

Presence vs. absence of the preposition *por* before topicalized infinitives in Spanish

The construction of the topicalized infinitive consists of repeating the same verb both in the topic position (as an infinitive) and as an inflected verb form in the comment. Most studies have focused on the variant with the bare infinitive: *doler duele* ‘as for hurting, it hurts’. In contrast, the variant introduced by the preposition *por* ‘to’ (e.g., *por doler duele* ‘as for hurting, it hurts’) has not yet been systematically analyzed.

The aim of this study was to contrast the contexts of use of the two structures in order to determine how they differ at the semantic-pragmatic level. The literature consulted considers that in cases with *por*, the comment expresses a minimum or maximum value relative to the information introduced by the topic. E.g., in sentences like *Por doler, me duele incluso el pelo* ‘As for hurting, even my hair hurts’, the scalar focus adverb *incluso* ‘even’ places the focalized element at the upper end of a (pragmatic) scale. In contrast, the bare infinitive (without *por*) does not seem to be associated as often with scalarity contexts. The context in which it has been most often analyzed is that of an adversative continuation.

Hypothesis: The variant with *por* is more likely to be associated with scalarity contexts (specifically, a very high degree or a maximum value on a scale), and the variant with the bare infinitive is more likely to be associated with adversarial contexts.

Independent variables and stimuli. In all sentences tested, the topicalized infinitive-construction is used to emphasize the assertion in the comment. Furthermore, two contextual factors were added: scalarity (high degree) and adversativity. These were therefore the two independent variables of the experiment, creating four conditions:

(i)	- scalarity	- adversativity	[<i>Por doler / Doler</i>], <i>duele seguro.</i>
(ii)	+ scalarity	- adversativity	[<i>Por doler / Doler</i>], <i>me duele incluso el pelo, y eso que soy calvo...</i>
(iii)	- scalarity	+ adversativity	[<i>Por doler / Doler</i>], <i>duelen, sí, pero no importa: ¡¡¡He ganado!!!</i>
(iv)	+ scalarity	+ adversativity	[<i>Por doler / Doler</i>], <i>me duele hasta el último centímetro del cuerpo, pero no importa: ¡He ganado!</i>

The examples were presented in broader contexts (colloquial dialogues) in which these structures sound natural. In all the target sentences, the two variants can alternate, i.e., both are grammatical. The aim was, therefore, to test the preference for one or the other variant. Four different lexicalizations (infinitives) – *cobrar*, *doler*, *ganar*, and *mentir* – were adapted to the four conditions, thus creating four groups of stimuli. In addition, two control sentences were added. At the end of the questionnaire, there was also an (optional) open field in which the participants could write their own impressions and intuitions about the difference between these two structures.

Participants. We conducted a *forced-choiced test*, for which participants (natives of peninsular Spanish) were recruited through Prolific. Out of a total of 108 participants, we obtained 102 valid responses. The participants were divided into 4 groups, so that each group only saw each condition only once and with four different lexicalizations.

Results. Overall, the variant “*por* + topicalized infinitive” was chosen less often (34,80%) than the bare topicalized infinitive (56,20%). This is consistent with some of the answers in the open field: some participants claimed to never use “*por* + infinitive” in the examples shown, or not to know this structure at all. The sociolinguistic variables did not show any significant difference.

The results partially confirm our initial hypothesis in that scalarity is a statistically significant variable (exact Fisher test, 2-tailed p-value = $1.9390252477606167e-36$), whereas adversativity is not (exact Fisher test, 2-tailed p-value = 1).

Only in + scalar contexts (conditions ii and iv), the variant “*por* + infinitive” is preferred over the bare infinitive, whereas in non-scalar contexts (conditions i and iii) it is clearly disfavoured. On the other hand, the variable +/- adversativity has no (statistically significant) influence on the choice of one variant over the other. This could be explained by the fact that the scalar element (*incluso...*) is found within the same sentence (i.e., it is part of the same proposition), and therefore has a stronger influence on the choice of the topic structure. Contrarily, adversativity only arises in contrast to the following sentence (i.e., in contrast to another proposition).

The free answers of some of the participants point in the same direction: although most of the participants claimed that they only chose intuitively between the two variants and could not explain why, some others claim that they use “*por* + infinitive”, for example, in contexts of “exaggeration” (i.e., in contexts of a very high degree on a scale).

Furthermore, there are differences between the four individual lexical items. Although they all follow the same tendency (i.e., a preference for “*por* + infinitive” in scalar contexts and a clear or even total preference for the bare infinitive in non-scalar contexts), the usage rate of “*por* + infinitive” is consistently (slightly) higher in some lexemes than in others: *cobrar* > *ganar* > *doler* > *mentir*. This can be explained by the different contexts created for each lexicalization: in some contexts, the high degree (or exaggeration) was probably more salient than in others.

The results of this experiment will be compared to and complemented with corpus data and presented at the *Hispanistiktag 2025* conference in Hamburg.