



UNIVERSITY OF HELSINKI



<https://helda.helsinki.fi>

Helda

On the Origins of the Shell Noun Construction in Portuguese

Granvik, Anton

2024-10

Granvik, A 2024, On the Origins of the Shell Noun Construction in Portuguese. in M Calderón Campos & G Vaamonde (eds), Linguistic Corpora and Big Data in Spanish and Portuguese. Humanidades Digitales y Big Data en Iberoamérica / Digital Humanities and Big Data in Ibero-America, vol. 4, de Gruyter, Berlin, pp. 179-206. <https://doi.org/10.1515/9783110781465-008>

<http://hdl.handle.net/10138/586999>

10.1515/9783110781465-008

cc_by_nc_nd

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Anton Granvik

On the Origins of the Shell Noun Construction in Portuguese

1 Introduction

Shell nouns (cf. Schmid 2000, 2018) are abstract nouns which serve the important function of encapsulating, or classifying, other, often more complex elements present in the discourse as simple concepts such as *idea*, *conclusion*, *fact*. Flowerdew & Forest (2015: 46) describe them as “open-class vocabulary items with discourse signalling functions, items which [. . .] exhibit a requirement for specification in context and which [. . .] act as discourse signposts.” Sanches Duran & Volpe Nunes (2019: §1) add that “In encapsulation there is an encapsulating word (pronoun or noun phrase) that is used in place of an entire sentence encapsulated by it.”¹

As shown in (1), *certeza* and *facto* make reference to the following passages, *que . . . a parte social iria continuar comigo* and *eu ser uma pessoa preocupada*, respectively. Importantly, the information presented in these passages can be considered a *certainty* and a *fact*; and, without the following passages we do not know what the certainty and the fact are.

- (1) Tenho a **certeza** de que se deixasse de trabalhar em televisão, ou se me reformasse da televisão, a parte social iria continuar comigo, é indissociável do **facto** de eu ser uma pessoa preocupada com as pegadas que deixo aqui quando me for embora; de ser mulher, portanto confrontada com as desigualdades que ainda existem cá e lá fora. (CP-NOW, <https://online.sapo.pt/641931>)

‘I’m sure that if I stopped working in television, or if I retired from television, the social part would remain with me, it’s inseparable from the fact that I’m a person who is concerned about the footprints I leave here when I leave; that I’m a woman, and therefore confronted with the inequalities that still exist here and there.’

In present-day Portuguese, shell nouns typically appear in three constructional formats, each with two variants, as shown in (2)–(4):

¹ My translation. The original reads: “No encapsulamento há uma palavra encapsuladora (pronome ou sintagma nominal) que é usada no lugar de toda uma oração por ela encapsulada.”

- (2) a. **(det) N (adj) *ser que* + oração**
 A **idéia apresentada é que** as pessoas dediquem um dia por mês para trabalhos voluntários.²
 ‘The idea is for people to dedicate one day a month to voluntary work.’
- b. **(det) N (adj) *ser* + (*de*) infinitive**
 A minha parte eu vou pagar e posso assegurar que **a idéia** de todos é **fazer** o mesmo
 ‘I’ll pay my share and I can assure you that everyone’s idea is to do the same’
- (3) a. **(det) N (adj) *de* + infinitive**
 Com[o] surgiu **a idéia de fazer** esse filme de guerra?
 ‘How did the idea of making this war film come about?’
- b. **(det) N (adj) *de* + infinitivo pessoal**
 –Ah! meu pai! **essa idéia de irmos** para tão longe, sem esperança de um dia podermos voltar...
 ‘Oh, my father, this idea of going so far away, with no hope of ever coming back. . .’
- (4) a. **(det) N (adj) *de que* + oração**
 –O senhor fala que o PDT quer espaço político, mas fica sempre **a idéia de que** espaço significa cargo.
 ‘You say that the PDT wants political space, but there’s always the idea that space means office.’
- b. **(det) N (adj) *que* + oração**
 Por isso se criou **a idéia que** uma economia saudável é aquela que cria megasuperávits, o que não é correto...
 ‘That’s why the idea has been created that a healthy economy is one that creates mega surpluses, which is not correct...’

In the medieval period, however, the available constructions were fewer, including only those in (2a), (3a), (3b) and (4b), as in (2c) to (4c):

- (2) c. Ca **não é razão que** aquele que recebe a entrega mingue nada do seu.
 ‘Because it is not right [reason] that the one who receives the gift shall be without all that is his’

² If not marked otherwise, all the examples are taken from Davies & Ferreira’s (2006-) *Corpus de Português* (CP).

- (3) c. Ca a Sennor que o atan ben dá non á ome **razon de lle furtar nen de roubar-ll'** o seu nen llo filar...‘Because a Lord who gives so well a man does not have the right [reason] to rob him or to take it from him. . .’
- (4) c. E don Vaasco fernandez Maestre do Tempre e os ffreyres dessa Ordim da outra **per Razom que** o dicto procurador delRey dizia que o dicto Maestre e ffreyres tragiã ascondudas...‘And Don Vaasco Fernandez, Master of the Templars, and the freedmen of that Order of the other, for the reason that the said procurator of the King said that the said Master and freedmen were carrying X hidden. . .’

Given these differences in construction format between the medieval period and the present-day language, in this paper I propose to investigate and describe

- A. how the shell-noun construction in Portuguese arises and evolves over the centuries;
- B. how the use of the constructional formats develops over time (*N ser que* + oração, *N de* + infinitive, *N (de) que* + clause);
- C. what the relationship is between the function (shell-noun use or not?), the form (construction) and usage context of these constructions.

