PRINCIPIA **30**(1): 169–193 (2026) doi: 10.5007/1808-1711.2026.e107229
Published by NEL — Epistemology and Logic Research Group, Federal University of Santa Catarina (UFSC), Brazil.

# NEUROIMAGING AND THE SEARCH FOR PHILOSOPHICAL GROUNDINGS IN LANGUAGE PROCESSING STUDIES

LAURA NIÑO-BUITRAGO
*Universidad del Valle*, COLOMBIA
laura.nino.buitrago@correounivalle.edu.co
https://orcid.org/0000-0001-9165-500X

JONATHAN RESTREPO-RODAS
*Universidad del Valle*, COLOMBIA
restrepo.jonathan@correounivalle.edu.co
https://orcid.org/0000-0001-6543-1774

MERCEDES SUÁREZ DE LA TORRE
*Universidad Autónoma de Manizales*, COLOMBIA
mercedessuarez@autonoma.edu.co
https://orcid.org/0000-0002-1223-2146

DILLIS ELIANA JIMÉNEZ-GUTIÉRREZ
*Universidad del Quindío*, COLOMBIA
dejimenezg@uqvirtual.edu.co
https://orcid.org/0000-0003-4698-0107

**Abstract.** Neuroimaging-based language studies face the challenge of properly selecting linguistic units to be used as stimuli. While the selection of simple terms is relatively straightforward, abstract concepts and polysemic markers pose greater difficulties, potentially affecting the validity of the association between stimuli and the neurocognitive structures. This study addresses this problem by focusing on modality, specifically through the validation of polysemic statements using the verb *tocar* in Colombian Spanish. A set of modal logic proofs was developed based on existing theoretical definitions. These definitions were coupled with diverse senses of *tocar* as expressed in linguistic terms, formalized as axioms and then expressed as formal deductive proofs to support the categorization of statements. This process helps in maintaining the philosophical contribution to studies in the neurocognitive realm. Through an elicitation task and expert validation, a final set of 198 statements was consolidated, achieving fair agreement according to Fleiss' Kappa statistic. It is concluded that these statements contribute to enhancing the validity of neurocognitive lexical decision tasks involving polysemic markers and their neurocognitive correlates. This work provides a deductive and replicable methodological framework that can be applied in studies framed in empirical, analytic and experimental philosophy, neurophilosophy and Cognitive Siences.

**Keywords:** neurophilosophy • polysemic sentence processing • content validity • modality • modal logic

RECEIVED: 28/05/2025 ACCEPTED: 17/10/2025

# 1. Introduction

Neuroscience has become an interdisciplinary field in which phenomena of several disciplines converge. This is the case of philosophy, which has transitioned to the *naturalizing enterprise*, a movement grounded, among others, in the foundational work of Patricia Churchland. She pioneered the field of neurophilosophy by proposing that philosophical questions about the mind must be answered in light of empirical findings about the brain (Churchland 1986; 2002). Her contributions have placed neural data at the center of philosophical theorization, highlighting its role as a fundamental component in understanding concepts such as consciousness, intention, and moral reasoning. This naturalizing shift is sometimes enacted by assuming partial or total epistemological replacement, and other times by applying methodologies and techniques typical of the neurosciences (Estany 2022a, p.296-298; 2022b, p.9-10).

In this way, language studies that pursue correlations between structures and functions of the neural mechanisms can offer empirical approximations to classical philosophical concepts, say consciousness, morality, intentionality, to mention just a few. A crucial step in this kind of research is the appropriate selection of linguistic elements that serve as a bridge to match philosophical matters with neurocognitive processes, or how it is commonly accepted, the union of philosophical ideas with neurocognitive hypothesis that can be tested empirically (Northoff 2004, p.92), as is the case of moral judgment (Estany 2022a; Young & Saxe 2008; 2009).

Such linguistic elements should contemplate philosophical groundings that delineate the conceptual dimensions to be tested. This is to say that the mere linguistic adequation of number of syllables, kinds of verb, or syntactic-semantic properties (New et al. 2001, Barrio-Cantalejo et al. 2008) for neurocognitive stimuli is not sufficient, leaving chances for philosophical principles, procedures and techniques to play an important role. Although conceptual analysis has been traditionally assumed as the quintessential philosophical approach (Knobe & Nichols 2008, p.23), logic appears to offer a basis for correctly typifying the linguistic components that aim to contribute to neurophilosophy. In particular, studies that report data on neurocognitive processes that are vigorously framed in philosophy can be sound to transdisciplinary endeavors, aspiring to better fit in either of the scientific movements ascribed to neurophilosophy, say empirical philosophy and experimental philosophy (Prinz 2008), or even currents of the Cognitive Sciences that explore the mind's workings (Felisatti & Fischer 2024).

A typical research domain where logic can be beneficial as a technique is the lexical decision paradigm. Neuroimaging studies in lexical decision and statement comprehension rely on techniques of the cognitive neurosciences to ensure academic rigor and replicability. This entails characterizing stimulus length, response time, processing speed, and cognitive load (Turkeltuab et al. 2002; Rogalski 2016; Bornkessel

et al. 2016; Beres 2017). However, many of the studies in this current, while presenting consistent linguistic criteria as protocols for the corpus selection, still lack conceptual groundings that stem from philosophy, especially when dealing with concepts of the field. If not exactly cared, the exclusive linguistic rationale behind stimuli preparation can bias experimental task construction and their replicability (Marsden 2020, p.19-20). Hence, applying diverse types of logic can be fruitful both for philosophy (Suárez-Rivero 2024, p.3) and neurophilosophy (Northoff 2004, p.95), assuring greater levels of validation and reliability of the linguistic-conceptual material, procedures, and techniques used in neuroimaging research.

For this reason, it is crucial to ensure the proper construction, selection, and validation of linguistic statements, thus consolidating internal validity (Gillioz & Zufferey 2021, p.34). This process entails a lexical, syntactic, semantic, *and conceptual* schematization that must be consistent with the theoretical framework of each study. Content validation is a fundamental process in the development of instruments aimed at defining and measuring abstract constructs, as it guarantees the quality and relevance of the stimuli presented to the participants (Escobar-Pérez & Cuervo-Martínez 2008, p.27). In turn, this guarantees that the data collected accurately reflect the phenomena of interest (Maldonado-Suárez & Santoyo-Telles 2024, p.8).

The validation of linguistic elements requires evaluating core dimensions to ensure their adequacy and relevance in relation to the objectives of each study. The following criteria reflect such values: *relevance* refers to whether the selected stimuli are appropriate and pertinent; *clarity* indicates that they are understandable and free of ambiguities; and *coherence* assesses whether all elements maintain a logical and consistent structure in both content and format (Vermeiren, Vandendaele & Brysbaert, 2023). It is also crucial to analyze *representativeness* (Luck 2014, p.23; Beres 2017, p.253), which determines whether the selected stimuli adequately encompass the phenomenon under study. Additionally, it is necessary to guarantee *Operationalization,* as it evaluates that theoretical concepts have been successfully converted into observable and measurable stimuli. In addition, *target population suitability* ensures that the language and complexity of instruments are appropriate for the target audience. Together, these properties contribute to the robustness and validity of the instrument for research applications (Maldonado-Suárez & Santoyo-Telles 2024, p.10).

