

Publication status: This preprint has not been published elsewhere.

On the semantics of em: polysemy, synonymy and competition in the prepositional system of Brazilian Portuguese

Thayse Ferreira

<https://doi.org/10.1590/SciELOPreprints.13776>

Submitted on: 2025-10-15

Posted on: 2025-11-11 (version 1)

(YYYY-MM-DD)

ON THE SEMANTICS OF *EM*: POLYSEMY, SYNONYMY AND COMPETITION IN THE PREPOSITIONAL SYSTEM OF BRAZILIAN PORTUGUESE

SOBRE A SEMÂNTICA DE *EM*: POLISSEMIA, SINONÍMIA E COMPETIÇÃO NO SISTEMA PREPOSICIONAL DO PORTUGUÊS BRASILEIRO

Theoretical Essay

Thayse Letícia Ferreira¹

ABSTRACT: The preposition *em* is the most frequently used spatial item in Brazilian Portuguese (Ilari et al., 2015), a fact often attributed to its high degree of polysemy (Oliveira & GoodGod, 2018; Rodrigues et al. 2020). This paper aims to investigate the polysemy of *em* within the theoretical frameworks of Nanosyntax and Formal Semantics, with a focus on understanding how its different interpretations arise from the interaction between formal features of syntactic computation and encyclopedic knowledge. Methodologically, the analysis combines nanosyntactic decomposition of the universal functional sequence with semantic diagnostics, enabling us to identify which spatial properties are formally encoded and which emerge from conceptual factors. We propose that the notions of adjacency, contact, and inclusion, widely recognized in the descriptive literature, should be treated as independent features in the *f-seq*, built incrementally. Within this perspective, the core interpretations associated with *em* can be derived through the Superset Effect (Caha et al., 2025): inclusion presupposes contact, which in turn presupposes adjacency. The findings further indicate that finer interpretive nuances emerge from the encyclopedic component, especially from the interaction between Figure and Ground. Grounds denoting volumetric entities license all three readings (adjacency, contact, and inclusion), while Grounds denoting surface entities restrict interpretation to *contact* only. Finally, we discuss cases of competition between *em* and complex prepositional phrases such as *dentro de*, showing how the interplay between nanosyntactic structure and encyclopedic knowledge accounts for the distribution of relevant spatial readings in Brazilian Portuguese.

KEYWORDS: Preposition *em*. Prepositional polysemy. Nanosyntax of space. Syntax-semantics interface.

RESUMO: A preposição *em* é o item espacial mais utilizado pelos falantes de português brasileiro (Ilari et al., 2015), possivelmente em razão de seu alto grau de polissemia (Oliveira & GoodGod, 2018; Rodrigues et al., 2020). Este trabalho tem como objetivo investigar a polissemia de *em* no quadro teórico da Nanossintaxe e da Semântica Formal, com foco em compreender de que modo suas diferentes interpretações resultam da interação entre traços formais da computação sintática e o componente enciclopédico. Metodologicamente, a análise

¹ PhD in Linguistic Studies (UFSCar), Federal University of Paraná (UFPR), Graduate Program, Curitiba, Paraná, tleticiaf@gmail.com, <https://orcid.org/0000-0003-2883-8721>. The author declares that there are no conflicts of interest in this work. Data Availability Statement: this study does not rely on external empirical data. All examples were constructed from the author's native speaker intuition, following standard practice in formal syntax and semantics. Therefore, no additional data are available.

combina a decomposição nanossintática da sequência funcional universal (*f-seq*) com diagnósticos semânticos, permitindo identificar quais propriedades espaciais são formalmente codificadas e quais derivam de fatores conceituais/enciclopédicos. Propomos que as noções de adjacência, contato e inclusão, amplamente descritas na literatura, devam ser tratadas como traços independentes na *f-seq*, construídos incrementalmente. Nessa perspectiva, as interpretações básicas associadas a *em* podem ser derivadas pelo Efeito do Superconjunto (Caha et al., 2025): inclusão pressupõe contato, que, por sua vez, pressupõe adjacência. Os resultados indicam, ainda, que algumas nuances de sentido emergem do componente enciclopédico, especialmente da relação entre Figura e Fundo. Fundos que denotam entidades volumétricas, por exemplo, licenciam as três leituras (adjacência, contato e inclusão), ao passo que Fundos que expressam entidades de superfície restringem a interpretação a contato. Por fim, discutimos também os casos de competição entre *em* e sintagmas complexos como *dentro de*, evidenciando como a interação entre a estrutura formal e o conhecimento enciclopédico molda a semântica do sistema preposicional do português brasileiro.

PALAVRAS-CHAVE: Preposição *em*. Polissemia preposicional. Nanossintaxe espacial. Interface Sintaxe-semântica.

ABSTRACT FOR NON-SPECIALISTS: The preposition *em* is one of the most frequently used words in Brazilian Portuguese, especially to indicate location. Although it seems simple, it can express many different meanings. For example, the sentence “o gato está no armário” (“the cat is in the cupboard”) may mean that the cat is inside, on top of, or next to the cupboard. In this study, we explore why such variation occurs and how it can be explained. To do so, we draw on linguistic theories that investigate how language organizes its most basic elements and how speakers interpret words depending on context. We argue that three main ideas help explain the meanings of *em*: *adjacency* (being close), *contact* (being in touch), and *inclusion* (being inside). These meanings are related: whenever there is inclusion, there is also contact, and whenever there is contact, there is adjacency. We also show that the meaning of *em* depends not only on the word itself but also on the nature of the objects involved. For instance, with volumetric objects like a drawer, *em* can mean “inside,” while with surface objects like a table, it usually means “on top of.” Finally, the study discusses how *em* interacts with longer expressions such as *dentro de* (“inside of”), shedding light on how the Brazilian Portuguese prepositional system works.

Introduction

The notion of “space” has attracted considerable attention across a range of disciplines, including philosophy, geography, and linguistics, insofar as it represents a cognitive domain “indispensable to our experience of the world” (Vandeloise, 2006). Yet, the concept of “space” remains elusive: although spatial relations are physically observable, they cannot be apprehended through direct sensory contact. As a result, this category occupies an intermediate position on the *continuum* between concrete and abstract entities/facts/events. From a socio-geographical perspective, Santos (2006) defines space as “an indissociable set of systems of objects and systems of actions”. This articulation between objects and events extends our

perception of physical spaces to encompass broader dimensions of human experience, such as emotions and temporal awareness, as the examples below illustrate.

- (1) a. O passado ficou para trás
 the.masc past stay.PST.PFV.3SG to back
 ‘The past was left behind.’
- b. Hoje eu estou para cima
 today I be.PRS.1SG to up
 ‘Today I am feeling up / in high spirits!’
- c. Hoje eu estou muito para baixo
 today I be.PRS.1SG very to down
 ‘Today I am feeling very down / depressed.’

In the linguistic domain, the interface between language and cognition has often been explored through processes such as analogy and metaphor, which give rise to expressions in which, for example, positive and negative emotions are conceptualized as opposite directions along a vector (an oriented line segment). As Jackendoff (2012) argues, any theory of the language faculty must ultimately account for “how we talk about what we see”. From this perspective, understanding how human languages encode spatial notions is crucial, since the extension of spatial concepts into other domains, such as time, can only be properly understood once the underlying conceptual basis is well described.

Although the ways in which human languages structure space have long been a central concern in linguistics (e.g., Jackendoff, 1983; Talmy, 2000; Ameka & Levinson, 2007), many questions remain open. For instance, what specific features license the metaphorical extensions illustrated in (1)? Which linguistic categories are most apt to convey spatial notions? And do we observe comparable realizations of space across the world’s languages? These are just a few of the questions at stake in the study of space and language (cf. Talmy, 1983; Landau & Jackendoff, 1993; Ursini, 2010; Svenonius, 2012). In this paper, we aim to advance this discussion by examining how languages frame and externalize spatial notions, with particular attention to the prepositional system of Brazilian Portuguese (BrP). More specifically, our goal is to analyze the semantic behavior of the locative preposition ‘*em*’ within the frameworks of Formal Semantics (Zwarts & Winter, 2000; Portner, 2005; Partee, 2011) and Nanosyntax (Starke, 2009; Baunaz et al., 2018).

The choice of theoretical framework, situated at the interface between syntax and semantics, is motivated by the issues addressed in this work: the alleged polysemy of the preposition *em* and its synonymy with expressions involving complex prepositional phrases,

such as *dentro de* (“inside of”), which leads to an apparent competition within the linguistic system regarding the choice of one form over another. The relation between the sentences in (2) and (3) below illustrates these issues.

- (2) O gato está no armário
 the.masc cat be.PRS.3SG in-the.masc cupboard
 ‘The cat is in the cupboard.’
- (3) a. O gato está dentro do armário
 the.masc cat be.PRS.3SG inside of-the.masc cupboard
 ‘The cat is inside the cupboard.’
 b. O gato está embaixo do armário
 the.masc cat be.PRS.3SG under of-the.masc cupboard
 ‘The cat is under the cupboard.’
 c. O gato está em cima do armário
 the.masc cat be.PRS.3SG on top of-the.masc cupboard
 ‘The cat is on top of the cupboard.’