To reach these goals the paper focuses on nine shell-nouns which are representative of the above constructions in three different periods: the 13th and 14th centuries, the 16th and 17th centuries and the 19th and 20th centuries. The analysis of the nine nouns, i.e. *caso*, *facto*, *ideia*, *mercê*, *questão*, *razão*, *sinal*, *temor* and *vontade*, is based on data retrieved from *Corpus do Português* (Davies & Ferreira 2006-), a large diachronic reference corpus of Portuguese. In line with the interests of Digital Humanities, the analysis is based on different quantitative analyses, such as collocation analysis (Stefanowitsch & Gries 2003) and logistic regression (Gries 2013, Winter 2020).

The interest of investigating the history and early development of the shell-nouns in Portuguese is further motivated by the fact that previous studies have mainly focused on the present-day language. Secondly, observations regarding these nouns in other languages suggest that their use is connected to the rise of informative and journalistic texts in the 17th century (Borreguero 2018, Borreguero Zuloaga & Octavio de Toledo 2007). Previous studies on Spanish, however, have shown that some nouns are used as shell nouns as early as the 13th century

(Granvik 2019).³ It is thus interesting to see whether this is the case in Portuguese as well.

The paper is structured as follows. Section 2 introduces the concept of shell nouns and their encapsulating function. What I call the shell-noun construction is also presented and defined. The corpus and methods used in the analysis are presented in Section 3. Section 4 focuses on analyzing the data first from a diachronic perspective, and then from the perspective of determining which criteria best explain the difference between encapsulating and non-encapsulating uses. Section 5 concludes.

2 Theoretical Background

The notion of *shell noun* was coined by Schmid in the late 20th century (Schmid 2018: 110–111) and is used to refer to what he calls a functionally defined category: “A noun is turned into a shell noun when a speaker decides to use it in a shell-content complex in the service of certain aims. Shell-nounhood is thus a functional property” (Schmid 2000: 13). Shell nouns are thus not a category a noun either belongs to or not, but rather many, often abstract nouns are turned into, or function as, shell nouns in certain uses, i.e. when they are used to encapsulate, classify, or signal another discourse element as what the designate, such as a *fact*, an *idea*, a *conclusion*, etc.

Schmid’s (2000) monograph on shell nouns approaches this group of nouns from three perspectives, highlighting their semantic, cognitive, and textual function. According to Schmid (2000: 14), the semantic function corresponds to “characterizing and perspectivizing complex chunks of information”, the cognitive function to “temporary concept-formation (encapsulate these complex chunks . . . in temporary nominal concepts)” and the textual function to “linking these nominal concepts with clauses.”

The text or discourse function of shell nouns had, of course, been noticed way before Schmid’s (1998, 1999, 2000) work. Within the discourse tradition these referring, signalling, linking, encapsulating, and characterizing functions (cf. Schmid 2018: 111) all refer to the fact that the nouns “act as labels, also known as encapsulators, [and] being compactors, they are responsible for changing or linking topics

³ See also Schmid & Mantlik (2015) for some observations regarding early shell-noun uses in English, dating from, at least, the 14th century.

and also contribute to preserving textual continuity by introducing new information within old information” (Paredes Silva & Bezerra 2013: 219–220).⁴

The essence of these nouns is the concept of encapsulation which, in the words of Guimarães (2011:30) “suggests a process of linguistic compacting of events around a lexical item ... This means that a word class summarizes content to be enunciated.” For Andrade (2019: 180), encapsulating nouns imply “a choice capable of summarising information present in the context and, at the same time, transforming it into a new object of discourse.”⁵ Flowerdew & Forest (2015: 48) also take encapsulation, i.e. “equative encapsulation with lexical specifics provided elsewhere in the text”, as the defining “criterion for SN membership”. Schmid (2000: 21) acknowledges that the contextual identification of what the shell noun encapsulates is central, i.e. what he calls *experiential identity*, the notion that the shell noun and the shell content express ideas about the same thing.

This is what happens in (5), where there are two nouns acting as shells, i.e. *conceção* and *ideia*. *Conceção* makes anaphoric reference to the immediately preceding textual segment —*Adquirir novos conhecimentos* ‘acquiring new knowledge’— whereas *ideia* encapsulates (by cataphoric reference) the following complete sentence —*o aprender é entendido como um meio para saber novos conteúdos* ‘learning is understood as a means of acquiring. . .’.

- (5) Como podemos observar, a *conceção* mais referida diz respeito ao Adquirir novos conhecimentos, sendo enunciada por 49,1% dos participantes. Esta **conceção** refere-se à *ideia* de que o aprender é entendido como um meio para saber novos conteúdos. (Carvalho e Ribeiro 2017: 210)

‘As we can see, the most mentioned conception relates to Acquiring new knowledge, which was mentioned by 49.1 per cent of the participants. This conception refers to the idea that learning is understood as a means of acquiring new content.’

Note that, what the presence of an encapsulator as *ideia* adds to the sentences in (5), is a characterisation of the proposition “learning is understood as a means of acquiring new content” as an ‘idea’, i.e. as a kind of mental entity. But we could

4 My translation. The original reads: “Os SNs que atuam como rótulos, também chamados encapsuladores, sendo compactadores, são responsáveis por mudar ou ligar os tópicos e contribuir, também, na preservação da continuidade textual ao introduzir as informações novas dentro das velhas” (Paredes Silva & Bezerra 2013: 219–220).