Building on this framework, this study develops and validates a set of statements that contain the verb *tocar* in its semantic value of will and intention, as derived from its prototypical form, and non-prototypical meanings, of judgment of obligation (Fernández 2014a, 116), while also validating alethic statements that do not contain the verb *tocar* (i.e., synthetic and analytical judgments). The statements will be later used to correlate volition and deontology in polysemic statements with neural activity. This process expands upon the study by Restrepo Rodas et al. (2022),

who analyzed and characterized several statements containing the verb *tocar* in the volitive, deontic and axiological dimensions. The proper validation of these statements is crucial to ensure that the neurocognitive conclusions to be drawn from the study are both accurate and valid. In Colombian Spanish, the verb *tocar* is highly polysemic and encompasses meanings that vary from physical concreteness (volitive, telic statements) to metaphorical extensions (deontic statements, both telic and atelic). These variations in meaning raise important questions about how the human brain processes different semantic uses of the same verb (Malaia et al. 2015, p.134).

From a neurocognitive perspective, the processing of *tocar* in its prototypical use (associated with volitive intention) would involve the activation of different brain areas related to physical action, including the primary and secondary somatosensory cortices, as well as the motor cortices (Feldman; Narayanan; Lakoff 2000; Pallier; Devauchelle; Dehaene 2010; Ghio & Tettamandi 2015; Pulvermuller & Fadiga 2016; Nozari & Thompson-Schill 2016; Bornkessel et al. 2016; Hauk 2016). However, for non-prototypical (deontic) uses, it is still unknown whether motor and sensory areas are also activated, and whether additional areas associated with abstract language processing and decision-making are involved (Yamada & Neville 2007; Malaia et al. 2015; Rogalski 2016).

In order to explore how humans process the different uses of the verb *tocar*, this study establishes a novel methodological protocol that can be applied in neurocognitive research involving electroencephalography (EEG) and event-related potentials (ERPs) during lexical decision tasks or reading comprehension. To this end, it is essential to validate the linguistic statements that make up these tasks, not only in linguistic terms, but also in conceptual (philosophical) dimensions, ensuring they accurately reflect the neurocognitive and linguistic nuances proposed in the study's hypotheses. By adopting forms of logic, like modal logic, this study embraces a transdisciplinary mechanism of philosophical naturalization (Northoff 2004, p.95-105), involving concepts like volition and deontology that are present in the polysemic meanings of the verb *tocar*.

## 2. Theoretical framework

Language is a fundamental starting point for the study of cognitive, linguistic, and social phenomena (Lupyan 2016, p.516). Within this context, the present study focuses on two polysemic uses of the verb *tocar*: 1) volition, expressed in statements that denote intentions such as "Mis amigos *tocan* la guitarra muy bien" (My friends *play* the guitar very well); "El chef *toca* los alimentos con las manos" (The chef *touches* the food with his hands), and 2) deontology, which conveys obligations and restrictions:

"Te *toca* pagar la cuenta esta vez" (You *have to* pay the bill this time); "Le *tocó* casarse en Junio" (He *had to* get married in June).

This kind of polysemy indicates, on the one hand, the possibility of carrying out a semantic revision of the verb *tocar* and, on the other, its potential usefulness as an experimental resource in neuroimaging studies. The modal properties of *tocar* have been studied in recent years, both in factual descriptions and in prescriptions (Fernández 2014; Restrepo Rodas et al. 2022). Unlike alethic modality, which refers to factual information, volitive and deontic modalities relate to the transition of actions or framing prospective states. They are often expressed by commissive acts or prescriptions, not by commands and suggestions (Restrepo Rodas et al. 2022, p.91), in the schematized force dynamics of agent-patient interactions (Fernández 2014, p.99). In this way, volitive and deontic modalities convey both the ongoing immediate reality and the projected ones, reflecting diversity in the semantic temporality that is frequently used in everyday discourse (Gosselin 2005, p.80).

Gosselin's (2010) Modular Theory of Modalities proposes several modal categories. The volitive modality refers to an agent's will or intention over something. The deontic modality involves obligations imposed on the agent or the absence of intention. The alethic modality, which includes analytic and synthetic judgments, refers to propositions that are unintentional, value-neutral, and non-subjective in nature. In Latin America, modalities have been studied in the field of interculturality (Loaiza et al. 2022) and in cognitive linguistics (Restrepo Rodas et al. 2022), although there is still no operational formalization of these categories applied to particular cases. The large polysemy of the verb *tocar* (Fernández 2014, p.116) offers a fertile niche for the formal exploration of volitive, deontic and alethic dimensions.

This study proposes a preliminary characterization of the axioms that account for the linguistic diversity of this verb across modalities. These axioms are expected to support validation of volitive, deontic and alethic statements, making them suitable for use in experimental studies involving cortical brain activity recording. The formalizations presented in Table 1 were derived from theoretical definitions of modal categories (Gosselin 2010) and empirical findings of the verb *tocar* (Fernández 2014; Restrepo Rodas et al. 2022). This allowed for the establishment of axioms and the emergence of deductive proofs inspired by modal logic, assuming key principles of the naturalizing project introduced by Quine (1969). These logical proofs aim to provide formal criteria for the coding and categorization of the statements used in the validation process. To the best of our knowledge, this deductive formal approach has not been previously applied in the content validation of linguistic stimuli for language processing in neurocognitive research.

| Number | Axiom |
|---|---|
| **1** | The verb *tocar* expresses the volitive modality in any verb tense. |
| **2** | The verb *tocar* expresses the volitive modality in the indicative mood. |
| **3** | The verb *tocar* expresses the volitive modality in conjunction with verbal periphrases located before the verb *tocar*. |
| **4** | The verb *tocar* does not express the volitive modality in the imperative, subjunctive and conditional moods. |
| **5** | The verb *tocar* does not express the volitive modality in the passive voice. |
| **6** | The verb *tocar* does not express the volitive modality in conjunction with verbal periphrases located after it. |
| **7** | The verb *tocar* expresses the deontic modality in any verb tense. |
| **8** | The verb *tocar* expresses the deontic modality in the pronominal or impersonal forms of the indicative mood. |
| **9** | The periphrasis *toca que* expresses the deontic modality in the imperative and subjunctive moods. |
| **10** | The verb *tocar* can express the deontic modality in conjunction with verbal periphrases before or after the verb *tocar*. |
| **11** | The verb *tocar* cannot express the deontic modality in the passive voice. |
| **12** | The alethic modality of synthetic and analytic judgments cannot be expressed with the verb *tocar*. |
| **13** | The verb *tocar* and its periphrases express the outcome of the intention or will in the volitive modality. |
| **14** | The verb *tocar* and its periphrases do not express the outcome of intention or will in the alethic modality. |

Table 1: Axioms for the delimitation of volitive, deontic and alethic meanings 'tocar'.

