be conceivable that the supplementary markers were of similar scalarity, as they modalize the same propositional content. Such hypothesis is verified in the repetition of the marker, term we have chosen to designate the assiduous duplication of the concerned exponent. Given this pattern of repetition, it’s hard to determine to what extent this is a reinforcement of conjecture or a repetition of some parts of speech, something common in orality, motivated by external factors and without any other intention on the part of the speaker. A part from this repetition, the analysis of maybe and certainly allows to verify that other markers in the sentence environment show a wide range of levels of uncertainty.

In any case, regardless of the degree of certainty of exponents, the proximity of other markers consolidates the mental space of possibility established by the analyzed exponent or favors the emergence of new spaces, more or less linked to the one built by the main marker. One of the texts of certainly exemplifies that:

• A few cases it is known more and in other cases it is known less. But, there is nobody who is perfect and I certainly must have, and I have, many faults, and that are known, what happens is that sometimes generously then the people do not mention them.

In the discursive context, the markers contribute to the creation or the consolidation of the mental space of uncertainty, where the studied operator is assembled in a contextual network, formed by the adjacent probability sentences.

---

4 It is essential to notice the approximate character of the translation of the exponents, given the larger number of lexical exponents of Spanish with regard to English. For example, maybe is translated sometimes as a lo mejor, sometimes as quizá(s) or tal vez (this last one being object of our study, along with seguro que - certainly -, also found as to be sure that).
3.3. ‘Certainly’ and ‘maybe’ with expressions of doubt and disjunctions

Expressions of doubt near the probability marker help to reaffirm the impression of uncertainty about what is said. The employment of these resources might qualify as a type of hedging, in order to protect the individual from the potential criticism that would follow an assertive act. However, it does not imply that the indication of insecurity carries, unavoidably, attenuation, since it can be real. The expressions of doubt, noticeable in 9.2% of occurrences of maybe and in 1.5% of the paragraphs of certainly may appear in the probability sentence, in the preceding, subsequent or in all of them and contribute to the backing of the mental space of uncertainty shaped by the marker.

As for disjunctions, we observe that the speaker opts sometimes to resort to diverse alternatives at the moment of verbalizing his insecurity. Bringing into play two or more possible worlds imply, in our view, most incertitude on the part of the speaker, unable to suggest only an approximate reasoning. In these cases, we are obviously dealing with disjunctions of exclusive type, that is to say, disjunctions where $p$ and $q$ cannot be true simultaneously (Escandell Vidal, 2004: 117-118).

- I have discussed with the people on this issue and, well, I have finished mad, really, because I have not known how to make them understand maybe because I don’t know how to explain myself or I do not have ease to express that I mean ... Y es, yes. ...

3.4. ‘Certainly’ and ‘maybe’ in assertive statements

The assertion is opposed to the probability statement, definable as a weak assertion, for offering a guarantee of approximate veracity of the propositional content. Given this essential distinction, the coexistence of both classes of statements in the same sentence turns out to be paradoxical. Nevertheless, and in spite of not being very frequent, some examples have been found. This phenomenon materializes in two different ways. On the one hand, by means of adverbs such as undoubtedly, evidently, etc. On the other hand, the probability exponent is inserted in the area of assertive predicates (it is true that, there is no doubt that, etc.) and factive predicates (to see, to happen, to realize, etc.). Let’s see these two examples:
• Whereas in CAI Saragossa have we Indian Díaz that they are just now in the bench, with seven points and Turping? With eight. Playing in assault Real Madrid. Maybe this change of Indian Díaz for Pepe Arcega owes, undoubtedly, to that it is the most sticky marking that of Pepe Arcega.

• Consequently, it is true that everything what means to answer questions on the part of the prime minister to answer questions in the chamber, certainly benefits the government, because it will allow us to explain the most questioned aspects of the governmental policy to the citizens in a better way.

The inventory of coincidences of assertion shows the superiority of cases in certainly (4.6 %) opposite to maybe (2.6 %). This finding is significant if we consider the greater certainty of the first exponent versus the second.

3.5. ‘Certainly’ and ‘maybe’ as hedges

The expression of uncertainty is one of the strategies used by the speaker to mitigate the illocutionary force of the assertion. That is due to its status as non-polite act, that is to say, an illocutionary act whose object does not benefit the interlocutor (Haverkate, 1994: 116). Mitigation is found in 13.2% of the occurrences of maybe and in 5.1% of the instances of certainly.

A phenomenon that has called our attention is the concentration in certainly of examples of attenuation in which threat is not perceived to self or others image. In all these texts, mostly belonging to radio or television programs, certain knowledge or desire to listen to the program is attributed to the audience or to someone who is interviewed. A possible explanation of such prevalence is that the image here defended is that of the listener, which would justify the recurrence to high certainty markers:

• You can see his work in the art room of Virgen del Amparo street, number five, the Provincial Caja de Guadalajara. He's been there a few days and certainly many of our listeners have known his work, but just in case, we will describe it.