5 My translation. The original reads: “O encapsulador é um processo referencial que merece destaque, pois se apresenta como uma escolha capaz de sumarizar informações presentes no contexto e, ao mesmo tempo, transformá-las em novo objeto de discurso.”

substitute *ideia* by a noun such as *facto* ‘fact’, and the overall meaning would not be very much altered, except that we would now be talking about something conceived of as a fact instead of as an idea. The function of *ideia* in (5) corresponds perfectly to Abad Serna’s (2015: 230) observation that “nouns used as encapsulators have the singularity of synthesising the reference of the textual segment to which they refer”.⁶

According to Schmid (2000: 21), the encapsulating function can be considered a prototypical category, the core of which is the experiential identity relation (Schmid 2000: 21; cf. López Samaniego 2011: 447; Abad Serna 2015), i.e. “the notion that the shell noun and the shell content express ideas about the same thing”. Returning to example (5), above, this experiential identity can be verified by noting that the clause “learning is understood as a means of acquiring new content” is an *idea*. However, if the core of the category of shell nouns is the notion of experiential identity, and the category is prototypical in nature, it is to be expected that there are less typical and even marginal cases around it. Thus, for Schmid (2000: 25–26), nouns used as part of “expanded predicates”, in combination with functional verbs of the type *to have* and *to do*, “cannot be considered good examples of encapsulating nouns”. Likewise, nouns of temporal or locative meaning, such as *time*, *place*... “are treated as marginal” (Schmid 2000: 26). Examples (6) to (8) illustrate different degrees of typicality, where the notion of “experiential identity” becomes increasingly difficult to maintain:

- (6) Por exemplo, **no caso de serem vistas de longe**, deveria imprimir força e monumentalidade. (CP, 19Ac:Br:Enc)
 ‘For example, in case they were to be seen from a distance, they should be strong and monumental.’
- (7) Que se **fizesse**, porém, grande **caso de juntar com a pregação os exercícios de humildade**. (CP, 15:Lucena:SFXavier)
 ‘However, it would be a good idea to combine preaching with exercises in humility.’
- (8) Também não **faça caso de que lhe faltam fôrças** para a batalha porque isto é soberba e não humildade (CP, 16:Chagas:Cartas)
 ‘Don’t pretend that they lack the strength for battle either, because that’s pride, not humility’

⁶ My translation. The original reads: “los sustantivos empleados como encapsuladores tienen la singularidad de sintetizar la referencia del segmento textual al que remiten.”

In (6), although *no caso de* is a lexicalized expression, it is transparent enough that it is clear that ‘to be seen from a distance’ is what constitutes the *case*, so there is clearly identification and encapsulation. In (7), on the other hand, *caso* is combined with the light-verb *fazer* ‘to do, make’, and the compound predicate *fazer caso* is combined with *de* which appears to maintain some of its sense of ‘aboutness’ (see Granvik 2014). It seems possible to interpret that the *case* is *to combine preaching with exercises*. . . , but this interpretation is clearly less salient than in (6). In (8) the experiential identity relation is even more peripheral, since the same compound predicate *fazer caso de* is used in another sense, that is, ‘to make an issue about’ or, as the translation suggest, ‘to pretend’. The complement clause in (8) thus corresponds more to a direct object than to an appositional clause (see Leonetti 1993, 1999).⁷

When determining whether an instance of an abstract noun constitutes a case of the shell noun use or not, it seems natural to take the notion of experiential identity as a defining criterion. That is, a noun is considered to have an encapsulating function if and only if there is “experiential identity” between the noun and another element present in the immediate context (typically, the complement clause following it). Indeed, this is how shell nouns, and indexical nouns, are often defined (see above; Flowerdew & Francis 2015). So, if the answer to the question “Is N equal to the (complement) clause? (or, is the clause (an) N?) is affirmative, it would count as a shell use.

However, as examples (6) to (9) indicate, the distinction between experiential identity and its absence is not always clear (let alone objective). There is, for example, no biunivocal relation between one or more syntactic constructions (e.g. *N de que* or *N é que*) and the relation of identity. As Samaniego’s (2011: 446) discussion of Leonetti’s (1999) examples reveals, the same noun can be combined with both argumentative (9) and appositive complement clauses (10) in the same syntactic construction:

- (9) La explicación [de que suspendas siempre] es que no te preocupas por entender bien la asignatura,
 ‘The explanation [that you always fail] is that you don’t bother to understand the subject well’

7 According to López Samaniego (2011: 446–447), “the discursive actualisation of the noun only appears in the complement clause when it maintains an appositive relation of identity with the noun”. The alternative is that the complement clause has a predicate function, constituting an argument of the nominal nucleus, reflecting Leonetti’s (1993, 1999) distinction between argumentative and appositive complement clauses.

- (10) La explicación [de que han aumentado los gastos] no es muy convincente.
‘The explanation [that expenses have increased] is not very convincing.’

In (9) there is no identification of the noun *explanation* as “you always fail”, whereas in (10) “that expenses have increases” is the explanation.

On the other hand, consider examples (11) and (12), involving the noun *temor* ‘fear’. In (11), it is difficult, if not impossible, to determine whether the complement clause “as minorias sérvias . . . seriam perseguidas” is what constitutes the ‘fear’ (*temor*), or whether it is the object of the fear, i.e. what the fear is about. Clearly, this is a question of degree, where the semantics of the noun *temor*, which parallels the modal verb *temer* ‘to fear’, is compatible with both an identifying and an object interpretation. In the first interpretation, we would say that the fear is that the Serbian minorities are being persecuted; in the second, we would say that this is what is feared. Compare this to the use of *temor* in the equative construction with *ser* in (12), where the identification of *que os Brasil suspende as exceções . . . as a fear* is made explicit.

- (11) Com a onda de independências a partir de 90, os nacionalistas sérvios disseminaram **o temor de que** as minorias sérvias em as outras repúblicas seriam perseguidas como ocorreu quando sérvios (CP, 19N:Br:Folha)
‘With the wave of independence from the 1990s onwards, Serbian nationalists spread the fear that Serbian minorities in the other republics would be persecuted as happened when Serbs’
- (12) **O grande temor** dos empresários argentinos **é que**, em outubro, o Brasil suspende as exceções às restrições às importações. (CP, 19Or:Br:Intrv:ISP)
‘The great fear of Argentine businesspeople is that in October Brazil will lift the exceptions to import restrictions’

Since the relation of experiential identity between the noun and the complement clause is often a question of degree and may thus be difficult to establish objectively, Granvik (2019) introduced an alternative —formal and therefore objective— approach to the analysis of shell nouns. Starting from Schmid’s (2018: 11) observation that “a limited number of lexico-grammatical patterns are particularly suitable for encoding the link of experiential identity between shell noun and shell content”, Granvik considered all combinations of N with a complement clause (i.e. the constructions included in points b) and c), below), a case of the encapsulating construction. The four grammatical constructions that, according to Schmid greatly facilitate that a noun is used with an encapsulating function are listed in the following:

- a) demonstrative + noun (e.g. *this dilemma...*)
- b) noun followed by modifying clause (*the dilemma + that/infinitive...*)
- c) noun as the subject of an identifying clause with the verb *ser* (*the dilemma is that...*)
- d) neuter demonstrative or neuter personal pronoun as subject and encapsulating noun as predicate (*this is a dilemma*) (following Rodríguez Espiñeira 2015: 658; the examples are mine)

Focusing on the constructions in b) and c) finds support in Schmid's (2018: 115) consideration that these two lexico-syntactic patterns (N *be* + *that*-clause and N + *that*-clause) "stand out from the others by providing the kind of environment in which shell nouns thrive".

I propose, then, that there is an encapsulating construction in the schematic sense given by Construction Grammar (cf. Croft & Cruse 2004; González-García 2012; Traugott & Trousdale 2013; Hilpert 2019). This shell noun construction has four variants, which all involve the presence of a complement clause in either infinitive or finite form: N *que* + clause, N *de que* + clause, N *de* + infinitive and N *ser* + clause (finite or infinitive) (see Schmid 2018: 124–125 for a discussion of the construction status of the encapsulating construction(s)).⁸ As will become evident below, one of the focal points of the analysis is to verify to what degree the formal characteristics of the shell noun construction corresponds to the encapsulating function of the nouns.

3 Data and Methods

The data analyzed in this paper was retrieved from Davies & Ferreira's (2006–) reference corpus of Portuguese called *Corpus do Português* (CP). The diachronic part of this corpus includes roughly 45 million words, ranging from the 13th to the 20th century.

The nouns were selected using a series of criteria. As a starting point I use Granvik's (2017b: 39) century-wise collocation analysis (Stefanowitsch & Gries 2003) of the most frequently attested nouns in the sequences N (*de*) *que*, N *ser que* and N *de* + infinitive.⁹ Second, among the most highly ranked nouns in each cen-

⁸ I thereby exclude the formats involving demonstratives, types a) and d) from consideration in this paper.

⁹ Note that the collocation analysis presented in Granvik (2017b) was performed on data ranging from the 1500s to the 1900s (due to scarcity of data in previous centuries) and focused on the sequences N *que*, N *de que* and N *de* + infinitive. For the 1200s to 1400s, I manually inspected

tury, I selected nouns that fall into different types of nouns, following Schmid's (2000) classification of shell nouns into factual, mental, modal, linguistic, and circumstantial ones. Third, inspecting the nouns and their usage frequencies across the construction formats and the centuries permitted selecting a total of nine nouns, three for each period as shown in the table below.

1200–1399 *mercê* (modal), *razão* (factual), *vontade* (mental)
 1500–1699 *caso* (factual), *sinal* (factual), *temor* (mental)
 1800–1999 *facto* (factual), *ideia* (mental), *questão* (mental)

Having limited the study to these nine nouns, the following step was to extract all instances of them in the following sequences:

N *de* N *de* + infinitive, N *de que* + clause
 N *que* N *que* + clause, *é* N *que* + clause
 N [dist=9] *é que* N *é que* + clause

This provided me with an initial dataset of a little over 9350 instances. As the figures in Table 1 show, the nine nouns appear in these constructions in a step by step fashion. In the 13th century only *mercê*, *razão*, *sinal* and *caso* are documented, with *vontade* and *temor* appearing in the 14th. Table 1 also reveals that the usage frequencies of the attested nouns go up significantly and the six nouns are used in all four constructions from the 15th century on. In the 18th century the first uses of *ideia* and *questão* appear, and in the 19th and 20th centuries, all nine nouns are attested. The usage frequencies continue to increase for all except *mercê*, which appears to lose its status as a salient shell noun in the last two centuries. Finally, as the percentages shown on the last row of Table 1 indicate, the relative frequency of the use of these nouns in the different formats of the shell-noun construction increases steadily, moving from around 2 per cent in the medieval period, to approximately 5 per cent from the 1400s to the 1700s only to surpass the ten per cent line in the last centuries.

Out of the 9300 concordances extracted from the CP I created a subsample by arbitrarily selecting up to 100 cases of each noun and period (1200s–1300s, 1400–1700s, and 1800s–1900s). Compared with the figures in Table 1, then, the sample which was used for more detailed analysis included all 58 cases from the

all nominal collocates of the same sequences as above as well as N *é* + clause and selected three nouns which are indeed used with an encapsulating function. See Granvik (2015: § 5) and Schmid & Küchenhoff (2013) for discussion of the details of the collostructional analysis. The partition of the data into centuries is due to architecture of the CP.

Table 1: Distribution of the nine nouns across the 13th to the 20th century in the five constructions.

Nouns	1200s	1300s	1400s	1500s	1600s	1700s	1800s	1900s	Sum
<i>mercê</i>	3	8	50	45	73	11	7	1	198
<i>razão</i>	8	17	66	106	105	45	169	201	717
<i>vontade</i>	0	3	95	65	26	41	407	936	1573
<i>sinal</i>	6	2	34	43	72	33	159	169	518
<i>caso</i>	1	7	55	111	103	39	342	273	931
<i>temor</i>	0	3	31	47	16	18	18	44	177
<i>facto</i>	0	0	0	0	0	0	204	2841	3045
<i>ideia</i>	0	0	0	0	0	8	674	1057	1739
<i>questão</i>	0	0	0	0	0	4	41	419	464
Sum	18	40	331	417	395	199	2021	5941	9362
Nro. de N	990	1590	5575	9487	6825	4561	21256	40693	90977
%	1.8%	2.5%	5.9%	4.4%	5.8%	4.4%	9.5%	14.6%	

13th and 14th centuries; all 112 cases of *temor*, all 8 cases of *ideia* and all 4 cases of *questão* as well as 100 arbitrarily selected cases of *mercê*, *razão*, *vontade*, *sinal* and *caso* in the period running from 1400 to 1799. Finally, all 62 cases of *temor*, all eight of *mercê* as well as 100 arbitrarily selected cases of the remaining seven nouns from 1800 to 1999. The total number of cases included in the qualitative analysis is 1446 (see Table 4 in Section 4, below).