## 2.1. Formal linguistic proofs

With the axioms defined in Table 1, it becomes necessary to provide a formal proof for each modal category as applied to the uses of *tocar*. This process warrants a clear-cut characterization of the statements where *tocar* can appear in its diverse conceptual dimensions.

| **Modality** | **Formal linguistic/modal proof** |
|---|---|
| **Volitive** | 1. If in a statement the verb *tocar* appears in volitive modality (intention or will) and the insertion or removal of negation leads to a transformation into either the deontic modality (obligation) or the alethic modality (factuality, analytic or synthetic judgment), then the original statement is classified as volitive. Let the expression be:<br>$\forall \mathbf{p}(\Diamond \mathbf{p} \wedge \boldsymbol{Vol}(\boldsymbol{p})((\neg \mathbf{p} \leftrightarrow \Box \mathbf{p}) \vee (\boldsymbol{p} \leftrightarrow \Box \mathbf{p})) \rightarrow \Diamond \mathbf{p})$ |
| **Deontic** | 1. If a statement contains the verb *tocar* in a deontic modality (obligation) and appears in any mood except the impersonal, and if *tocar* can be replaced by a periphrasis involving *tener que* ("to have to") while maintaining the deontic reading, then the statement is classified as deontic. Let the expression be:<br>$\forall \mathrm{p}(\Box \mathrm{p} \wedge T(p) \wedge \Box \mathrm{p}' \rightarrow \Box \mathrm{p})$<br>2. If a statement contains the verb *tocar* in the deontic modality and it appears in the impersonal mood, and if *tocar* can be replaced by a modal construction involving "must" while deriving in the deontic modality, then the statement expresses a deontic modality. Let the expression be:<br>$\forall \mathrm{p}(\Box_i \mathrm{p} \wedge H(p) \wedge \Box \mathrm{p}' \rightarrow \Box \mathrm{p})$<br>This enables the integration of both expressions, let the expression be:<br>$\forall \mathbf{p}((\Box \mathbf{p}\boldsymbol{p} \wedge \boldsymbol{T}(\boldsymbol{p}) \wedge \Box \mathbf{p}') \vee (\Box \boldsymbol{ip} \wedge \boldsymbol{H}(\boldsymbol{p}) \wedge \Box \mathbf{p}') \rightarrow \Box \mathbf{p})$ |
| **Alethic** | 1. If a statement describes a synthetic or analytic truth, then it is an alethic modality. Let the expression be:<br>$\forall \mathbf{p}(\boldsymbol{Syn}(\boldsymbol{p}) \vee \boldsymbol{An}(\boldsymbol{p}) \rightarrow \Box \mathbf{ap})$<br>2. If a statement contains the verb *tocar* expressing intention, will, or obligation (or a similar lexeme), then the modality is not alethic. Let the expression be:<br>$\forall \mathrm{p}((\Diamond \mathrm{p} \vee \Box \mathrm{p}) \rightarrow \neg \Box \mathrm{ap})$<br>3. If a statement contains any state or action, and the insertion or removal of a negation results in a deontic, volitive, appreciative, axiological, or epistemic modality, then the statement does not express an alethic modality. Let the expression be:<br>$\forall \mathrm{p}(N(p) \wedge (\Box \mathrm{p} \vee \Diamond \mathrm{p} \vee Ap(p) \vee Ax(p) \vee Ep(p)) \rightarrow \neg \Box \mathrm{ap})$<br>That is, if neither condition 2 nor 3 is met, then the modality is alethic. Since the alethic modality does not convey deontic, volitive, appreciative, axiological, or epistemic values, if any of these arises only through the insertion or removal of a negation marker, then the statement expresses an alethic modality. Let the expression be:<br>$\forall \mathbf{p}(\neg(\Diamond \mathbf{p} \vee \Box \mathbf{p}) \wedge \neg(\boldsymbol{N}(\boldsymbol{p}) \wedge (\Box \mathbf{p} \vee \Diamond \mathbf{p} \vee \boldsymbol{Ap}(\mathbf{p}) \vee \boldsymbol{Ax}(\boldsymbol{p}) \vee \boldsymbol{Ep}(\boldsymbol{p}))) \rightarrow \Box \mathbf{ap})$ |

Table 2: Formal proofs for the volitive, deontic, and alethic modalities derived from the verb 'tocar'. **Note:** the modal expressions in bold are categorical and best represent the modal category at hand.

The formalization of modal categories through axioms and formal proofs facilitates the coding and categorization of statements in cases such as the polysemic verb *tocar*.

## 3. Methodology

The present study is framed in quantitative guidelines (Loewen & Godfroid 2020); it sought to ensure the validity of the content of statements with volitive, deontic, and alethic uses of *tocar* through the judgment of experts. Likewise, the study falls within the scope of instrument validation (Soriano 2014), which involves two phases: an exploratory phase aimed at identifying the real use of linguistic markers, and a validation phase, in which these markers are subjected to the scrutiny of expert judges (Supo 2013, p.5). The first is aimed to elicit the verb *tocar* in its various meanings (Gillioz & Zufferey 2021, p.66-8) while the second intends to provide a framework of validity that supports the use of the statements in future studies (Maldonado-Suárez & Santoyo-Telles 2024, p.15).

### 3.1. Phases of the study

The validation process took place in two main moments. First, an exploratory phase was conducted to empirically identify the uses of the verb *tocar*. During this phase, an elicitation task was carried out, after which the resulting statements were coded and categorized (Saldaña 2021) according to their prototypical and non-prototypical meanings and based on the deductive approach above-mentioned, resulting in a preliminary list of categorized statements.

Second, the statements were evaluated by a panel of expert judges. This entailed analyzing the conceptual dimensions of the statements, selecting the expert judges, and adapting and providing the evaluation questionnaire (McAlister et al. 2017). As a result, expert evaluations were collected, and the final list of statements was defined. The first phase led to the identification of the statements based on participants' actual language use, while the second enabled the quantitative analysis of expert assessments.

### 3.2. Participants

Each phase of the study involved a specific population. For the exploratory phase, 20 language teachers from a public institution in Dosquebradas, Risaralda, participated. In the validation phase, a three expert judges panel was formed, meeting the minimum number suggested by Maldonado-Suárez and Santoyo-Telles (2024, p.9). The panel was composed of experts with master's level training in interdisciplinary

fields: psychology, applied linguistics, and cognitive linguistics. Informed consent was obtained from both participant groups, outlining the study's objectives and the procedures to be carried out. The study, identified under ID 1195, received ethical approval from the University of Quindío, as stated in Act 031 dated June 1st, 2023.

## 3.3. Information collection instruments

### 3.3.1. Elicitation task

Each phase applied instruments of different scopes. In the exploratory phase, a rubric was used to carry out the elicitation task involving the use of the verb *tocar* within the Colombian population. This is one of the most widely used techniques in linguistics to gather linguistic corpora (Baker 2018) as it provides evidence of the mental representation of syntactic structures and their complexity (Gillioz & Zufferey 2021, p.61-2). It can be applied in one or more languages, depending on the characteristics of the participants (Serratrice et al. 2015, p.991). Additionally, it allows for empirical observation of structures that shape language at the level of concepts (Gillioz & Zufferey 2021, p.40-1) providing information about individual's cultural and linguistic contexts, as well as their world knowledge and beliefs, all of which shape the construction of concepts and the variety of terms they use to express them (Bowen 2019).

The elicitation task was divided into several stages, each with a 10-minute time limit (Creswell & Poth 2018, p.204). Participants were first asked to write ten sentences using the verb *tocar* in any of its different meanings. Next, they were asked to group the sentences by similarity of meaning. Then they were told to translate their own sentences to corroborate the intended meanings. Finally, they were invited to exchange the rubrics with another participant: while one participant read the statements, the other represented them through gestures, sounds, etc. As can be seen, the task of elicitation begins individually and subjectively, and ends collectively and intersubjectively, since only in this way can it be verified that the resulting statements correspond to everyday use and conventionality (Gillioz & Zufferey 2021, p.66).

### 3.3.2. Validation form by expert judges

The validation form by expert judges proposed by Escobar-Pérez and Cuervo-Martínez (2008, p.35-6) was used. It consists of three sections: a summary of the study, a description of the instrument to be validated, and an assessment rubric. For this study, the summary included the research problem, objectives, methodological approach, study hypothesis, data collection instruments and the analysis strategy. The description section presented the theoretical foundations that support the modal dimensions expressed in the statements, the procedures for defining and selecting statements,

and the preliminary list of statements to be validated. The evaluation grid includes quantitative rating criteria, accompanied by a space for qualitative comments.