The sentence provided in (2) is ambiguous, as it may describe (at least) three different situations: given a locative relation between “the cat” and “the cupboard,” the proposition can be true if the cat is (i) inside the cupboard; if the cat is located (ii) at the bottom of the cupboard, that is between the base of the piece of furniture and the floor; or, alternatively, if the cat is (iii) on top of the cupboard. It should also be noted that each of these situations could be more precisely described by a complex expression, which may or may not contain the preposition *em*, as illustrated by the sentences in (3). From this contrast, two central questions arise: is there something in the structure of *em* that gives rise to this range of interpretations for spatial locative PPs? And, given that speakers could choose other structures to convey the same interpretation, what regulates this competition in terms of the syntactic derivation?

To address these questions, concerning the phenomena of polysemy, synonymy, and competition, a well-articulated theoretical framework is required, and this is precisely what we aim to achieve by adopting the assumptions of Formal Semantics (Zwarts & Winter, 2000; Portner, 2005; Partee, 2011) and Nanosyntax (Starke, 2009; Baunaz et al., 2018). More specifically, Nanosyntax enables a fine-grained investigation of the grammatical features involved in broad linguistic-cognitive domains, such as space, by rigidly organizing them into the universal functional sequence (*f-seq*) (Cinque & Rizzi, 2010). This hierarchy of features is furthermore encoded in lexical items, which mediate the process of lexicalization by mapping the abstract syntax onto phonological and semantic content. Within this framework, the

semantics of formal features can constrain how natural language sentences are composed, working in tandem with a rigid externalization algorithm that, in turn, regulates cases of competition. In addition, the data examined in this study also appear to require an incursion into the domain of encyclopedic knowledge, since the interpretations available for *em* are guided by our world knowledge about the objects involved in the spatial relation. The sentence below, for instance, cannot be true under the same conditions as the sentence “the cat is in the cupboard,” discussed above, even though, in terms of syntactic-semantic categories, the structure is precisely the same as in (2).

- (4) O quadro está na parede
 the.masc painting be.PRS.3SG em-the.fem wall
 ‘The painting is on the wall.’

In this case, there is only one interpretation available: the one of contact between the picture and the surface of the wall; the readings of *on top of*, *underneath*, and *inside* are strongly blocked. Nanosyntax provides a framework for understanding how encyclopedic knowledge can interfere with syntactic derivations, which makes it particularly well suited to the analysis developed in this paper. Within this framework, we argue that it is possible to closely examine the spatial notions conveyed by the preposition *em*, as well as to account for how polysemy relates to the syntactic derivation, guided by principles of competition among lexical items.

With the aim of examining the semantic functioning of the preposition *em* and its high degree of polysemy, particularly in relation to synonymy and competition with analytic PPs, the paper is organized as follows: Section 1 presents a description of the behavior of the preposition *em* in Brazilian Portuguese, focusing on what the literature has discussed regarding this item, especially in relation to its syntactic-semantic functioning with spatial interpretation. Section 2 introduces the theoretical framework, outlining the core assumptions of Nanosyntax (Caha et al., 2025) and the spatial features that have been identified as linguistically relevant by researchers working within this model (Pantcheva, 2011; Svenonius, 2012; Ferreira, 2021). As we shall see, Nanosyntax provides an articulated hierarchy for the spatial domain, incorporating cognitive notions that have been explored in non-formal perspectives, such as Figure, Ground and Place. The existence of this hierarchy and its terminal nodes will be addressed in Section 3, where we offer a nanosyntactic analysis of the preposition *em*, highlighting its potential range of interpretations, based on topological notions identified in the previous literature. Finally, in Section 4 we present the concluding remarks.

1. The Preposition *em* in Brazilian Portuguese (BP)

Historically, the weakening of the Latin case system led to the emergence of prepositions in the Romance languages (Bassetto, 2010), from Vulgar Latin. Within this context, *em* is a peculiar preposition, as it has remained virtually unchanged throughout diachronic development, both semantically and phonologically. While the phonological connection between the Latin form *in* and the Portuguese form *em* is quite evident, it is important to note that the Latin *in* conveyed two distinct interpretations, associated respectively with the accusative and ablative cases: a spatial boundary, marking the end of a movement through space, and static location — most notably with the meaning of *inside of*. According to Ilari et al. (2015), synchronically, *em* stands out as the most frequently used spatial preposition in Brazilian Portuguese. Its versatility allows it to appear both with verbs of motion and in contexts that express topological location, thereby retaining the core semantic values inherited from Latin.²

- | | |
|------------------------------------|--|
| (5) Pedro foi no mercado. | [‘em’=final place of movement] |
| ‘Pedro went to the market’ | |
| (6) Os documentos estão na pasta. | [‘em’=topological relation of containment] |
| ‘The documents are in the folder.’ | |

Ilari et al. (2015), working within the framework of Cognitive Linguistics, describe the preposition *em* in terms of two spatial categories, namely a position relative to a horizontal axis and a spatial arrangement in a containment schema. The first category is associated with the interpretation of a goal in a motion event, i.e. the end of a path, as the example (5) above, while the second one encodes a containment schema and conveys relations of the type *inside of*, as illustrated in (6). Schematically, the authors represent these relations as follows:

² Some studies suggest that the preposition *em* is ambiguous and may convey both the notion of static location and that of path (Wiedemer, 2013). In this work, we follow the line of Rammé (2017) and Ferreira (2021) and assume that the Place and Path interpretations associated with *em* are, in fact, a case of false syncretism. Although *em* may occur as the complement of a verb of motion, the notion introduced by this item in the derivation is necessarily that of a static location, which is then interpreted as the endpoint of the motion event or, alternatively, as the Goal of the Path.

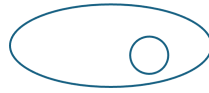


Figure 1. Image scheme of *em* as a container in static space (Ilari et al., 2015, p. 251)



Figure 2. Image scheme of *em* marking the end of a path (Ilari et al., 2015, p. 251)

In addition to conveying the Latin senses of *in*, related to the semantic roles of “inside” (Figure 1) and “goal” (Figure 2) (Ilari et al., 2015, p. 205), it is assumed in the literature that the preposition *em* has also come to encode other topological notions, such as support, boundary, and adjacency, which describe spatial interactions between objects. Notably, the study by Oliveira and GoodGod (2018) suggests that the preposition *em* conveys the topological notions of inclusion (a relation in which one object is contained within another), contact (a relation of connectivity between two objects), and adjacency (a relation of proximity between two objects). These concepts were mapped by the authors through theoretical research and examined in a psycholinguistic experiment “based on perceived similarities in the meaning of the preposition” (Oliveira and GoodGod, 2018, p. 12), which led to a list of four possible senses for the preposition *em* in Portuguese.

(7) Static Inclusion

- a. O pão está no saco de papel pardo.
‘The bread is in the brown paper bag.’
- b. Há muitos icebergs no mar.
‘There are many icebergs in the sea.’
- c. Tem um buraco na parede.
‘There is a hole in the wall.’

(8) Inclusion in a motion event

- a. Pedro jogou os livros na bolsa.
‘Pedro threw the books into his bag.’
- b. Eu fui na padaria.
‘I went to the bakery.’

(9) Contact

- a. As faixas estão estendidas no gramado.
‘The banners are stretched out on the lawn.’
- b. Tem um retrato na parede.
‘There is a portrait on the wall.’
- c. A placa está no acostamento.
‘The sign is on the shoulder.’

- (10) Adjacency
- a. Ela estava sentada na mesa.
'She was sitting at the table.'

There is, admittedly, a subtle semantic distinction between examples such as (7b) “iceberg no mar” (“iceberg in the sea”) and (7c) “buraco na parede” (“hole in the wall”). Nevertheless, as Oliveira and GoodGod (2018) argue, such nuances do not appear to influence speakers’ judgments concerning the broader categories encoded by *em*. This observation supports the decision to describe the preposition in terms of only three topological concepts. These topological relations are likewise examined in Rodrigues, Santos, and Lopes (2017) and Rodrigues et al. (2020). The key distinction between these approaches lies in their theoretical orientation: whereas Oliveira and GoodGod (2018) adopt a cognitive perspective, Rodrigues, Santos, and Lopes (2017) and Rodrigues et al. (2020) employ a formal calculus designed for computational applications. As a result of this orientation, the latter works incorporate into their analysis certain semantic nuances associated with the preposition *em* that, by contrast, are filtered out in the cognitive account proposed by Oliveira and GoodGod (2018).