The attenuating function is obvious as well in texts where the speaker comments on daily activities, reports on his preferences or his character or he's censoring himself somehow:5

• They do not approach the young people so maybe as they had to. Exactly. My parents, in this sense, are traditional, and maybe what I do miss a bit is a little dialogue, right? It's not that it doesn't exist, but having a clash of generations between my parents and mine, and education, a clash of education, then, it doesn't exist, probably, the dialogue that I would like. Also partly my fault, right? or because of them, I don't know, I think that it is it.

In these occurrences the operator is located in a declaration that concerns squarely the perception that the speaker has of himself: personality, capacities, interests ... Questions, definitively, seemingly more inclined to be formalized in assertive sentences or at least in factive ones, for which mitigation is the ultimate purpose of the expression of uncertainty.

5 The intervention of the subject abounds in probability markers in order to modalize declarations about what the speaker lives at home and his relation with his parents, that is to say, comments on something that he clearly knows. Such a modalization would have as purpose not to be excessively critical in these personal aspects. Note also the presence of mitigating expressions of assertion (“in that sense”) or the immediate precision of the speaker after the probability sentence, possibly motivated by the fear of being misunderstood (“It’s not that it doesn’t exist, but ...”). We believe that the same purpose lies in the final sentences of the fragment (with the conditional tense with a hypothetical value), where the speaker does not know who to put the blame, finished off by a sentence with a cognitive verb.
4. Conclusions

The observation of certainly and maybe in the oral corpus of CREA shows that certainly does not differ from maybe in many of its contexts of appearance, despite their different position in the scale of uncertainty. The majority of phenomena analyzed in maybe take place likewise in certainly, though usually in low proportion. Now then, certainly is characterized by its behavior in a series of environments that ratify their higher scalar position versus maybe. Thus, it presents less frequency of expressions of doubt and disjunctions that maybe, along with a greater emergence in assertive contexts. To a lesser extent, their greater certainty is shown in the lower presence of operators in its sentence and discursive environment. Such a pattern would be indicative of the top position of the adverb in the probability scale. Once underlined these peculiarities of the adverb, we can affirm that the variability in the expression of the probability, if we look at the results provided by the corpus, must be relativized to a certain extent, since the degree of certainty conveyed by certainly does not always trigger remarkable effects in its sentence or in the discursive framework that surrounds it.

References

Barrios Sabador, M. J. (2012). La expresión de la probabilidad en el español nativo y no nativo. Una aproximación descriptiva, gramatical y pragmática. PhD dissertation. Universidad Nebrija.
De Saeger, B. (2008). Speaker Involvement through Cognition Verbs in Spanish. Belgian Journal of Linguistics, 22 (1), 63 - 81.
Escandell Vidal, M. V. (2004). Fundamentos de semántica composicional. Barcelona: Ariel.
Fauconnier, G. (1985/1994). Mental Spaces. Aspects of Meaning Construction in Natural Language. Cambridge: Cambridge University Press.
Fauconnier, G. (2007). Mental Spaces. In D. Geeraerts and H. Cuyckens (Eds.), Oxford Handbook of Cognitive Linguistics (pp. 351–376). Oxford: Oxford University Press.
Flamenco García, L. (1999). Las construcciones concesivas y adversativas. In I. Bosque and V. Demonte (Dirs.), Gramática descriptiva de la lengua española, volume 3 (pp. 3805–3878). Madrid: Espasa Calpe.
Haverkate, H. (1994). La cortesía verbal: estudio pragmalingüístico. Madrid: Gredos.
Lakoff, G. (1973). Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. Journal of Philosophical Logic, 2 (4), 458 – 508.
Langacker, R. W. (1990). Subjectification. Cognitive Linguistics, 1 (1), 5 – 38.
Langacker, R. W. (1991/2002). Concept, Image, and Symbol. The Cognitive Basis of Grammar. Berlin: Mouton de Gruyter.
Nuyts, J. (2001). Subjectivity as an Evidential Dimension in Epistemic Modal Expressions. Journal of Pragmatics, 33, 383 – 400.
Olmos, S., and Ahern, A. (2009). Contrast and Propositional Attitude: A Relevance Theoretic Analysis of Contrast Connectives in Spanish and English. Lengua, 119, 51 – 66.
Real Academia Española. Banco de datos (CREA) [online]. Corpus de referencia del español actual. <http://www.rae.es>
Real Academia Española. (2009). Nueva gramática de la lengua española. Madrid: Espasa Libros.
Sweetser, E. (1990). From Etymology to Pragmatics: The Mind-body Metaphor in Semantic Structure and Semantic Change. Cambridge: Cambridge University Press.
Zadeh, L. A. (1975). Fuzzy Logic and Approximate Reasoning. Synthese 30 (3-4, 407 - 428.)