### 3.3.3. Description of the statements

The instrument is based on three main components: definitions, axioms, and formal proofs. The strategy used is the articulation of the categories of linguistic modalities with modal logic in favor of deductive reasoning. Definitions were based on Gosselin's (2010) Modular Theory of Modalities and adapted to the uses of *tocar* in Colombian Spanish (Restrepo Rodas et al. 2022). Axioms are propositions accepted as true that articulate the minimal linguistic principles (e.g., mood, voice, tense) in which each modal category can be expressed. Formal proofs are deductive procedures that articulate the axioms according to the conceptual definitions. This framework served as the basis for drafting the provisional list of statements, which was later submitted for expert validation.

### 3.3.4. Assessment rubric grid

The grid was structured horizontally into four sections: criterion, score, justification, and suggestions. Vertically, four criteria were established for the evaluation. First, we reviewed the rationale behind each statement, assessing the relevance of conceptual definitions, axioms, and formal proofs, in other words, whether the statements clearly reflected the appropriate use of the verb *tocar* in the given contexts. Second, the operationalization of the statements was assessed to verify whether the linguistic mechanisms used for the formal proofs were relevant and effectively represented the intended modal categories (volitive, deontic, alethic).

The third criterion corresponded to the representativeness of the stimulus, ensuring that the sentences adhered to the grammatical and semantic norms required to express the designated modal category. This confirmed whether each sentence was representative of a prototypical or non-prototypical use of the verb without violating linguistic principles. Finally, the adequacy of the statements to the experimental protocol was evaluated, focusing on whether the statements allowed for the collection of clear and useful data for the neurocognitive analysis of each category (the subsequent study). A four-point Likert scale was used, with 1 meaning not relevant and 4 meaning very relevant. Additionally, a dedicated space was also included to allow judges to provide qualitative feedback on each item (Escobar-Pérez & Cuervo-Martínez 2008, p.36). Total scores were obtained by summing the ratings across the four criteria, with statements scoring above 11 (out of a possible 16) considered valid.

## 3.4. Procedures

### 3.4.1. Phase 1

**Stage 1: Sentence coding process**

Based on the elicitation task, the qualitative analysis was carried out through the category construction strategy. In the first instance, open coding (Saldaña 2021) was applied to verbal periphrases containing the verb *tocar* in its traditional (volitive), non-traditional (deontic) and other (alethic) meanings. The statements were compiled in a Notion-based database created by the research team. Each entry was labeled and assigned a participant-specific code, allowing for the identification of linguistic patterns (Creswell & Poth 2018) and facilitating the selection of provisional statements for the experiments (Gillioz & Zufferey 2021). This analysis ensured that the chosen statements reflected linguistic variations relevant to the context of the study, including voice, mood, and tense. In addition, a space for notes was added to the categorization database to include relevant observations, such as potential ambiguities, problems of interpretation, or justification for inclusion or exclusion from the provisional list.

**Stage 2: Statement selection process**

An operational table was used to select the statements, including the statement number, the corresponding text, the formal proof, and lexical density (Proverbio 2023). Completion of each section indicates whether the statement is relevant and is selected, thus forming the final list of statements to be validated. As part of this process, formal proofs were carried out through two linguistic procedures from Hispanic grammaticology:

(1) Substitution by synonyms in verbal periphrases containing the verb *tocar* (Martí et al. 2016), and

(2) Insertion or removal of the negation marker in the main verbal copula (secondary in some cases) or the inversion of polarity in nouns, adjectives, or adverbs (Zapico & Vivas 2015).

These procedures allowed the identification of change in modal category expressed in the verb *tocar*. For example, if a statement that expresses a volitive modality (“Mi hija *toca* el piano y la guitarra” “My daughter *plays* the piano and the guitar”) is negated, then it expresses an alethic modality (“Mi hija *no toca* el piano y la guitarra” “My daughter *does not play* the piano and the guitar”) or a deontic modality (“*No toque* el piano y la guitarra” “*Do not* play the piano and the guitar”). The volitive intention in the first statement becomes a factual assertion devoid of intentionality in the second and a command or obligation in the third, hence the importance of

carrying out formal proofs of specific modal categorizations of each statement. Additionally, sentences presenting medium lexical density (Hagoort & Indefrey 2014; Rogalski 2016), were considered optimal and retained in the final list of statements to be evaluated (Table 3).

The following table presents a selection of statements originally elicited in Colombian Spanish and translated into English for the purpose of illustrating modal categorization. The English versions aim to accurately reflect the polysemy of *tocar*, preserving the intended modal meaning (volitive, deontic, alethic), speech register, and the contextual nuance of each statement. These choices ensure that the formal proofs remain consistent and valid across both linguistic versions.

| **Nº** | **Volitive Statements** | **Formal Proof** | **Apply** | | **Lexical** | **Selection** | |
|---|---|---|---|---|---|---|---|
| | | | Yes | No | **Density** | Yes | No |
| 13 | He *plays* very well (Él *toca* muy bien) | He **doesn't** *play* very well – alethic (Él **no** *toca* muy bien) | Yes | | 4 | Yes | |
| 14 | Someone *touched* my hair on the bus (Alguien *tocó* mi cabello en el bus) | **No one** *touched* my hair on the bus – alethic (**Nadie** *tocó* mi cabello en el bus) | Yes | | 7 | Yes | |
| 15 | You *touched* the still fresh paint (*Tocaste* la pintura aún fresca) | **You didn't** touch the still fresh paint – alethic | Yes | | 5 | Yes | |
| **Nº** | **Deontic Statements** | **Formal Proof** | **Apply** | | **Lexical** | **Selection** | |
| | | | Yes | No | **Density** | Yes | No |
| 19 | I ended up having to wait in a very long line (*Me tocó* hacer una fila muy larga) | I ***had to*** wait in a very long line (***Tuve que*** hacer una fila muy larga) | Yes | | 7 | Yes | |
| 20 | My dad will *have no choice but* to work tonight (A mi papá *le tocará* trabajar esta noche) | My dad will ***have to*** work tonight (Mi papá ***tendrá que*** trabajar esta noche) | Yes | | 8 | Yes | |
| 21 | *You're up for* buying the cake (*Toca que* ustedes compren la torta) | You ***have to*** buy the cake (Ustedes ***tienen que*** comprar la torta) | Yes | | 6 | Yes | |
| **Nº** | **Alethic Statements** | **Formal Proof** | **Apply** | | **Lexical** | **Selection** | |
| | | | Yes | No | **Density** | Yes | No |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 52 | Speaking is a human act (Hablar es un acto humano) | Speaking **is not** a human act – alethic (Hablar **no** es un acto humano) | Yes | | 5 | Yes | |
| 53 | The white guitar is bigger (La guitarra blanca es más grande) | The white guitar is **not** bigger – alethic (La guitarra blanca **no** es más grande) | Yes | | 6 | Yes | |
| 54 | The dining table is round (La mesa de comedor es circular) | The dining table is **not** round – alethic (La mesa de comedor **no** es circular) | Yes | | 6 | Yes | |

Table 3: Example of the selection process

### 3.4.2. Phase 2

**Stage 1: Selection of expert judges**

To define the profile of expert judges, criteria were established to guarantee both heterogeneity among judges and a high degree of expertise in the evaluation of the statements. It was stipulated that the panel would consist of at least three judges (Maldonado-Suárez & Santoyo-Telles 2024, p.9) each with a distinguished background in one of three fields: cognitive linguistics, applied linguistics, or cognitive sciences. Efforts were also made to promote diversity within the panel, by including at least one international expert, gender variety (men and women), different levels of postgraduate training, and a range of professional trajectories in terms of university teaching experience (Escobar-Pérez & Cuervo-Martínez 2008, p.27). To avoid conflicts of interest, none of the authors of this study participated in the validation of the instruments.