The central idea pursued by these authors is that the preposition *em* is a polysemous item that can be treated similarly to vague terms, whose meanings are determined by contextual variables. This proposal is based on the observation that the precise meaning of the preposition depends on the context established by the arguments involved in the locative relation, referred to as Figure and Ground:

The spatial configurations elicited by the prepositional phrases ‘A nail in the wall’ and ‘A nail in the box’ bring out an important positional difference between the nails. If the first one is graphically depicted or verbally described, the nail will most probably be represented in a perpendicular position to the wall. In the latter, the nails in the box take positions inside it. In the first scenario, the wall serves as a surface to accommodate the nail, whereas in the latter, the box is usually a container for the nails. (Rodrigues et al., 2020, p. 638)

The contrast between sentences such as “a nail in the wall” and “a nail in the box” illustrates that some meaning distinctions cannot be fully accounted for within a strictly compositional semantics. For this reason, a semantic model that considers the speaker’s belief states or possible world scenarios may be more appropriate for addressing polysemy in the prepositional domain. Based on this idea, Rodrigues, Santos, and Lopes (2017), as well as Rodrigues et al. (2020), suggest that in order to capture the high degree of polysemy associated with *em*, it is necessary to posit a semantics of precisification, grounded in a supervaluationist semantics, which “explain linguistic indeterminacy in terms of a collection of possible precise

interpretations of the terms of the language” (Bennett, 2011). In a certain sense, all possible interpretations are contained within the lexical item, and the context determines which reading is selected in each situation. This context-dependence is captured by a viewpoint semantics (Bennett, 2011), which “characterizes the range of possible worlds and interpretations that are considered plausible or acceptable” (Rodrigues et al., 2020, p. 648).

In addition to viewpoint semantics, Rodrigues et al. (2020) adopt, for the description of *em*, a formalism known as Region Connection Calculus (RCC), based on a primitive and symmetric relation between pairs of spatial regions. This calculus begins with a primitive relation $C(x, y)$, read as *x is connected to y*, where *x* and *y* are two regions of an Euclidean space (a real vector space that allows for distance measurement) with at least one point in common. Below, we provide a graphical representation of how this calculus operates over spatial regions:

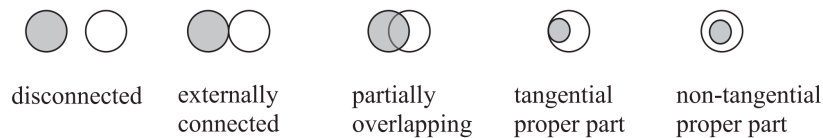


Figure 3. Region Connection Calculus (Zwarts, 2017, p. 4)

As can be observed in the relations above, the RCC formalism operates with concepts such as disconnection, connection only to the exterior of regions, overlap, and inclusion, the latter modeled as the property of proper part. A structure such as “as faixas estão no gramado” (“the banners are stretched out on the lawn”), for example, in which the Figure is located on top of the Ground and the regions are partially overlapping, could be described as follows:

- (11) a. As faixas estão no gramado.
 b. $PO(x, y) \equiv_{\text{def}} O(x, y) \wedge \neg P(x, y) \wedge \neg P(y, x)$
 c. $PO(\text{faixas}, \text{gramado}) \equiv_{\text{def}} O(\text{faixas}, \text{gramado}) \wedge \neg P(\text{faixas}, \text{gramado}) \wedge \neg P(\text{gramado}, \text{faixas})$

In words, (11b) states that a relation of partial overlap (PO) between two regions, *x* and *y*, can be defined by the existence of an overlap between regions *x* and *y*, $O(x, y)$, such that *x* is not a part of *y* ($\neg P(x, y)$) and *y* is not a part of *x* ($\neg P(y, x)$). Namely, the regions partially overlap if there is a portion of space shared by both, provided that the two regions do not coincide at all their points. Based on this type of relation, Rodrigues et al. (2020, p. 651) provide the following

list of possible interpretations for the preposition *em*. Each of these readings constitutes a precisification of the relation between Figure and Ground, represented as $p_n(x, y)$.

$$\begin{aligned}
 p_1(x, y) &\equiv_{def} \text{GEO-INSIDE}(x, y) \\
 p_2(x, y) &\equiv_{def} \text{P-INSIDE}(x, y) \\
 p_3(x, y) &\equiv_{def} \text{CONT-INSIDE}(x, y) \\
 p_4(x, y) &\equiv_{def} \text{OUTSIDE}(x, y) \wedge \exists z((\text{GEO-INSIDE}(z, y) \vee \text{P-INSIDE}(z, y)) \wedge \text{EC}(x, z)) \\
 p_5(x, y) &\equiv_{def} \text{PO}(x, y) \\
 p_6(x, y) &\equiv_{def} \text{TPP}(x, y) \vee \text{NTPP}(x, y) \\
 p_7(x, y) &\equiv_{def} \neg\text{INSIDE}(x, y) \wedge \text{VERT}(x, y) \wedge \text{EC}(x, y) \\
 p_8(x, y) &\equiv_{def} \text{TOP-INSIDE}(x, y).
 \end{aligned}$$

Figure 4. Meanings of *em* according to Rodrigues et al. (2020, p. 651)

These relations can be informally described as follows: the first precisification relation accounts for readings in which the Figure is contained or included within the Ground, whose structure is geometric and convex. This captures the reading of *em* in a sentence such as “os panos estão na bacia” (“the cloths are in the basin”). The precisification in p_8 , in turn, also captures an inclusion reading of the Figure within the Ground. The difference between the *GEO-INSIDE* and *TOP-INSIDE* relations is that the former captures geometrically complex Grounds, while the latter incorporates Grounds with closed topological properties, yielding a reading of encapsulation of one object by another, as in “a semente está na uva” (“the seed is in the grape”). Interpretation p_2 suggests that the Figure may be partially encapsulated by the Ground: located inside it, but in contact with other objects. A closely related case is p_3 , in which the Figure is nearly outside the Ground, but remains in contact with it due to its proximity/connection with other objects contained within it. This scenario would be exemplified by a sentence such as “a maçã está na fruteira” (“the apple is in the fruit bowl”), where the Figure “the apple” may be positioned on top of other fruits, completely disconnected from the Ground, but still considered to be, in some sense, spatially contained.

The precisification in p_4 indicates that the Figure may be situated outside the physical boundaries of the Ground, provided there exists a spatial point of external connection between the two entities. An illustrative case is the sentence “o carro está na farmácia” (“the car is at the pharmacy”), in which the Figure is geometrically external to the Ground, yet still falls within a spatial region that can be construed as constituting “the pharmacy”. In addition to these connection-based regions, Rodrigues et al. (2020) propose three other precisifications for the meanings of *em*: *PO* (p_5), in which the Figure and Ground regions partially overlap (“o lençol está na cama”/“the sheet is on the bed”); *TPP* (p_6), in which the Figure is a tangential proper

part of the Ground (“a maçaneta na porta”/ “the doorknob on the door”), meaning that the Figure is located at the boundary of the spatial configuration of the Ground; and $\neg INSIDE(p7)$, which licenses a support relation between Figure and Ground (“o copo na mesa”/ “the glass on the table”). It is worth noting how this proposal engages with the notions of inclusion, contact, and adjacency, while the use of Region Connection Calculus enables more precise and diversified interpretations for the preposition.

The proposals reviewed in this section offer robust accounts of the semantic behavior of the preposition *em*. However, one relevant issue remains insufficiently addressed in explaining the considerable degree of polysemy in the prepositional domain — though it is briefly mentioned in Rodrigues, Santos, and Lopes (2017) and Rodrigues et al. (2020): the role played by the elements denoting Figure and Ground. As the preceding discussion has shown, structurally similar sentences may nonetheless give rise to distinct interpretations. This suggests that the source of the polysemy, or rather, the proliferation of spatial meanings associated with *em*, may not, in fact, reside just in the preposition itself, but instead in the other elements that compose the sentence structure. Rodrigues et al. (2020) acknowledge that this phenomenon cannot be accounted for compositionally and therefore propose that *em* carries eight distinct meanings, with the relevant reading being determined by context.

In this study, we aim to explore the hypothesis that the contribution of *em* remains relatively stable across its various precisifications. Accordingly, we assume that compositional constraints may arise from the encyclopedic component, associated with the lexical entries of the Figure and Ground. In this sense, we also acknowledge that context determines the relevant reading of the structure; however, these possible interpretations are not embedded in the formal architecture of the preposition itself and, to some extent, can be derived “compositionally”. Furthermore, it is important to highlight that previous studies on *em* in Brazilian Portuguese have not considered the competition between the preposition and its analytic counterparts. Thus, this research differs on two fronts: first, by seeking to minimize the number of meanings attributed to the preposition *em*, and second, by aiming to understand how speakers choose between a simple or complex structure when describing a scene involving spatial location. To this end, we now present the theoretical framework that will guide our analysis.

2. Nanosyntax and the Decomposition of Space in the model

Nanosyntax is a recent model of grammatical architecture, developed methodologically in the wake of Cartography, given its concern with the number and order of components in the

universal functional sequence (*f-seq*), and operationally in the wake of Distributed Morphology, since it shares the principle of Late Insertion and assumes the need to posit competition rules for Spell-out.

The model finds its inaugural formulation in Starke (2009), who argues that the syntactic structures/mental representations attributed to speakers had undergone considerable changes since *Syntactic Structures* (Chomsky, 1957). This empirical shift, however, was still being treated within the same theoretical framework established at the beginning of the program, with only minor, mainly terminological, modifications. Broadly speaking, although empirical evidence pointed to the fact that syntax operated with items smaller than syntactic terminals, the generative component was still conceived as a mechanism for organizing the lexicon (composed of morphemes and words) into complete sentences (cf. Caha et al., 2025).