This subsample was annotated in detail by means of the variables presented in Table 2. As the table shows, for each concordance line extracted from the corpus I registered the construction involved (N *de* + infinitive, N *que* + clause, N *de que* + clause or N *é que* + clause), the noun (*caso*, *facto*, *ideia*. . .), the century, the immediate usage context of the noun (involving three structural properties, see Table 2), a formal characterisation of the usage as typical, less typical or untypical (based on the three subparts of the usage context variable), and, finally, whether it functions as a shell noun in the strict sense of there being an identifying relation between it and another element in the (immediate) usage context. The aim of this annotation was to capture the essential characteristics of the usage context of the nouns in order to determine to what degree they affect the way the nouns are used.

The key variable of the annotation is Identity relation, which distinguishes between the encapsulating uses of shell nouns, and the use of the nouns in other functions, particularly as part of complex predicates in combination with light-verbs, such as *fazer* ‘to do, make’, *dar* ‘to give’, *haver* ‘to be, exist’, *ser* ‘to be’, etc. The distinction between the two functions, shell vs. non-shell use, was made with the help of two questions:

Table 2: Variables and variable values used in the analysis.

Variable	Values
Construction (cxn)	inf(itive), <i>que, de que, é que</i>
N(oun)	<i>mercê, razão, vontade, caso, sinal, temor, facto, ideia, questão</i>
Century	1200s to 1900s
Usage context	a) Determiner or not; b) syntactic function (object, prepositional complement, predicative, subject); c) the head element of the noun (lexical verb, noun, etc; light verb, preposition, no head)
Formal typicality	typical, less typical, untypical ¹⁰
Identity relation	0 (no relational identity), 1 (identification with previous element or with a following complement clause)

- 1) Is the noun used in a verb-like fashion? Does the subordinate clause function as if it were a direct object?
- 2) Are the paraphrases “the noun is the (subordinate) clause content” or “the clause content is a N” acceptable?

If the answer to the first question is “Yes”, the function is considered non-encapsulating, i.e. “predicate”; if the answer to the second question is “Yes”, then the noun is considered to have a shell function (cf. Schmid 2000: 30). This means that the shell function of a noun is equated to there being a relation of identity or identification between the noun and the subordinate clause, as is the case in (13) as compared to (14), where there is no identification.

- (13) Parece um contra-senso, mas não é, **pela** simples **razão de que**, no meio da anarquia, a lógica não sobrevive! (CP, 19N:Pt:Jornal)
‘It sounds like nonsense, but it isn’t, for the simple reason that in the midst of anarchy, logic can’t survive!’
- (14) A Rosa **tem** toda **a razão de** acender uma vela aos pés de santo Antônio. (CP, 19:Fic:Br:Resende:Braco)
‘Rosa is absolutely right to light a candle at the feet of St Anthony.’

¹⁰ See the discussion of Table 3, below, for more details.

The formal traits making up the Usage context variable, determiner, syntactic function and head element, were used by Granvik (2019) to capture the formal typicality of the use of the nouns. The degree of Formal typicality was determined numerically by applying the value system presented in Table 3. Each example was assigned a number, and sum values from 0–1 were considered typical usage contexts, values 2–3 were considered less typical, whereas the highest values, 4 and 5, constitute untypical usage contexts. These correspond to usage contexts where the nouns bear no determiner, are used as direct objects of highly frequent verbs, often light verbs, such as *dar*, *fazer*, *haver*, *ter* or as prepositional complements of grammatical prepositions such as *com*, *sem*, *a*, *em*, *por*.

Table 3: Values assigned to formal properties to determine the degree of typicality of the usage context of each noun.

Criterion / level	0	1	2
Determiner	definite article demonstrative pronoun possessive determiner	indefinite article or pronoun	no determiner
Syntactic function	absolute, subject	(in)direct object, predicative, prepositional complement, unknown	
Head	lexical verb, non-grammatical prepositions no head	verb <i>ser</i> grammatical prepositions (<i>a</i> , <i>de</i> , <i>em</i> , <i>para</i> , <i>por</i>)	light verb (<i>haver</i> , <i>ter</i> , <i>fazer</i> , <i>dar</i>), prepositions <i>com</i> and <i>sem</i>

In comparison to Granvik (2019), where the formal characterisation was used directly to capture the differences in use of the course of the centuries, in this paper, an additional goal is to verify to what degree the formal characterisation proposed correlates with differences in function, i.e. Identity relation or not. The hypothesis is that formally typical uses will correlate with uses in which the nouns have an encapsulating function, involving an identity relation and establishing an anaphoric or cataphoric relationship to another contextual element.

4 Analysis of the Results

4.1 Diachronic Considerations

From a historical perspective, the main question is, naturally, how the nouns inserted into the shell noun construction are used in the medieval period and how the construction evolves over time. I will address these questions from the perspective of first the nouns and then the construction formats as a reflex of the encapsulating function. As was already seen in Table 1, above, the nine nouns are introduced subsequently into the different construction formats, and their usage frequency increases as we approach the present-day language. This is, of course, totally expected given that the nouns were selected based on their representativity (as measured by usage frequency) of different time periods.