**Stage 2: Statistical analysis of expert judgement**

In lexical decision studies that use Event-Related Potentials (ERPs), it is recommended that the number of statements of the same nature range between 40 and 60, as this interval ensures reliable statistical power (Luck 2014, p.23). For this reason, it is more efficient, in terms of time and effort, to evaluate the criteria of sufficiency, coherence, relevance and clarity that shape the statements, rather than assessing all the items individually ($n$=198). This approach reduces the cognitive load and fatigue associated with the validation task (Maldonado-Suárez & Santoyo-Telles 2024; Beres

2017, p.253), increasing confidence in the rate of agreement among judges on the evaluated criteria.

Inter-rater agreement was measured using Fleiss' Kappa statistic (Fleiss 2017), which accounts for the proportion of agreement attributable to chance (Gisev; Bell; Chen 2013, p.335). However, in order to prevent the inclusion of incorrect statements, i.e., statements identified as inadequate or inconsistent, the expert judges were asked to record relevant observations in the evaluation form. This allowed for the replacement of problematic statements with similar ones based on previously accepted criteria. The validation analysis followed three key procedures: 1) the mean rating across the three judges, (2) the degree of inter-rater agreement, and (3) the qualitative analysis of comments and suggestions related to criteria that received lower scores or statements requiring modification. Figure 1 illustrates the flow of the statement selection and validation process.

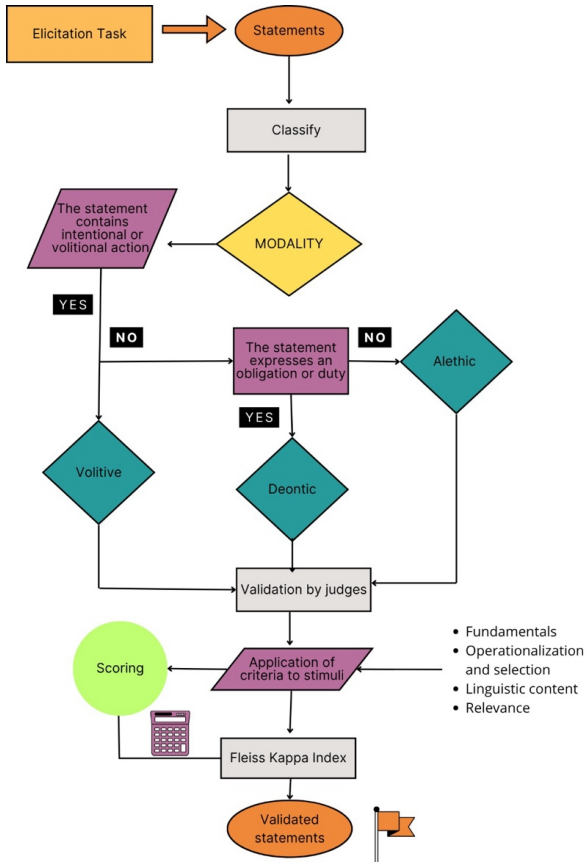


Figure 1: Flow diagram of the statement selection and validation process

# 4. Results

## 4.1. Phase 1, Stage 1: Elicitation Task

The elicitation task yielded the following results. Of the 20 participants, only 12 filled out the form correctly. The initial instruction was to provide 10 sentences with different uses of the verb *tocar,* which projected a potential corpus of 120 statements. From this, a total of 102 valid statements were obtained. Of the statements collected, 60 were categorized as volitive (58.82%), 23 as deontic (22.55%), and 19 as other uses (18.63%). After analyzing the lexical units to determine statement density and acceptability, the sample was refined to 26 volitive statements (25.48%), 13 deontic statements (12.75%), and 19 figurative or non-classifiable expressions (18.63%), with no alethic statements identified (0%). In broad terms, the total number of volitive and deontic statements retained for validation was 39 (38.23%). Table 4 presents a selection of statements that passed the validation filters (e.g., 50, 23, 59, 64, 17) and others that were excluded (e.g., 7, 12, 69, 86, 100).

| ***P*** | ***Statement*** | ***#*** | ***Category*** | ***Decision*** | ***Explanation*** |
|---|---|---|---|---|---|
| 3 | He *played* a wonderful song in the evening (Él *tocó* una maravillosa canción en la velada) | 50 | Volition | YES | Change from volitive to alethic modality; adequate lexical density |
| 3 | He *played* a wonderful song in the evening (Él *tocó* una maravillosa canción en la velada) | 50 | Volition | YES | Change from volitive to alethic modality; adequate lexical density |
| 4 | The boy *rang* the doorbell (El niño *tocó* el timbre) | 23 | Volition | YES | Change from volitive to alethic modality; adequate lexical density |
| 5 | I *have to* leave early for work (*Me toca* salir temprano a trabajar) | 59 | Obligation | YES | The deontic modality remains with the insertion of synonyms (I have to); adequate lexical density |
| 6 | *You are required* to present today (*Te toca* exponer hoy) | 64 | Obligation | YES | The deontic modality remains with the insertion of synonyms (I have to); adequate lexical density |
| 10 | She *had to* dance all night (*Le tocó* bailar toda la noche) | 17 | Obligation | YES | The deontic modality remains with the insertion of synonyms (I have to); adequate lexical density |

| | | | | | |
|---|---|---|---|---|---|
| 1 | They always *bring up* the same question about leaving 10 minutes early (Ellos siempre *tocan* el mismo tema de salir 10 minutos antes) | 7 | Diverse cases | NO | Figurative expression; Greater lexical density |
| 2 | I'm going to a musical *gig* (Voy a un *toque* musical) | 12 | Diverse cases | NO | Figurative expression |
| 9 | We must let our hearts *be touched* by such a difficult situation (Debemos *tocarnos* el corazón con esa situación tan difícil) | 69 | Diverse cases | NO | Figurative expression; Greater lexical density |
| 11 | He got very *touchy* about what I said (*Se tocó* mucho por lo que dije) | 86 | Diverse cases | NO | Figurative expression |
| 12 | They *bring up* the subject and it makes me angry (Ellos *tocan* el tema y me da rabia) | 100 | Diverse cases | NO | Figurative expression |

Table 4: Some selected and unselected statements. **Note:** P stands for participant.

## 4.2. Phase 2

### Stage 1: Selection of the expert judges

The panel of experts included two judges holding doctoral degrees and one with a master's degree, whose professional experience range from 3 to 15 years. In terms of geographical location, two are located in Colombia and one in the United States of America, 66.6% are women dedicated to the study of language, particularly research on modal markers within discourse analysis in Colombian Spanish. The third judge, an expert in the field of Cognitive Sciences, has extensive experience in the design, validation and implementation of instruments for measuring cognitive processes. The diversity of the panel added rigor to the process, by integrating perspectives that enhanced the delimitation of the dimensions expressed in the statements, their syntactic and semantic characteristics, as well as their psychometric properties (Maldonado-Suárez & Santoyo-Telles 2024).

| Judge | Gender | Context | Institutional affiliation | Area of expertise | Academic level | Experience |
|---|---|---|---|---|---|---|
| **1** | Female | National | University of Quindío | Applied linguistics | PhD | 15 years h-index 5 |
| **2** | Male | National | San Buenaventura University | Cognitive Sciences | PhD candidate | 8 years h-index 4 |
| **3** | Female | International | Florida Atlantic University | General Linguistics | Master's Degree | 3 years h-index 1 |

Table 5: Characterization of the panel of experts

### Stage 2: Analysis of expert judgement

The analysis yielded three levels of information. First, a general average score of 15.33 was obtained, corresponding to 95.83% relevance. The dispersion of the evaluations is low (*SD*: 0.47) and reflects high consistency among the judges' evaluations. By reviewing each criterion separately, a perfect score was obtained for Criteria 1 (Rationale of the stimuli) and 3 (Linguistic content of the stimuli) and a near-perfect score for Criteria 2 (Operationalization and selection of stimuli) and 4 (Relevance of the stimuli according to the experimental protocol).