Furthermore, with the advent of Syntactic Cartography (Cinque & Rizzi, 2010), as syntactic structures were expanded, their ingredients became progressively smaller. The nominal domain, in its extended projection, for instance, previously consisting solely of the D label, now comprises a series of heads, such as Number (NumP: singular, plural), Gender (GenderP/GenP), Case (KP: nominative, accusative, dative, genitive, etc.), Person (PersonP/PersP: first person, second person, third person), among others.

This cartographic expansion has ultimately shown that lexical items are in fact fully-fledged syntactic objects, and thus syntax can no longer be conceived merely as an operational device for concatenating words into sentences. Lexical items appear to carry information that is itself constructed by the syntactic component. Inflection in Romance languages, for instance, often bundles within a single morpheme a constellation of features that properly belong to the computational system. Take, for example, the morpheme *-i* in Brazilian Portuguese, which simultaneously encodes first person singular, perfective aspect, indicative mood, and past tense. Each of these features is represented as an independent syntactic terminal in the universal functional hierarchy (*f-seq*). Within a derivational model where each lexical item corresponds to a terminal node, it becomes problematic to capture this many-to-one mapping between lexical items (morphemes and words) and the set of abstract features generated by the syntax. The central challenge, then, is how a single morpheme can be spelled out as the realization of an entire syntactic span.

- (12) a. Eu corri ontem.
b. [-i] = 1SG + PRF + IND + PST

The discussion within Nanosyntax begins from this type of observation and pursues the cartographic agenda, seeking to determine which functional heads constitute the architecture of the universal feature sequence (*f-seq*) and in what order they are to be arranged. On the one hand, then, Nanosyntax aligns itself with the Cartographic program by investigating the mapping of the *f-seq*; on the other hand, it introduces a series of derivational rules designed to constrain the generative power of an increasingly fine-grained inventory of syntactic terminals, which would otherwise risk overgeneration. The formulation of such rules converges with certain assumptions of Distributed Morphology, such as the principle of Late Insertion, according to which Vocabulary Items are inserted only after syntax has built the relevant structures, that is, phonological and conceptual content are introduced only once the syntactic derivation is complete. This assumption is explicitly reflected in the architecture of the Nanosyntactic model.

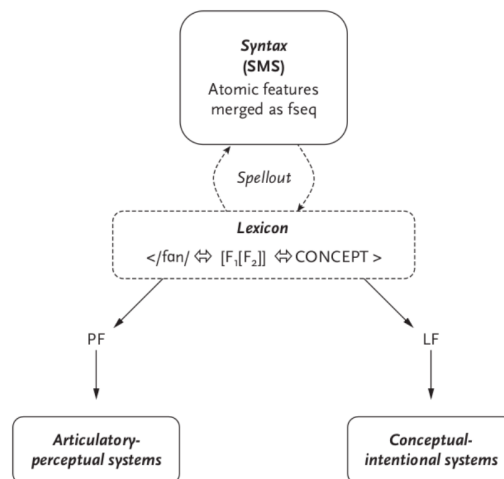


Figure 5. Grammar Architecture in Nanosyntax (Baunaz & Lander, 2018, p. 11)

As illustrated in Figure 5, within Nanosyntax the computational component (syntax) operates over features drawn from domains traditionally labeled as semantics, morphology, and syntax (henceforth SMS). These features are realized as terminals such as Definiteness, Aspect, Agreement, Case, among others. The lexicon, in turn, contains both phonological information (/fon/) and encyclopedic knowledge (CONCEPT), linked to syntactic-hierarchical structures ([F₁ [F₂]]) previously constructed by syntax (SMS).

Crucially, the SMS feature set is organized into a finely articulated hierarchy, and lexical items are conceived as storing fully fledged portions of this hierarchy, i.e. phrasal constituents. The assumption that lexical entries may spell out complex syntactic objects (Phrasal Spell-out) directly addresses the many-to-one mismatch between the abstract syntactic features generated

by the computational system and the morphemes or words realized in natural languages. An inflectional morpheme such as *-i*, discussed earlier, can therefore be represented as a lexical entry of the following type, following Cortiula (2023).

(13) ‘-i’ = </i/, [Asp [Mood [Ind [T [Pst [# [ϕ [Part [Speaker]]]]]]]]]>

The structure of this lexical entry³ must be read hierarchically, from the lowest feature on the left to the highest feature on the right. In this case, the lowest terminals correspond to the semantic domain of perfective aspect and indicative mood; the intermediate features encode past tense [T [Pst]], on top of which the projections of singular number ([#]) and person are built. These features are privative in nature, that is, “their presence indicates the presence of a particular grammatical meaning, whereas their absence indicates the absence of that meaning” (Cortiula 2023: 48). The composition of syntactic terminals can thus be seen as cumulative (Caha 2024): interpretation is reached incrementally, according to the construction of each semantic domain (space, tense, aspect, etc.).

From this example, it becomes clear that the conception of the lexicon is arguably one of the most distinctive traits of Nanosyntax. In this model, the lexicon — like in Distributed Morphology (DM) — is conceived as a postsyntactic component. Crucially, however, unlike in DM, the lexicon is not distributed across multiple lists but concentrated as a repository pairing information built in other components. Note also that, in addition to being fed by syntax (as represented by the arrow from syntax to lexicon in Figure 5), the lexicon also relates back to derivational syntax (as indicated by the arrow from lexicon to syntax in Figure 5), thereby mediating the process of lexicalization. This mediation allows abstract syntax to be clothed with phonological and conceptual content and subsequently transferred to the PF and LF interfaces.

The lexicon in Nanosyntax, therefore, is not merely a passive repository of information (Starke, 2004). While it does not contain derivational rules of its own, it plays an active role in the construction of natural language structures. This process of mediation in lexicalization is governed by a set of principles and rules, some of which will be detailed in the following.

The relation between the structure under construction and the structure stored in the lexicon is established through a checking procedure that applies at each step of *Merge*: the idea is that every structure created by *Merge* must be matched with an existing structure in the

³ It should be noted that in this lexical entry there is no conceptual content associated with phonology and syntax, since it is a strictly functional item.

“lexical entries of the particular language” (Caha et al., 2025, p. 17). This matching is not restricted to terminal nodes: “[...] if the constituent built by syntax is identical to a constituent stored in the lexicon, the entry of that constituent can be used to ‘lexicalize’ the syntactic representation” (Caha et al. 2025: 12). This condition of compatibility, also named *identity*, gives rise to the Superset Effect, also referred to in the literature as the Superset Principle, namely, a lexical entry may lexicalize any phrase contained within it.

(14) Superset Effect/Principle

A lexically stored tree L matches a syntactic node S iff L contains the syntactic tree dominated by S as a subtree.

(Starke, 2009, p. 3)

The Superset Effect allows certain features of a lexical item to remain “unused” during the derivation, provided that the lowest terminal of the lexical entry is always matched to the target/abstract structure (Anchor Condition) and no locality constraints are violated. The Superset Principle has proven highly effective in accounting for morphological phenomena such as syncretism, allomorphy, and root suppletion (cf. Baunaz et al. 2018; Caha et al. 2025).

An illustration of the phrasal Spell-out and the Superset Effect can be found in English irregular verb forms such as *put*, which is syncretic between present and past readings. Such a lexical item arguably stores a larger syntactic structure of the type [T [Pst]], enabling it to appear both in environments of the form [T [Pst]] (yielding the past interpretation) and [T] (yielding the present interpretation). By contrast, an irregular form like *gave* cannot also appear in the present tense, since English lexicon contains a more specific entry (*give*) specified for the [T] head. In other words, more specific lexical entries, those that leave no or fewer features unmatched, win the competition.

Given these basic principles, the lexicalization algorithm proceeds as follows, guided by the movement possibilities outlined in Cinque (2005).

(15) The Lexicalization algorithm

- a. Merge F and lexicalize.
- b. If fail, try a spec-to-spec movement and lexicalize.
- c. If fail, try a movement of the complement of the newly inserted feature and lexicalize.
- d. If fail, go back to the previous cycle, and try the next option for that cycle.
- e. If fail, spawn a new derivation providing feature X and merge that

with the current derivation, projecting feature X to the top node.
 (Caha et al., 2025, p. 18 adapted from Starke 2018)

To illustrate, in simplified terms, how the algorithm operates in association with the principles outlined above, consider the lexicalization of a regular plural form such as *houses* and an irregular form such as *mice* in English, which lexicalizes both Noun [NP] and Plural [# [PI]]. First, the nominal projection is built by syntax and can be lexicalized by an item like *house*. When the projections carrying plural information are introduced, however, there is no single lexical item that matches the entire structure (i.e. the already lexicalized *house* plus plural). In the absence of such an item in the lexicon, the derivational syntax may be manipulated via movement as part of the lexicalization process: step (b) of the algorithm (spec-to-spec) does not apply here, since the structure contains no specifier; step (c) (comp-to-spec), however, generates a constituent that finds a perfect match in the lexicon, allowing it to be lexicalized by the morpheme *-s*, yielding *houses*.