As the figures in Table 4 show, there is a clear difference between the nouns with regard to their use as shell nouns or not. Thus, of the nouns which are attested already in the medieval period, *caso*, *mercê*, and *sinal* are mainly used as shell nouns, whereas *razão*, *temor* and *vontade* are more often used with a non-encapsulating function (*haver/ter razão de* + infinitive ‘have reason to’, *haver/ter/com/sem/por temor de* + infinitive ‘have fear of/that’ and *haver/ter/com vontade de* + infinitive ‘have the will to’). In the following centuries, 1400s to the 1700s, *temor* and *vontade* continue to be used mainly in a non-encapsulating function, *razão* and *sinal* are used in both functions, whereas *caso* and *mercê* are mostly used as shell nouns. In the 19th and 20th centuries, *facto*, *ideia* and *questão* are introduced. While *facto* and *ideia* are almost exclusively used as shell nouns, *questão* is used with both functions. In the 1900s *questão* appears mainly in combinations where no encapsulation is involved (*fazer questão de* + inf). The above considerations, then, allow dividing the nouns into three groups:

- A. Typical shell nouns: *caso*, *mercê*, *ideia* and *facto*
- B. Multi-use nouns: *questão*, *razão* and *sinal*
- C. Non-encapsulating nouns: *temor* and *vontade*

Table 4: Use of the nine nouns with an encapsulating or non-encapsulating function across the centuries.

Nouns / identity relation	1200	1300	1400	1500	1600	1700	1800	1900	Grand total
caso	1	7	17	37	33	12	43	57	207
identity	1	7	17	34	32	12	43	57	203
no				3	1				4
facto							3	97	100
identity							3	97	100

Table 4 (continued)

Nouns / identity relation	1200	1300	1400	1500	1600	1700	1800	1900	Grand total
ideia						8	32	67	107
identity						8	32	66	106
no								1	1
mercê	3	8	23	27	47	3	7	1	119
identity	2	8	22	26	46	3	6	1	114
no	1		1	1	1		1		5
questão						3	9	90	102
identity						3	5	30	38
no							4	60	64
razão	7	17	22	26	36	16	49	51	224
identity	2	6	3	15	17	2	23	19	87
no	5	11	19	11	19	14	26	32	137
sinal	6	2	18	24	41	16	51	49	207
identity	3	2	11	18	22	11	8	32	107
no	3		7	6	19	5	43	17	100
temor		3	30	47	16	17	18	44	175
identity					1	1	2	17	21
no		3	30	47	15	16	16	27	154
vontade		3	51	23	13	15	30	70	205
identity			5	3		1	2		11
no		3	46	20	13	14	28	70	194
identity (sum)	8	23	58	96	118	41	124	319	787
no (sum)	9	17	103	88	68	49	118	207	659
Grand total	17	40	161	184	186	90	242	526	1446

Examples (15) to (25), below, illustrate typical uses of each noun in the century in which its use is the most salient as compared both to the other nouns and to other centuries. In examples (16), (18), (22) and (25), *mercê*, *caso*, *ideia* and *facto* all function as shell nouns, encapsulating the information of the underlined passages. The three multi-use nouns appear once in each usage type, as shells in (19), (20), (24) and not shells in (15), (23), (26), in the century when this use is most salient. In examples (17) and (21), finally, *vontade* and *temor* are used in compound predicate-like expressions where the underlined complement clauses are comparable to direct objects.

- (15) Ca a Sennor que o atan ben dá non á ome **razon de lle furtar** nen de roubar-ll' o seu (CP, Mettman:CantigasSM3)
 'From the Lord who gives it so well a man has no reason to steal it or take it from him.'
- (16) Senhor, **seja vossa mercee que mandees entregar** o seu aos que son com ho Cide. (CP, 13:CIPM:CGEsp)
 'Sire, may it be your honour that you order theirs to be delivered to those who are with the Cide.'
- (17) Danubre, que bem vio que morreria e que **havia vontade de vingar sa morte**. . . (CP, 14:CIPM:Demanda)
 'Danubre, who knew that he would die and wanted to avenge his death...'
- (18) & ser lhe ha tido em segredo,**em caso que elle não queria ser accusador**, para que mais liurementemente o possa descobrir: (CP, 15:Liao:Leis)
 '& it shall be kept secret from him, in case he didn't want to be the accuser, so that he could find out more easily'
- (19) cuidando que se podiam salvar na mesquita, acabaram nela, e assi **era razão que** no lugar onde tinham perdido as almas, dessem sepultura aos corpos. (CP, 15:Barros:Asia2)
 'Thinking that they could save themselves in the mosque, they ended up there, and so it was right that in the place where they had lost their souls, they should bury their bodies.'
- (20) e não se sabe até agora mais que haverem-se ouvido tiros pela madrugada, **sinal de que** foram sentidos. (CP, 16:Vieira:Cartas)
 'and so far no more is known than that shots were heard in the early hours of the morning, a sign that they were felt.'
- (21) se casaria de boa vontade com huma destas damas **sem temor de ser alguma noite degolado** (CP, 17:Macedo:Antidoto)
 'he would willingly marry one of these ladies without fear of being beheaded one night'
- (22) Laura estremeceu de pudor **com a idéia de que em breve estaria totalmente despida** e descomposta (CP, 18:Azevedo:Demônios)
 'Laura shuddered at the thought that she would soon be completely naked and uncomposed.'

- (23) Ela meneou a cabeça afirmativamente, e ele **fez-lhe sinal** de que o esperasse por detrás do cortiço, no capinzal dos fundos. (CP, 18:Azevedo:Cortiço)
‘She nodded, and he made a sign to her to wait for him behind the tene-ment, in the back garden.’
- (24) Venham donde vierem as ideias. A origem pouco importa, **a questão é que elas sejam boas.** (CP, 18:Dinis:Fidalgos)
‘Wherever the ideas come from. It doesn’t matter where they come from, the point is that they should be good.’
- (25) O ascendente que o PC teve no Alentejo deveu-se fundamentalmente **ao facto de o Partido Comunista ter conseguido uma certa mobilização** das massas trabalhadoras (CP, 19Or:Pt:Intrv:Pub)
‘The rise of the CP in the Alentejo was fundamentally due to the fact that the Communist Party managed to mobilise the working masses to a certain extent’
- (26) Esta é a convicção do treinador-adjunto vitoriano, Romeu, que, todavia, **fez questão de salvar a importância** com que todos os jogos são encarados (CP, 19N:Pt:Jornal)
‘This is the conviction of Vitoria’s assistant coach, Romeu, who nevertheless made a point of emphasising the importance of every match.’