Second, regarding inter-rater agreement, Fleiss' Kappa coefficient was calculated at $k = 0.402$ indicating a fair agreement, on a scale from <0 (deficient agreement) to 1 (almost perfect agreement). This result is considered significant, particularly given the lack of similar instruments reported in the literature for validating this type of statements. Table 6 summarizes the values obtained.

| Judge/K | Criterion 1 | Criterion 2 | Criterion 3 | Criterion 4 | Overall average | SD |
|---|---|---|---|---|---|---|
| **Judge 1** | 4 | 4 | 4 | 3 | **15.33, (95.83%)** | **0.47** |
| **Judge 2** | 4 | 3 | 4 | 4 | | |
| **Judge 3** | 4 | 4 | 4 | 4 | | |
| ***K*** | 1 | 0.66 | 1 | 0.66 | | |
| ***K (total)*** | **0.5714** | | | | | |

Table 6: Descriptive statistics of the evaluation

Finally, a qualitative analysis matrix was carried out to compile the arguments that complemented the quantitative evaluations. The ideal scores highlight the linguistic properties (both syntactic and semantic) of the statements, as well as the psychometric consistency in the overall process in general and the individual statements.

On the other hand, criteria 2 (Operationalization and selection of stimuli) and 4 (Relevance of stimuli according to the experimental protocol) received suggestions for improvement, which were addressed through concrete actions to optimize the proposal (Table 7).

| **Criterion** | **Judge 1** | **Judge 2** | **Judge 3** | **Action** |
|---|---|---|---|---|
| **1** | "The stimuli are completely aligned with the theoretical references presented in the annex." | "The definitions provided are clear and well-founded. It shows a solid basis for the categorization of the linguistic stimuli." | "The theoretical foundations contribute to unraveling the brain mechanisms underlying language processing." | Not applicable |
| **2** | "The operationalization process is well presented, and the selection of stimuli is consistent with the theoretical framework" | "The identification and selection of the statements are meticulous, and the classification and verification of criteria are clearly delineated. And, without affecting the quality of the exercise in this regard, greater detail in psychometric reasoning would have allowed for a better understanding of this process." *** | "The criteria of demonstrability and lexical density are relevant for selecting statements, ensuring their verifiability and representativeness in linguistic complexity. Table 1 facilitates the implementation of these criteria in a consistent manner." | It is worth noting that the researchers assumed the understanding of the process without giving explicit examples as it was considered redundant, it can be considered as a weakness in the socialization of information, without compromising the integrity of the validation process. |

| | | | | |
|---|---|---|---|---|
| **3** | "The linguistic content of the stimuli is adequate and relevant. It respects the grammatical and syntactic rules of Spanish." | "The linguistic content of the stimuli is well developed, broadly covering the uses of the verb *tocar*. The statements are varied and appropriately categorized." | "The linguistic content of the stimuli clearly demonstrates the categories to be studied and helps prevent rote memorization." | Not applicable |
| **4** | "Some statements within the 'Prototypical statements of will' of the Volitive category do not clearly express will. Such statements are highlighted: 42, 48, 54, 55, 57." *** | "The stimuli are pertinent and aligned with the objectives of the study. In addition, this content review process by linguistics experts will ensure its relevance and accuracy." | "The stimuli are relevant to the extent that they ensure an accurate analysis of neurocognitive activity for each of the categories to be studied." | Following Gosselin (2010), in statements 42 and 48, 55 and 57, meta rule 69 prevails, which establishes that every appreciative modality entails a volitive one. As for statement 54, it is not expressed in the passive voice as the judge points out. |

Table 7: Comments made by the expert judges

## 5. Discussion

The validation process carried out in this study highlights the importance of ensuring the alignment between the theoretical framework and linguistic stimuli used in experimental studies involving neurocognitive measurements (Proverbio 2023). The content validation of volitive, deontic and alethic statements by expert judges, based on the formalization of axioms and formal proofs proposed *a priori*, led to two key stages: the exploratory phase and the validation phase. Regarding the exploratory phase, the use of elicitation tasks that prompt statements directly from language users proved to be highly valuable, as suggested by (Gillioz & Zufferey 2021, p. 66). This procedure ensures that the statements genuinely represent the semantic values of the verb *tocar* as reported in the literature (Restrepo Rodas et al. 2022) and formalized in this study.

The *in vivo* performance of the elicitation task highlights the relevance of the systematic processes carried out (coding, categorization, and selection) to compile

the final linguistic corpus (Creswell & Poth 2018; McAlister et al. 2017; Saldaña, 2021). These procedures help prevent linguistic stimuli from being exclusively derived from theoretical/intuitive assumptions or selected in an *ad hoc* fashion (Gibbs 2007). For instance, although the elicitation task initially yielded 102 statements, only 39 were ultimately selected. Through a rigorous analysis process, statements that were deemed as not theoretically and methodologically relevant to the study were discarded. This approach enabled the consolidation of a robust list, consistent with previous findings reported in the literature (Proverbio 2023; Serratrice; Hervé; Corley 2015). Specifically, it drew upon previously validated syntactic structures and incorporated commonly used lexical items, all while preserving the intended semantic distinctions of the verb *tocar* (Fernández, 2014; Restrepo Rodas et al. 2022).

For neuroimaging studies, it is essential to consider the lexical density of stimuli (statements), as this feature complements theoretical coherence (Almeida & Manouilidou 2015; Proverbio 2023). For this reason, stimuli must have comparable lexical and syntactic characteristics, an aspect that is fostered through naturalistic elicitation tasks and the systematic refinement of information, grounded in clear methodological principles and procedures (Gillioz & Zufferey 2021). In this case, a variety of syntactic dimensions were considered, all using frequently occuring words of medium lexical density (Hagoort & Indefrey 2014; Rogalski 2016). This feature offers stability to obtain data through neuroimaging techniques, by reducing the probability of recording irrelevant information (noise) during neurocognitive measurement and enables more reliable statistical filtering through averages weighting during data processing (Luck 2014; Beres 2017).

The transition to the expert validation phase was significantly strengthened by the solid foundation established during the exploratory phase (Supo 2013). Notably, the panel of experts was diverse in terms of areas of expertise, fields of study, academic productivity, and gender, which expanded the scope of the evaluation beyond a strictly linguistic perspective (Soriano 2014; Escobar-Pérez & Cuervo-Martínez 2008). For instance, several characteristics of the psychometric properties of the statements were identified and assessed, demonstrating a strong interrelation between the stimuli under validation and the summary sheet of the study, thus providing theoretical and methodological coherence.

As shown in the results section, the quantitative values yielded by the assessment of the panel of experts are both sufficient and consistent, an essential step toward achieving greater precision in language processing studies (Loewen & Godfroid 2020). On the one hand, the descriptive statistics support the validity of the statements, with an average score of 15.33, with a low dispersion (0.47) which indicates a high consistency of the perceptions of the panel, indicating a high level of consistency in the panel's assessments. On the other hand, the obtained agreement coefficient ($k=0.402$) represents a relevant outcome, as it reflects not only a level

of inter-rater agreement but also eliminates the possibility that this agreement is the product of chance. Although classified as fair, this level of agreement is acceptable for the validation of a construct that, to date, has no precedent in the field. Taken together, the quantitative results and qualitative feedback support the conclusion that the content validity of the statements is positively affirmed, fulfilling the necessary criteria for their use in future experimental studies, as suggested by Larson-Hall and Plonsky (2015) and Marsden (2020).