By contrast, the form *mice* is stored in the lexicon as a more specific entry and therefore wins the competition over a structure like *mouses*.

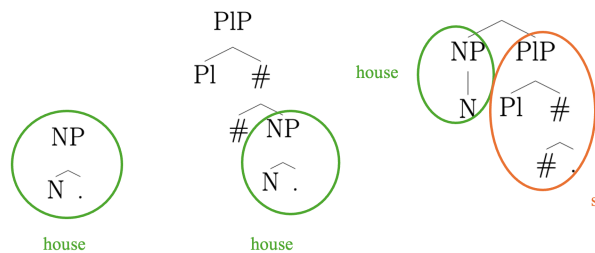


Figure 6. Simplified lexicalization of *houses* by complement movement

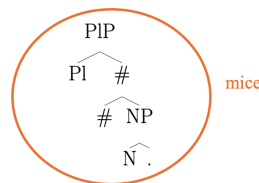


Figure 7. Lexicalization of *mice* by Phrasal Spell-out

In the next section, we explore how these principles and derivational rules can account for the behavior of the preposition *em*. To this end, it is necessary to first illustrate the feature hierarchy for the stative spatial domain as suggested in nanosyntactic research. Below we provide a structure mapping a locative configuration. This hierarchy is based on Romeu (2014) and Ferreira (2021). It should be noted, however, that alternative features and hierarchies have

also been proposed in the literature (cf. Svenonius, 2012). In this paper, we take the hierarchy below to represent the *minimal* structure required for the analysis of the data under discussion, leaving aside features such as Deixis, for instance, that capture items such as *here* and *there*, also locatives.

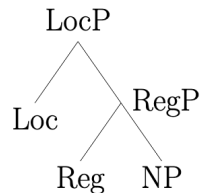


Figure 8. Minimal *f-seq* related to the spatial domain

Assuming, following Wunderlich (1991), that a spatial location relation is asymmetric and holds between an individual (type <e>) and a space (type <l>), the structure above can be interpreted as follows: first, the Region terminal, originally proposed in Romeu (2014), is responsible for taking an ordinary individual, given by the NP, and returning the space occupied by that individual, thus creating the Region of the Ground (Talmy, 2000). The presence of [Reg] in the structure can be confirmed by the fact that certain languages, such as Ainu, Tairora, and Bará, exhibit independent morphology dedicated precisely to expressing the notion of region (Cinque, 2010 apud Romeu, 2014). The example below illustrates this: in Tairora, the suffix *-ra* indicates that the structure in question does not refer to the entity <e>, house, but rather to the spatial region occupied by the entity house, an entity of the type <l>.

- (16) Naabu-qi-**ra** bai-ro.
house-in-**place** be.pres.3sg-he

Crucially, this kind of data provides strong evidence that [Reg] must be represented as a syntactic feature. Moreover, from this example it is noteworthy that region morphology (*-ra*) may co-occur with spatial prepositions (*-qi-*), further supporting the claim that [Reg] occupies an independent position in the functional sequence. From the presence of [Reg], what is being morphologically marked is not a relation of place/location (Loc) or path, but the more primitive notion of region, a domain that underlies and makes possible higher spatial distinctions. Semantically, the notion of region can be understood as a topological space, an unstructured set of spatial points that is not yet divided into measurable units, directions, or coordinates

(Wunderlich, 1991; Zwarts & Winter, 2000). Formally, this head can be represented by the following semantics:

$$(17) \quad [[\text{Reg}]] = \lambda x_{\langle e \rangle}. \chi x_{\langle l \rangle}. \lambda p [\text{EIGEN}(x_l, p) \wedge \text{GROUND}(x_l) \wedge x_e = x_l]$$

Formula (17) states that the denotation of [Reg] is a function that, given an individual of type $\langle e \rangle$ returns a spatial entity of type $\langle l \rangle$ via the chorion operator (χ). This spatial entity is mapped onto a set of points that is occupied by it (the entity's EIGENPLACE), which is identified as the Ground.

Above [Reg], there is the locative terminal [Loc], which plays a crucial role in transforming an unstructured set of points (the region introduced by [Reg]) into a linguistically interpretable *place*. The idea is that while [Reg] simply supplies an unstructured domain of points, [Loc] imposes a geometric organization to it, arranging these points into an ordered space where notions of direction and relative distance, for example, become accessible. Ferreira (2021), following Zwarts & Winter (2000), describes this process as the construction of a vector space: once [Loc] is merged, the system can represent locations as positions along dimensions. In other words, [Loc] converts the topological material provided by [Reg] into a structured domain where Figures can be anchored geometrically in relation to Grounds. Without [Loc], spatial semantics would lack the vector-like properties needed for modification (e.g., “exactly in the corner,” “near the wall”) and for the composition of more complex path structures. The interpretation of this head can be given as in (18).

$$(18) \quad [[\text{Loc}]] = \lambda Q. \exists V. \exists v [|v| \geq 0 \wedge \forall p, q \in \text{Ground} \rightarrow \exists w \in V]$$

The locative terminal [Loc] takes as its complement a predicate denoting a region, i.e. an unstructured set of spatial points, and maps it into a vector space ($\exists V$). In this configuration, the points of the region are organized as vectors (v, w), conceived as oriented line segments, which are measurable entities: they support a norm function $|v|$ that assigns to each vector a magnitude ≥ 0 , thereby introducing the geometric dimension of distance into the derivation. Moreover, for every pair of points (p, q) belonging to the Ground, there must exist some vector (w) in the vector space V that organizes or relates them. In this way, [Loc] enriches the topological material supplied by [Reg] with the geometric properties necessary for the interpretation of place.

In sum, the assumptions of Nanosyntax, together with the minimal architecture of the spatial domain outlined above, provide the theoretical foundation for the analysis that follows. By decomposing spatial structure into hierarchically ordered terminals, this framework allows us to capture the fine-grained mechanisms underlying the polysemy, synonymy, and competition observed in the preposition *em*. In the next section, we build on these premises to show how the interplay between structural decomposition and lexicalization principles can account for the diverse interpretive possibilities associated with this preposition.

3. An Analysis of the Preposition *em* through the Lens of Nanosyntax

The phenomena investigated in this article are the high degree of polysemy of the preposition *em* and its apparent synonymy with complex prepositional phrases such as *em cima de* ('on top of') and *dentro de* ('inside of'), which can be viewed as a case of competition within the computational system. The discussion therefore centers on the relations among the following sentences, presented in the introduction of the article and repeated below:

- (19) O gato está no armário.
 the.masc cat be.PRS.3SG in.the.masc cupboard
 'The cat is in the cupboard.'
- (20) O gato está dentro do armário.
 the.masc cat be.PRS.3SG inside of.the.masc cupboard
 'The cat is inside the cupboard.'
- (21) O gato está embaixo do armário.
 the.M cat be.PRS.3SG under of.the.M cupboard
 'The cat is under the cupboard.'
- (22) O gato está em cima do armário.
 the.M cat be.PRS.3SG on.top of.the.M cupboard
 'The cat is on top of the cupboard.'

The sentence in (19) may have at least three interpretations, represented by the complex phrases *dentro de* ('inside of'), *embaixo de* ('underneath'), and *em cima de* ('on top of'). In the literature, the number of meanings associated with the preposition *em* is even broader: Oliveira and GoodGod suggest that *em* carries four distinct meanings, while Rodrigues et al. (2020) propose as many as eight senses related to this predicate. Descriptively, the fact is that the preposition *em* seems to be linked to a complex semantics, which allows for multiple interpretations of sentences containing PPs headed by this item. The question that follows from this observation is: which formal features can be taken as the source of this polysemy? And

further, we may ask: in what precise ways do these features contribute to the spatial interpretations associated with *em*? More specifically, to what extent can the attested meanings be derived directly from the formal features, and are all interpretations necessarily accounted for in this way? This second question is particularly important when we consider that the same range of interpretations available in (20) “o gato está no armário” (‘the cat is in the cupboard’) is not observed in a sentence such as (23) below, even though the formal architecture of the clause is essentially the same: definite count nouns in Figure and Ground positions, a stative verb, and the preposition *em*.

- (23) O envelope está na gaveta
 the.masc envelope be.PRS.3SG em-the.fem drawer
 ‘The envelope is in the drawer.’

For a sentence such as “o envelope está na gaveta” (‘the envelope is in the drawer’) to be true, only one scenario is possible: the container interpretation described by Ilari et al. (2015), or what Oliveira and GoodGod (2018) term an inclusion relation. If the envelope were merely adjacent to the drawer or placed on the table containing it, sentence (23) would not be felicitous. These data indicate that the preposition *em* does not, on its own, account for the full range of interpretive possibilities; rather, certain readings emerge from the interaction between the arguments Figure and Ground. Crucially, even a minimal change in the structure of (23) yields a configuration in which both the adjacency and inclusion readings become available. For instance, the sentence below can truthfully describe two distinct scenarios: either when the key is located inside the drawer (inclusion) or when it is positioned in its lock (adjacency).

- (24) A chave está na gaveta
 the.fem key be.PRS.3SG em-the.fem drawer
 ‘The key is in the drawer.’