Apart from the individual nouns, the formal typicality of the usage context and the construction format also experience changes over time. Consider, first, the figures in Table 5 and the graphs illustrating different facets of these figures in Figures 1a and 1b. As, the figures in Table 5 show, the typical uses show a steady increase over time, reaching approximately a third of all the uses in the 19th and 20th centuries. The less typical uses show comparatively stable frequencies, whereas the frequency of the untypical uses goes down from over 70 per cent until 1699 to approximately a third of the cases in from 1700 on. These figures are directly reflected in the graphs of Figure 1a.

Table 5: Distribution of the three degrees of typicality of the usage context, 1200s to 1900s.

Typicality	1200	1300	1400	1500	1600	1700	1800	1900	Total
typical	0	6/15%	11/7%	11/6%	19/10%	17/20%	74/31%	183/35%	323
less typical	4/22%	5/13%	44/28%	36/20%	36/19%	35/38%	68/28%	173/33%	405
untypical	13/78%	29/73%	106/64%	137/74%	131/70%	38/42%	100/41%	170/32%	725
Total	17	40	161	185	186	90	242	526	1446

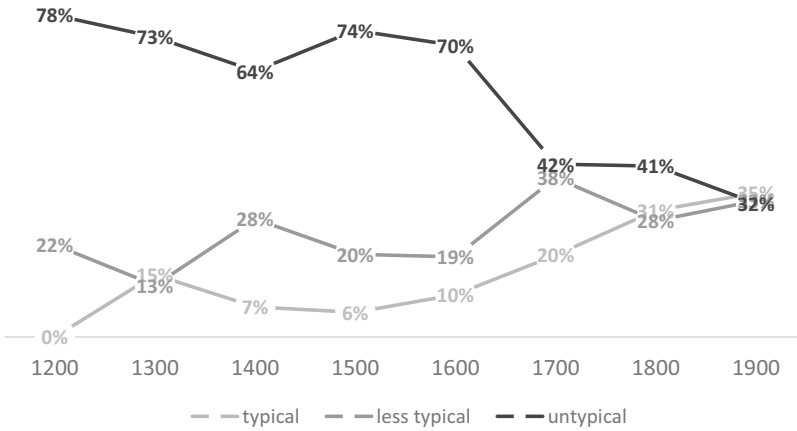


Figure 1a: Distribution of the three degrees of typicality of the usage context, 1200s to 1900s.

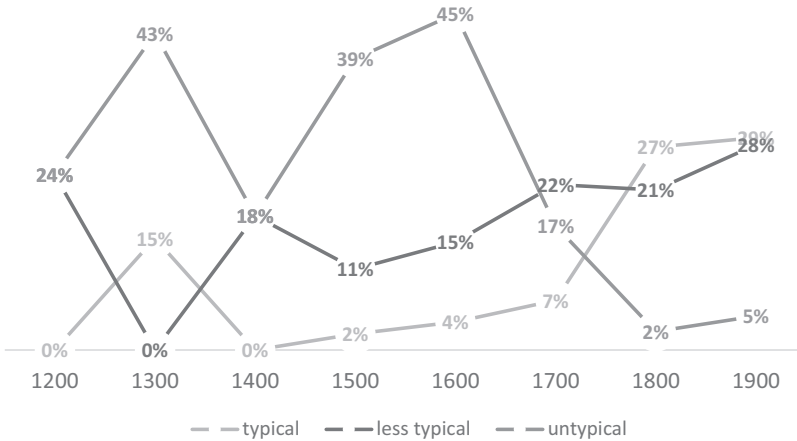


Figure 1b: Proportion of the three degrees of typicality of the usage context of the shell noun uses (involving an identity relation), 1200s to 1900s.

The graphs in Figure 1b, on the other hand, reflect the same tendencies as above, but considering only shell noun uses (with identity relation), illustrating the proportion of the three usage contexts of the cases with an identity relation of all uses (hence the percentages do not add up to 100%). In Figure 1b, the formally typical uses are shown to be of minor importance for the shell noun function all up to the end of the 18th century (0–15 % of all uses); but in the 19th and 20th centuries, the shell noun uses in typical contexts stand for about 30 per cent of all

uses. The line for the untypical uses shows the opposite evolution, with an average of around 40 per cent until 1699 and less than 5 percent from 1800 on. The correlation coefficient between the typical usage context and the shell uses is a convincing 0,95, and for the relationship between the untypical contexts and the non-shell uses it also reaches 0.85. This indicates that the formal usage context is a good predictor of the encapsulating function: when the nouns appear without determiners, in combination with light verbs and as direct objects or prepositional complements they are not used as shell nouns.

Table 6 and Figure 2a portray the distribution of the four construction formats over the centuries, with Figure 2b adding the impact of the construction format on the shell noun uses only. As Table 6 and Figure 2a show, both the infinitive (*N de* + infinitive) and the finite complement clause (*N de que*) construction show an increasing usage frequency over time. The equative construction (*N é que*) seems to maintain its frequency (with the exception of the 20th century), while the finite *N que* construction steadily loses the importance it had in the medieval period. The loss of usage frequency of the *N que* construction format is of course to be expected on the basis of previous studies (see Bogard & Company 1989, Pountain 2014 for Spanish, and Granvik 2015, 2017a, 2017b for Spanish and Portuguese) and must be put in relation to the development of the *N de que* format, which also involves finite complement clauses.

Table 6: Distribution of the four different construction formats, 1200s to 1900s.