## 6. Conclusions

The study provides theoretical novelty through the formalization of the modal categories expressed by the verb *tocar*. It embodies the transdisciplinary ethos of neurophilosophy (Churchland, 1986) by proposing polysemous linguistic expressions as entry points to empirically grounded philosophical theorization. Moreover, as noted by Churchland (2013) language is not merely a product of conscious deliberation, but a system largely governed by unconscious neural processes. Therefore, by formalizing and validating polysemous linguistic expressions through deductive logic, this work aims to reveal how conscious judgments (e.g., obligations or intentions) are rooted in neural processes that remain largely hidden yet essential to the structure of thought. This approach contributes to conceptual clarity from a deductive perspective to the understanding of polysemous statements containing *tocar*, that operate across the following dimensions:

- Volitive: $\forall p(\Diamond p \wedge Vol(p((\neg p \leftrightarrow \Box p) \vee (p \leftrightarrow \Box p)) \rightarrow \Diamond p))$,
- Deontic: $\forall p((\Box pp \wedge T(p \wedge \Box p') \vee (\Box ip \wedge H(p \wedge \Box p') \rightarrow \Box p))$,
- Alethic: $\forall p(Syn(p \vee An(p \rightarrow \Box ap)$ / $\forall p(\neg(\Diamond p \vee \Box p) \wedge \neg(N(p \wedge (\Box p \vee \Diamond p \vee Ap(p \vee Ax(p \vee Ep(p)) \rightarrow \Box ap))$.

It should be noted that, among these categories, only the alethic-deontic modalities have been previously studied in the same linguistic marker (Wu 2024; Fernández 2014; Restrepo Rodas et al. 2022) and in classical term-modal logics in models of obligation, evidence and abilities (Frijters & Van De Putte 2021), not yet in realistic and natural elicited cases such as the verb *tocar* as reported in this study. This work also contributes methodological novelty, since it implements a multi-phase-stage process that led to the creation and validation of volitive, deontic, and alethic statements. In this process, the elicitation task stands out as a crucial step that allowed generating, analyzing and selecting the statements based upon actual use, therefore providing ecological validity to the validated statements (Gillioz & Zufferey 2021).

Validation of statements has proven to be a key step in ensuring quality and accuracy for recording cortical activity when processing polysemic statements with *tocar*.

The process developed here offers a replicable model for content validation in lexical decision and reading comprehension tasks that deal with philosophical concepts, pointing out at the importance of rigorous statistical analysis to support the integrity of experimental stimuli (Larson-Hall & Plonsky 2015; Marsden 2020). This work addresses the need for enhanced methodological rigor in the design of linguistic (and philosophical) stimuli (Loewen & Godfroid 2020; Gillioz & Zufferey 2021), including the validation of the statements and the transparency of the methodology employed. For future studies, it is recommended that the reliability of these statements be explored, particularly through their transformation into indicators of measurement of volition, deontology and factual (alethic) reality in human cognition as expressed in polysemic markers. In addition, researchers are encouraged to extend the deductive approach in modal logic here used and the validation process to other modal dimensions proposed in Gosselin's (2010) Modular Theory of Modalities.

## References

Almeida, R.; Manouilidou, C. 2015. *Cognitive Science Perspectives on Verb Representation and Processing*. London: Springer.

Baker, P. 2018. *Sociolinguistics and Corpus Linguistics*. Cambridge: Cambridge University Press.

Barrio-Cantalejo, I.M.; Simón-Lorda, P.; Melguizo, M.; Escalona, I.; Marijuán, M.I.; Hernando, P. 2008. Validation of the INFLESZ scale to evaluate readability of texts aimed at the patient. *Anales del Sistema Sanitario de Navarra* 31(2): 135-152.

Beres, A. 2017. Time is of the Essence: A Review of Electroencephalography (EEG) and Event-Related Brain Potentials (ERPs) in Language Research. *Appl Psychophysiol Biofeedback* 42: 247–255.

Bornkessel, I.; Staub, A.; Schlesewsky, M. 2016. The Timecourse of Sentence Processing in the Brain. In: G. Hickok; S. Small (ed.), *Neurobiology of Language*, p.607-621. Amsterdam: Elsevier.

Bowen, A. 2019. 'You Don't Have to Say Anything': Modality and Consequences in Conversations About the Right to Silence in the Northern Territory. *Australian Journal of Linguistics* 39(3): 347-374.

Churchland, P.S. 1986. *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge, MA: The MIT Press.

Churchland, P.S. 2002. *Brain-Wise: Studies in Neurophilosophy*. Cambridge, MA: Th MIT Press.

Churchland, P.S. 2013. *Touching a Nerve: The Self as Brain*. New York: W. W. Norton & Company.

Creswell, J.; Poth, C. 2018. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. 4th ed. Thousand Oaks, CA: SAGE Publications.

Escobar-Pérez J.; Cuervo-Martínez Á. 2008. Content Validity and Expert Judgment: An Approach to Its Use. *Av In Measurement* 6: 27-36.

Estany, A. 2022a. Naturalización de la ética y la moral. *Revista De Humanidades De Valparaíso* 19: 293-312.

Estany, A. 2022b. Guest Editors' Introduction: Challenges from Neuroscience to Philosophy. *Revista De Humanidades De Valparaíso* 20: 9-12.

Feldman, J.; Narayanan, S.; Lakoff, G. 2000. Active Schemas for Generation, Recognition, and Understanding of Action. *International Conference on Mirror Neurons and the Evolution of Language and Cognition*, Delmenhorst.

Felisatti, A.; Fischer, M-H. 2024. *Experimental Methods in Embodied Cognition How Cognitive Psychologists Approach Embodiment*. London and New York: Routledge.

Fernández, J. 2014a. Aspectos cognitivos y construccionales de la evolución semántica del verbo tocar. *Anuari de Filologia Estudis de Lingüística* 4: 93-118.

Fleiss, J. L. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76(5): 378–382.

Frijters, S.; Van De Putte, F. 2021. Classical term-modal logics. *Journal of Logic and Computation* 31(4): 1026-1054.

Ghio, M.; Telamandi, M. 2015. Grounding Sentence Processing in the Sensory-Motor System. In: M. Gazzaniga (ed.), *The Cognitive Neurosciences*, p.647-661. London: The MIT Press.

Gibbs, R. 2007. Why cognitive linguists should care more about empirical methods. In: V. Evans; B. Bergen; J. Zinken (ed.), *The Cognitive Linguistics Reader*, p.40-56. London: Equinox.

Gillioz, C.; Zufferey, S. 2021. *Introduction to Experimental Linguistics*. Cognitive Science Series. London: ITSE.

Gisev, N.; Bell, J. S.; Chen, T. F. 2013. Interrater agreement and interrater reliability: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy* 9(3): 330-338.

Gosselin, L. 2005. *Temporalité et modalité*. Bruxelles: Duculot-de Boeck.

Gosselin, L. 2010. *Les modalités en français. La validation des représentations*. Etudes Chronos. Amsterdam-New York: Rodopi.

Hagoort, P.; Indefrey, P. 2014. The neurobiology of language beyond single words. *Annual Review of Neuroscience* 37: 347-362.