Therefore, in addition to identifying which formal features are encoded in the preposition, it is crucial to discuss how the physical properties of objects in the world influence the range of interpretations available for sentences containing *em*. With these considerations in mind, we now turn to the first stage of our analysis of the preposition *em* within the theoretical framework of Nanosyntax. The starting point is the observation that lexical items store information constructed in other components of grammar: phonology, syntax, and encyclopedic

knowledge, often referred to as CONCEPT. The question, then, is what the lexical entry associated with *em* might look like.

As previously described, two functional heads in the *f-seq* are central to the construction of spatial relations: [Reg], which builds a region out of an individual, and [Loc], which organizes the spatial points contained within that region. As a working hypothesis, however, we can assume that additional features participate in this configuration and are lexicalized by *em*. Oliveira and GoodGod (2018), Rodrigues, Santos, and Lopes (2017), and Rodrigues et al. (2020) point to a set of topological notions (adjacency, contact, and inclusion) that seem particularly relevant from the perspective of syntactic composition. These three properties recur across the multiple interpretations associated with *em* and can be ordered in a relational hierarchy of proximity between Figure and Ground: adjacency does not necessarily entail physical contact with the boundaries of the Ground, whereas inclusion implies contact at all relevant points. This distribution provides empirical motivation to treat them as independent features within the universal functional sequence.

In Nanosyntax (Starke, 2009), as in Cartography (Cinque & Rizzi, 2010), the existence of a syntactic head must be supported by independent morphophonological evidence. In English, such evidence is clear: distinct prepositions lexicalize each of these notions, with *in* encoding inclusion, *on* encoding contact, and *at* encoding contact and adjacency. A similar pattern emerges in Tsez (Comrie & Polinsky, 1998), where dedicated morphemes are associated with the essive case, interpreted as temporary spatial location, for each of these notions: *-ā* marks inclusion, *-x(o)* and *-de* mark proximity/adjacency, and *-q(o)* conveys contact. In Tsez, additional spatial meanings such as path can be derived compositionally from these morphemes. For example, a marker of movement away from a source, corresponding to the ablative case, is morphologically constructed on top of these spatial markers, further confirming their status as independent building blocks in the functional sequence.

| | Place (essive) | Path (ablative) |
|-----------|-------------------|--------------------|
| Inclusion | -ā | -āy |
| Contact | -q(o) | -q -āy |
| Adjacency | -x(o) | -x-āy |

Table 1. Morphological contenance relations in Tsez (Comrie & Polinsky, 1998, p. 7)

Beyond the morphological observation that these three topological properties, that is, adjacency, contact, and inclusion, find direct lexical counterparts in unrelated languages, there is also a compelling semantic argument for their place in the *f-seq*, grounded in the idea that features in the hierarchy are privative (Cortiula, 2023). Inclusion appears to necessarily involve contact, while contact is closely tied to adjacency, which in turn seems to be the most basic locative terminal. Adjacency requires only that the Figure and the Ground be near one another, without any necessary point-to-point connection between their regions. Once such a connection is established, however, the relation is no longer mere adjacency but a distinct and more complex property, contact. This decomposition is supported by topological reasoning: adjacency presupposes only proximity, whereas contact requires this proximity to be realized specifically at a boundary, thereby converting proximity into connection. Inclusion, in turn, is even more complex, as it involves contact between Figure and Ground at all relevant points, as stated above. In this sense, inclusion presupposes both adjacency and contact. From these considerations, the following hierarchy of terminals emerges:

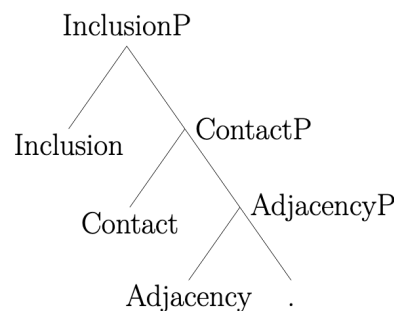


Figure 9. Topological features for the locative domain

Building on the works of Ilari et al. (2015), Oliveira and GoodGod (2018), Rodrigues, Santos, and Lopes (2018), and Rodrigues et al. (2020), it is possible to outline an articulated hierarchy for topological notions that appear crucial in the description of lexical items encoding spatial relations. The next step is to determine the appropriate height of these elements in the *f-seq*, in a way that is consistent with the already established terminals [Reg] and [Loc]. Once again, the argument can be framed in terms of the semantic properties of the spatial hierarchy. Since inclusion, contact, and adjacency are topological notions (independent of scalarity or measurement) it is natural to situate these terminals between [Reg] and [Loc]. The rationale is that [Reg] provides an unstructured set of spatial points, upon which topological notions can be defined, while [Loc] organizes these points and their relations into a vector space. On this basis,

the full hierarchy for the locative portion of the spatial domain can now be proposed as in Figure 10. Below we provide the interpretation of [Adjacency] (the points of the Figure are not the same as the points of the Ground), [Contact] (some points are shared between Figure and Ground) and [Inclusion] (the points of Figure and Ground coincide).

- (25) a. $[[Adjacency]] = \lambda x. \exists y [Adj(x, y) \leftrightarrow \forall p \in Ground(p_x \neq p_y)]$
- b. $[[Contact]] = \lambda x. \exists y [Cont(x, y) \leftrightarrow Adj(x, y) \wedge \forall p \in Ground(p_x \cap p_y = 1)]$
- c. $[[Inclusion]] = \lambda x. \exists y [Incl(x, y) \leftrightarrow Cont(x, y) \wedge \forall p \in Ground(p_x = p_y)]$

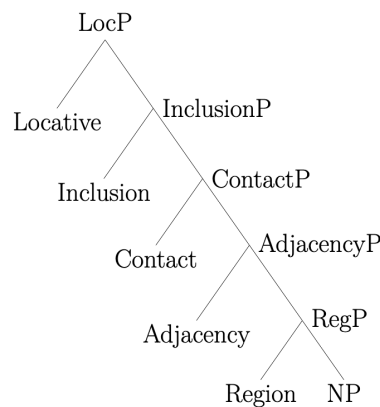


Figure 10 The Nanosyntax of static space

Given that this represents the maximal decomposition of the locative spatial domain, the next step is to determine which features of this structure are actually stored by the preposition *em*. Our hypothesis is that the structure stored by *em* goes from the adjacency terminal [Adj] up to the locative phrase [LocP], such that the region head [Reg] is associated exclusively with the NP that is the complement of the preposition, namely the Ground. Based on this, we propose the following lexical entry for *em*:

- (26) $\langle /\text{ẽj}/ \Leftrightarrow$

```

graph TD
    LocP --> Locative
    LocP --> InclusionP
    InclusionP --> Inclusion
    InclusionP --> ContactP
    ContactP --> Contact
    ContactP --> AdjacencyP
    AdjacencyP --> Adjacency
    
```

 \Leftrightarrow RELATION OF TOPOLOGY >

With this lexical entry, the prediction is that the preposition *em* can appear in at least four syntactic contexts: [Adj], by virtue of the Anchor Condition, which requires a match between the lowest feature in the lexical entry and the target structure, [Cont[Adj]], [Inc[Cont[Adj]]], and [Loc[Inc[Cont[Adj]]]]. By the Superset Effect, therefore, *em* can lexicalize any abstract structure compatible with the constituents it contains. This accounts for the three main spatial interpretations associated with *em*: when the speaker lexicalizes a structure in which the relevant relation is mere proximity between Figure and Ground, *em* surfaces by contributing only the adjacency feature [Adj]; when the intended interpretation is that of contact between Figure and Ground, *em* lexicalizes the structure [Cont[Adj]]; and when the relevant interpretation is inclusion of the Figure within the Ground’s region, *em* is again a suitable candidate, as it contains [Inc[Cont[Adj]]].

This proposal successfully explains the polysemy of *em*, but it also raises a theoretical issue that requires further consideration. Specifically, if [Loc] is included as the highest feature in the lexical entry, strict locality would predict that it cannot be lexicalized independently of the lower features. This creates a tension within the nanosyntactic model, since lexical entries are expected to respect the hierarchy of features while still allowing flexibility in the spell-out. The challenge, then, is to reconcile the explanatory power of this entry with the theoretical constraints of locality. According to Caha (2025, p. 283), “since lexical entries may link any well-formed structure to a phonological representation, it is possible to also consider entries which lexicalize structures that contain movement (Blix 2022)”. Thus, we can assume, through the Lexicalization Algorithm, that the configuration of terminals associated with *em* might be slightly different. Specifically, when [Loc] is built, it can trigger movement of its complement for purposes of lexicalization, yielding the following complex structure (Figure 11). The lexical entry of *em*, then, can be assumed to contain a complex specifier/complex left branch (cf. Caha et al., 2025), which allows [Loc] to be mobilized in the derivation independently of the lower terminals, since now we have two sides to me matched against the target structure.

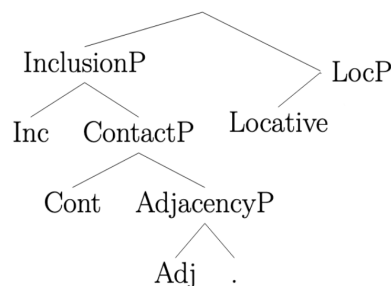
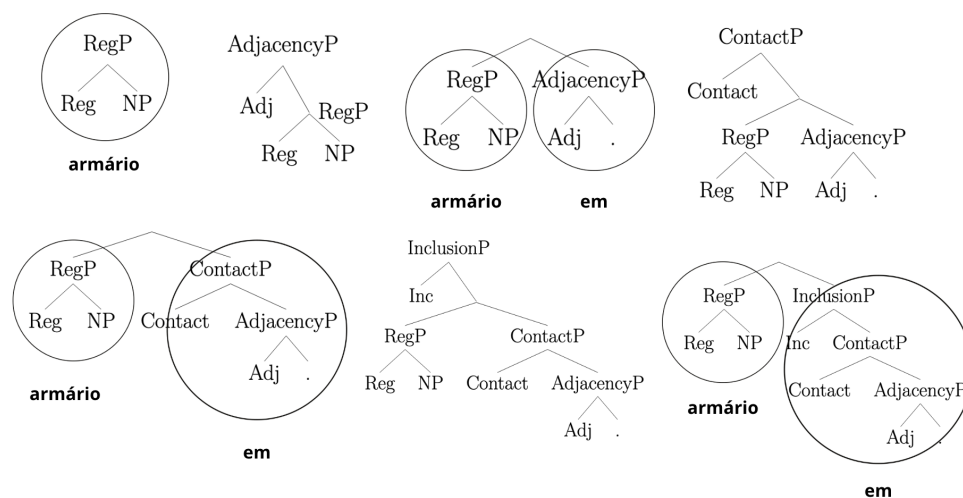


Figure 11. Updated structure of *em* with a complex left branch

This proposal has an interesting consequence: it predicts that the topological heads need not be restricted to strictly locative contexts. In other words, once [Loc] can be spelled out without requiring the simultaneous lexicalization of [Adj], [Cont], or [Inc], the same structural ingredients that define spatial relations may surface in other domains, extending the interpretive potential of *em* beyond purely spatial readings.

Given that, the derivation of a sentence such as “o gato está no armário” (“the cat is in the cupboard”) allows for three distinct readings of the preposition *em* because all three topological features are available in the derivation. If the structure is lexicalized only at [Adjacency] phrase, the interpretation then corresponds to proximity/adjacency between Figure and Ground. If the derivation proceeds through [Cont[Adj]], the resulting reading is one of contact. Finally, when the derivation reaches [Inc[Cont[Adj]]], the interpretation is that of full inclusion of the Figure within the Ground. To sum up, if the inclusion feature is matched against the target structure, by default all the other meanings are available. Thus, the polysemy of *em* in this context follows from the incremental activation of the three topological features in the functional sequence. The detailed stages of this derivation are provided below (Figure 12), considering that the Ground was already build.

The derivation follows the Lexicalization Algorithm, whereby at each Merge-F the structure is checked against the lexicon for a compatible item. If no such item is found, the derivation proceeds through the steps outlined in (15), involving spec-to-spec and comp-to-spec movements. Ultimately, through the application of these two types of movement, the maximal locative spatial structure can be lexicalized by *em*, which constitutes a perfect match, both regarding the features and the shape needed.



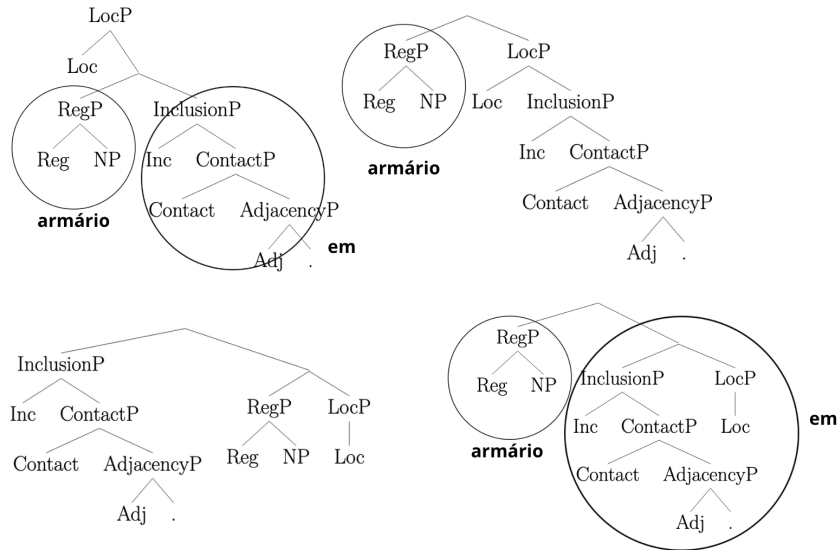


Figure 12. Lexicalization of “o gato está no armário”

The next question to be addressed is the fact that not all interpretations are available all the time. For a sentence such as “o envelope está na gaveta” (“the envelope is in the drawer”) the same syntax as above applies, but by contrast, in a sentence like “o prato está na mesa” (“the plate is on the table”), the only interpretation that emerges is the one of contact. We argue that this difference reflects the encyclopedic knowledge: the Ground *mesa* (“table”) is conceptualized as a surface entity, and therefore only the contact reading is licensed, while the Ground *gaveta* (“drawer”) is conceptualized as a volume, thus allowing the inclusion interpretation. Crucially, the choice of the preferred interpretation does not follow from the formal features introduced in the syntactic computation itself, but rather from the conceptual properties attributed to the entities Figure and Ground. In other words, it is the physical format assigned to these entities, at the level of encyclopedic knowledge, that determines which interpretation of *em* is available. The CONCEPTs relevant to the spatial domain are essentially VOLUME and SURFACE. Thus, a sentence such as “o prato está na mesa” (“the plate is on the table”) guide the construction of [Contact[Adjacency]] due to the SURFACE concept stored in the lexical entry of *mesa*.

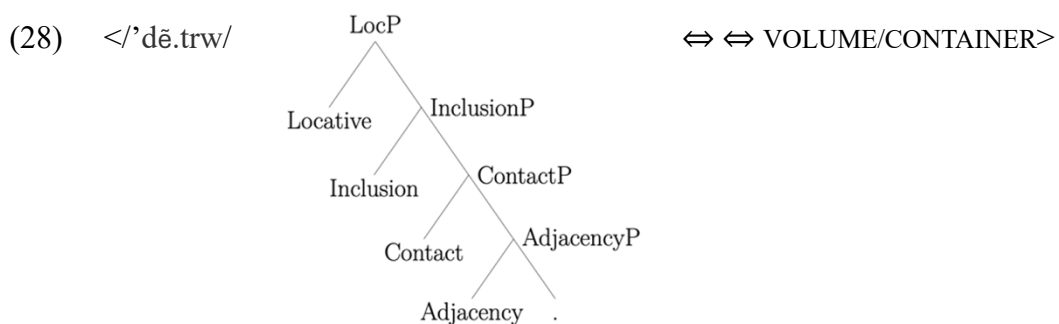
$$(27) \quad </'me.za/ \Leftrightarrow \begin{array}{c} \text{CountP} \\ \swarrow \quad \searrow \\ \text{Count} \quad \text{NP} \end{array} \Leftrightarrow \text{SURFACE}>$$

The constraint of the contact reading, therefore, arises from the presence of the conceptual feature SURFACE encoded in the lexical entry of the Ground. Since the contact

reading builds on the adjacency terminal, this relation also helps to explain the so-called “functional” uses of the preposition *em*. An example is “Maria está sentada na mesa” (“Maria is sitting at the table”), where the relevant interpretation is not one of strict physical contact with the Ground but rather one of proximity that licenses a functional association. In conclusion, volumetric Grounds allow the inclusion interpretation, while surface ground block this reading, allowing contact and adjacency. Since lexical items supply the syntactic structure with both phonology and encyclopedic content during the process of lexicalization, we can now identify the precise *locus* of the “interpretive restriction”.

What initially seemed arbitrary is, in fact, a consequence of the interaction between the syntactic features targeted by lexicalization and the conceptual layer (CONCEPT) that is activated at the interface. This perspective clarifies why certain readings are systematically licensed or blocked, depending not only on the formal features encoded in the functional sequence but also on the conceptual properties contributed by the Ground.

The final Issue to be addressed concerns the synonymy between the preposition *em* and complex prepositional phrases (PPs), which can be understood as a case of competition for the lexicalization of the same portion of the spatial *f-seq*. Considering that the relevant features are Locative, Inclusion, Contact, and Adjacency, the only direct competition appears to be between *em* and *dentro de*, since *embaixo de* and *em cima de* lexicalize a different type of information, more specific and associated with the presence of a vertical axis. *Dentro de*, on the other hand, seems to encode precisely the notion of Inclusion. If this assumption is correct, the structure underlying *dentro* would be equivalent to that of inclusive *em*.