Hauk, O. 2016. What Does It Mean? A Review of the Neuroscientific Evidence for Embodied Lexical Semantics. In: G. Hickok; S. Small (ed.), *Neurobiology of Language*, p.777-790. Amsterdam: Elsevier.

Knobe, J.; Nichols, S. 2008. An experimental philosophy manifesto. In: J. Knobe; S. Nichols (ed.), *Experimental philosophy*, p.3-14. New York: Oxford University Press.

Larson-Hall, J.; Plonsky, L. 2015. Reporting and Interpreting Quantitative Research Findings: What Gets Reported and Recommendations for the Field. *Language Learning* 65(1): 127-159.

Loaiza, N.; Ortiz, C.; Restrepo, J. 2022. *Estilos de aprendizaje, interculturalidad y modalidades en el texto argumentativo escrito bilingüe*. Armenia, Colombia: Editorial Kinesis.

Loewen, S.; Godfroid, A. 2020. Advancing quantitative research methods. In: J. Mckinley; H. Rose (ed.), *The Routledge Handbook of Research Methods in Applied Linguistics*, p.98-107. New York: Routledge.

Luck, S. 2014. *An introduction to the event-related potential technique*. Cambridge, MA: The MIT Press.

Lupyan, G. 2016. The centrality of language in human cognition. *Language Learning* 66(3): 516-553.

Malaia, E.; González, J.; Weber, C.; Talavae, T.; Wilbur, R. 2015. Neural Processing of Verbal Event Structure: Temporal and Functional Dissociation Between Telic and Atelic Verbs. In: R. Almeida; C. Manouilidou (ed.), *Cognitive Science Perspectives on Verb Representation and Processing*, p.131-140. London: Springer.

Maldonado-Suárez, N.; Santoyo-Telles, F. 2024. Content validity by expert judgment: Quantitative and qualitative integration in the construction of measurement instruments. *REIRE Revista d'Innovació i Recerca en Educació* 17(2): 1-19.

Marsden, E. 2020. Methodological transparency in applied linguistics and its consequences for the quality and scope of research. In: J. McKinley; Heath, R. (ed.), *Methods in Applied Linguistics*, p.15-28. London: Routledge.

Martí, M. A.; Martín-Valdivia, M.; Taulé, M.; Jiménez-Zafra, S.; Nofre, M.; Marsó, L. 2016. Denial in Spanish: Analysis and typology of denial patterns. *Natural Language Processing* 57: 41-48.

McAlister, A.; Lee, D.; Ehlert, K.; Kajfez, R.; Faber, C.; Kennedy, M. 2017. Qualitative Coding: An Approach to Assess Inter-Rater Reliability. *American Society for Engineering Education*, Paper ID #18189.

New, B.; Pallier, C.; Ferrand, L.; Matos, R. 2001. Une base de données lexicales du français contemporain sur Internet : LEXIQUE. *L'Année Psychologique* 101 : 417–462.

Northoff, G. 2004. What Is Neurophilosophy? A Methodological Account. *Journal for General Philosophy of Science* 35: 91-127.

Nozari, S.; Thompson-Schill, S. 2016. Left Ventrolateral Prefrontal Cortex in Processing of Words and Sentences. In: G. Hickok; S. Small (ed.), *Neurobiology of Language*, p.569-586. Amsterdam: Elsevier.

Pallier, C.; Devauchelle, A.; Dehaene, S. 2010. Cortical representation of the constituent structure of sentences. *PNAS* 108(6): 2522-2527.

Proverbio, A.M. 2023. EEG and ERPs in the Study of Language and Social Knowledge. In: P. Boggio; T. Wingenbach; M. da Silveira; E. Comfort; L. Murrins; M. Alves (ed.), *Social and Affective Neuroscience of Everyday Human Interaction*, p.195-211. Cham: Springer.

Prinz, J. 2008. Empirical Philosophy and Experimental Philosophy. In: J. Knobe; S. Nichols (ed.), *Experimental philosophy*, p.189-208. New York: Oxford University Press.

Pulvermuller, F.; Fadiga, L. 2016. Brain Language Mechanisms Built on Action and Perception. In: G. Hickok; S. Small (ed.), *Neurobiology of Language*, p.311-321. Amsterdam: Elsevier.

Quine, W. V. 1969. Epistemology naturalized. In: *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Restrepo Rodas, J.; Niño Buitrago, L.; Suárez, M. 2022. Naturalizing Morality. From Alethic to Deontic and Axiological Values: The Case of Tocar, a Colombian Spanish Verb. *Valparaíso: Revista de Humanidades de Valparaíso* 20: 77–99.

Rogalski, C. 2016. Sentence Processing. In: G. Hickok; S. Small (ed.), *Neurobiology of Language*, p.587-596. Amsterdam: Elsevier.

Saldaña, J. 2021. *The Coding Manual for Qualitative Researchers*. Thousand Oaks, CA: SAGE Publications.

Serratrice, L.; Hervé, C.; Corley, M. 2015. Dislocations in French-English bilingual children: An elicitation study. *Bilingualism: Language and Cognition* 19(5): 987-1000.

Soriano, A. 2014. Design and validation of measurement instruments. *Dialogues* 8(14): 19-40.

Suárez-Rivero, D. 2024. Is Logic Useful for Doing Philosophy?. *Principia* 28(1): 1-4.
Supo, J. 2013. *Cómo Validar un instrumento*. Lima: Biblioteca Nacional de Perú.
Turkeltaub P.; Eden G.; Jones K.; Zeffiro T. 2002. Meta-analysis of the functional neuroanatomy of single word reading: method and validation. *Neuroimage* 16(3): 765-780.
Vermeiren, H.; Vandendaele, A.; Brysbaert, M. 2023. Validated tests for language research with university students whose native language is English: Tests of vocabulary, general knowledge, author recognition, and reading comprehension. *Behavior Research Methods* 55: 1036-1068.
Wang, H. 2018. Thinking for speaking and gesturing motion events: A typological perspective. *Linguistic Approaches to Bilingualism* 8(4): 1-21.
Wu, Q. 2024. Alethic Modality is Deontic. *Mind and Language* 39(4): 561-577.
Yamada, Y.; Neville, H. 2007. An ERP study of syntactic processing in English and nonsense sentences. *Brain Research* 1130(1): 167-180.
Young, L.; Saxe, R. 2008. The neural basis of belief encoding and integration in moral judgment. *Neuroimage* 40(4): 1912–1920.
Young, L.; Saxe, R. 2009. An FMRI investigation of spontaneous mental state inference for moral judgment. *Journal of cognitive neuroscience* 21(7): 1396–1405.
Zapico, G.; Vivas, J. 2015. Synonymy from a linguistic-cognitive perspective. Measurement of semantic distance. *Onomazein* 32, 198-211.

## Acknowledgments

The project was part of the internal research call 014 at Universidad del Quindío and was registered under ID 1195. This research was also supported by Universidad del Valle through the allocation of institutional research hours to one of the authors.

We extend our sincere gratitude to the validation panel who enriched the statements with their thoughtful insights and comments during the evaluation process: Jesús Goenaga, Alexandra Botero, and Leslie Moreno. We are especially thankful to the teachers from SENA Dosquebradas, whose generosity and enthusiasm during the exploratory phase added meaning and depth to this project. We wish to thank the anonymous reviewers for their valuable time and feedback. Special thanks to Universidad del Quindío for institutional support, and to Juan Felipe Salcedo for his unwavering assistance, without which this project would not have been possible.

The authors declare that there is no conflict of interest regarding the publication of this article.