This competition between *em* and complex prepositional phrases such as *dentro de* illustrates the intricate interface between syntax, formal semantics, and pragmatics. From a syntactic perspective, both items are capable of lexicalizing overlapping portions of the spatial *f-seq*, with *dentro* arguably encoding the feature [Inclusion] in a more specific way, because it

doesn't need successive movements in order to lexicalize the structure. Computationally, this raises the expectation that *dentro* should always prevail, since specificity generally outcompetes underspecified forms. Yet, the fact that *em* remains highly productive suggests that the choice is not determined solely by formal features. At the pragmatic level, the use of *dentro* reflects the speaker's intention to assert precision and certainty about the Figure–Ground relation, while the use of *em* leaves room for underspecification, allowing interpretation to be negotiated through context. It is important to highlight that this context is also quite limited, based only on the conceptual properties of Figure and Ground, such as VOLUME and SURFACE.

The persistence of both options in the grammar, therefore, highlights how Nanosyntactic principles of lexicalization interact with encyclopedic knowledge and pragmatic reasoning, offering a fuller explanation for the polysemy and apparent synonymy of the Brazilian Portuguese prepositional system. Therefore, the preposition *em* can be taken as less specific than its complex counterparts (*em cima de*, *embaixo de*, *dentro de*) because it does not involve any axial information. In this sense, *em* can indeed be considered a polysemous item; however, its apparent complexity does not resist formal analysis.

By decomposing spatial relations into independent features within the functional sequence, we are able to explain why *em* gives rise to multiple interpretations while maintaining a principled connection among them. The nanosyntactic model, in dialogue with formal semantics, captures the incremental contribution of features such as [Adjacency], [Contact], and [Inclusion], while encyclopedic knowledge accounts for the restrictions imposed by the physical and conceptual properties of Figure and Ground. Thus, the high degree of polysemy associated with *em* is not a case of unstructured lexical ambiguity, but rather the predictable outcome of the interaction between syntax, semantics, conceptual knowledge and pragmatics.

4. Final remarks

The analysis developed in this paper has shown that the high degree of polysemy associated with the Brazilian Portuguese preposition *em* can be systematically accounted for within a nanosyntactic framework enriched by insights from formal semantics. By treating adjacency, contact, and inclusion as independent features in the functional sequence, we have demonstrated that the major spatial interpretations of *em* are not arbitrary, but rather the predictable outcome of incremental feature composition and the Superset Effect.

At the same time, the study highlights the indispensable role of encyclopedic knowledge in shaping interpretation. The distinction between volumetric and surface Grounds, for instance,

illustrates how conceptual properties of Figure and Ground restrict which features are accessible in a given derivation. This interplay between formal computation and conceptual content is further reinforced by cases of competition between *em* and complex prepositional phrases such as *dentro de*, where pragmatic considerations determine whether speakers opt for a more specific or a less committed expression of spatial relations.

Taken together, these findings suggest that the apparent synonymy and polysemy of *em* are not signs of lexical idiosyncrasy, but rather the result of principled interactions between nanosyntactic structure, semantic composition, and encyclopedic knowledge. More broadly, the case of *em* demonstrates how nanosyntax can offer a powerful tool for understanding prepositional systems, bridging formal features and conceptual interpretation in a unified account.

5. References

BASSETTO, Bruno Fregni. **Elementos de Filologia Românica: História Interna das Línguas Românicas**, v. 2. São Paulo: Editora da Universidade de São Paulo, 2010.

BAUNAZ, Lena et al. (Ed.). **Exploring nanosyntax**. Oxford University Press, 2018.

BENNETT, Brandon. Standpoint semantics: a framework for formalising the variable meaning of vague terms. **Understanding Vagueness—Logical, Philosophical and Linguistic Perspectives**. College Publications, 2011.

CAHA, Pavel. A default theory of default case. **Glossa: a journal of general linguistics**, v. 9, n. 1, 2024.

CAHA, P. et al. Nanosyntax: state of the art and recent developments. In: CAHA, P. et al. **Nanosyntax and the lexicalization algorithm**. Oxford University press, 2025.

CAHA, Pavel. A local analysis of an apparent nonlocal allomorphy in Tamil. In: In: CAHA, P. et al. **Nanosyntax and the lexicalization algorithm**. Oxford University press, 2025.

CINQUE, Guglielmo; RIZZI, Luigi. The Cartography of Syntactic Structures. In: HEINE, B. & NARROG, H. (ed.). **The Oxford Handbook of Linguistic Analysis**. New York, Oxford University Press, p. 51-65, 2010.

COMRIE, Bernard; POLINSKY, Maria; RAJABOV, Ramazan. Tsezian languages. **The languages of the Caucasus**. London: Curzon, 1998.

CORTIULA, Maria. **The nanosyntax of Friulian verbs**. An analysis of the present and past in Tualis Friulian. PhD diss., Masaryk University, 2023.

FERREIRA, Thayse Leticia. Preposições espaciais como predicados instanciadores de kind: sobre a existência de entidades espaciais. **ReVEL**, edição especial, v. 19, n. 18, 2021.

ILARI, R. et al. **As preposições**. In: ILARI, R.; NEVES, M. H. M. (Orgs.) Gramática do português culto falado no Brasil. Vol. II - Classes de Palavras e Processos de construção. São Paulo: Contexto, 2015.

LANDAU, Barbara; JACKENDOFF, Ray. Whence and whither in spatial language and spatial cognition?. **Behavioral and brain sciences**, v. 16, n. 2, p. 255-265, 1993.

OLIVEIRA, Aparecida de Araújo; VIEIRA GOODGOD, Pedro Ivo. Validade empírica das redes de polissemia para o significado preposicional. **Revista de estudos da linguagem**, v. 26, n. 1, 2018.

PANTCHEVA, Marina Blagoeva. **Decomposing path**: The nanosyntax of directional expressions. PhD Thesis. University of Tromsø, 2011.

PARTEE, Barbara H. Formal semantics: Origins, issues, early impact. **Baltic International Yearbook of Cognition, Logic and Communication**, v. 6, n. 1, p. 13, 2011.

PORTNER, Paul H. **What is meaning?: Fundamentals of formal semantics**. John Wiley & Sons, 2005.

RODRIGUES, Edilson; SANTOS, Paulo E.; LOPES, Marcos. Pinning down polysemy: A formalisation for a Brazilian Portuguese preposition. **Cognitive systems research**, v. 41, p. 84-92, 2017.

RODRIGUES, Edilson J. et al. Standpoint semantics for polysemy in spatial prepositions. **Journal of Logic and Computation**, v. 30, n. 2, p. 635-661, 2020.

SANTOS, Milton. **A natureza do espaço: técnica e tempo, razão e emoção**. São Paulo: Edusp, 2006.

STARKE, Michal. Nanosyntax: A short primer to a new approach to language. **Nordlyd**, v. 36, n.1, p. 1-6, 2009.

STARKE, Michal; CORTIULA, Maria. Bearing the Brazilian cross. **Revista Virtual de Estudos da Linguagem-ReVEL**, v. 19, n. 18, p. 9-35, 2021.

SVENONIUS, P. Structural decomposition of spatial adpositions. **Work. Papers Tromsø**, Nor., 2012.

TALMY, Leonard. How language structures space. In: **Spatial orientation: Theory, research, and application**. Boston, MA: Springer US, 1983. p. 225-282.

URSINI, Francesco-Alessio. How space gets into language: a novel approach. In: **Conference of the Australasian Society for Cognitive Science (9th: 2009)**. Macquarie Centre for Cognitive Science, 2010. p. 348-355.

VANDELOISE, Claude. Are there spatial prepositions?. In: **Space in Languages**. John Benjamins Publishing Company, 2006. p. 139-154.

ZWARTS, Joost; WINTER, Yoad. Vector space semantics: A model-theoretic analysis of locative prepositions. **Journal of logic, language and information**, v. 9, n. 2, p. 169-211, 2000.

This preprint was submitted under the following conditions:

- The authors declare that the necessary Terms of Free and Informed Consent of participants or patients in the research were obtained and are described in the manuscript, when applicable.
- The authors declare that the preparation of the manuscript followed the ethical norms of scientific communication.
- The authors declare that they are aware that they are solely responsible for the content of the preprint and that the deposit in SciELO Preprints does not mean any commitment on the part of SciELO, except its preservation and dissemination.
- The authors declare that the data, applications, and other content underlying the manuscript are referenced.
- The deposited manuscript is in PDF format.
- The authors declare that the research that originated the manuscript followed good ethical practices and that the necessary approvals from research ethics committees, when applicable, are described in the manuscript.
- The authors declare that once a manuscript is posted on the SciELO Preprints server, it can only be taken down on request to the SciELO Preprints server Editorial Secretariat, who will post a retraction notice in its place.
- The authors agree that the approved manuscript will be made available under a [Creative Commons CC-BY](#) license.
- The submitting author declares that the contributions of all authors and conflict of interest statement are included explicitly and in specific sections of the manuscript.
- The authors declare that the manuscript was not deposited and/or previously made available on another preprint server or published by a journal.
- If the manuscript is being reviewed or being prepared for publishing but not yet published by a journal, the authors declare that they have received authorization from the journal to make this deposit.
- The submitting author declares that all authors of the manuscript agree with the submission to SciELO Preprints.