Construction	1200	1300	1400	1500	1600	1700	1800	1900	Total
<i>N é que</i>	2/12%	6/15%	17/11%	32/17%	27/15%	1/1%	40/17%	39/7%	163
<i>N de inf</i>	7/41%	23/58%	103/64%	97/53%	92/49%	66/73%	135/56%	360/68%	883
<i>N de que</i>	0	0	0	2/1%	15/8%	6/7%	27/11%	94/18%	144
<i>N que</i>	8/47%	11/28%	41/25%	53/29%	52/28%	17/19%	40/17%	33/6%	255
Total	17	40	161	184	186	90	242	526	1446

When the four construction formats are observed from the perspective of the encapsulating function of the nouns, Figure 2b shows the same trends as above but with some important differences. First, it shows that the equative construction (*N é que*) is, overall, a format where almost all uses are encapsulating ones, with frequencies ranging from 100 to 83 per cent. Overall, 154 out of 163 cases (94%) involve encapsulation; and only *mercê*, *questão* and *razão* are used in this construction format without there being an identity relation. Second, the proportions of *N de que* and *N de* + infinitive construction formats of all shell noun uses also increase over time in Figure 2b. For *N de que* format, the rise starting in the 1600s is much more pronounced as compared to Figure 2a. Third, the trendlines also reveal that *N de* +

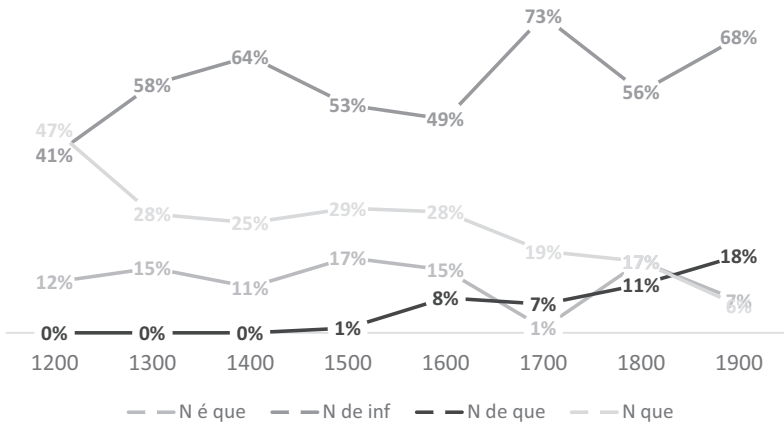


Figure 2a: Distribution of the four different construction formats, 1200s to 1900s.

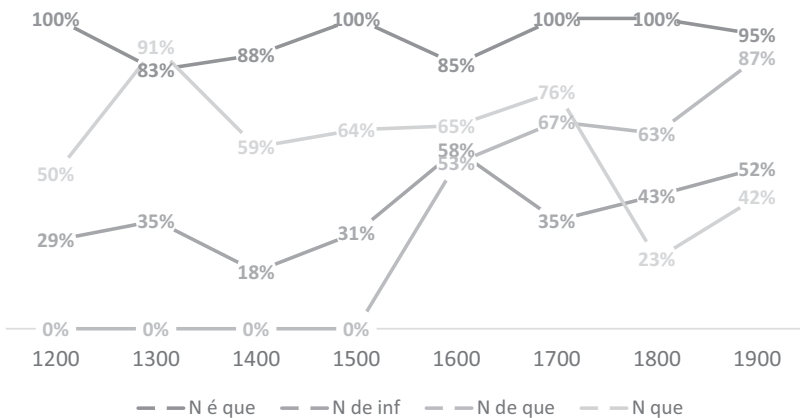


Figure 2b: Proportion of shell uses (involving an identity relation) of the four different construction formats, 1200s to 1900s.

infinitive and the *N que* construction formats run almost parallel until the 17th century; in the 18th century, the *N de + infinitive* format loses some of its importance, only to rise again in the last two centuries, while the descent of *N que* starts in the 19th century.

Figure 2b clearly shows that, overall, *N é que* is a construction format where encapsulation is almost guaranteed. It is clearly the least frequent of the four construction formats, but also the one most specialized on the encapsulating function. This is, of course, a very natural consequence of the construction itself,

which literally equates the noun with its complement clause given the structure *N é que* + clause. The *N de que* format, on the other hand, also shows a growing tendency to involve encapsulating uses of the nouns which culminates in the 20th century. This is due to two things: first, the typical shell nouns *facto*, *ideia* and the multi-use *senal*, often have an encapsulating function when followed by a *de que* clause. Second, almost two thirds of all uses of the *N de que* format are concentrated in the 19th and 20th centuries (see Table 6), when the overall frequency of the encapsulating function reaches its maximum.

The *N de* + infinitive and the *N que* + clause formats, on the other hand, are the least prone to encapsulation. For the *N de* + infinitive format the average of shell noun uses lies below 50 per cent in the whole period analyzed, and 504 out of the 659 cases (74%) of non-encapsulating uses in the data set are in the *N de* + infinitive format. The slight increase that can be observed between the 18th and the 20th centuries (in Figure 2b) is thus best explained by the apparition of the nouns *facto* and *ideia* in the data set, since these nouns are almost uniquely used in an encapsulating function, even when followed by infinitives. The *N que* construction format, finally, is the most frequent of the four construction formats until the 1500s, and from 1600–1700 it is the second most frequent after *N de* + infinitive, but it starts to lose ground with the rise of the other finite construction format, *N de que*, in the 19th century. In the first six centuries, *N que* is more frequent with encapsulating uses than non-encapsulating, but both uses are always attested, and from the 19th century on *N que* is more often used when the nouns do not have an encapsulating function.

4.2 Factors Determining the Encapsulating Function

In this subsection, I present the results of a mixed-effects logistic regression analysis performed to verify which criteria best help determine whether a noun is used in an encapsulating function or not. As we will see, the results align nicely with the tendencies presented in Section 4.1. The dependent variable is the identity relation (see Table 2, above), distinguishing between identification and no identification. The independent variables whereby the model tries to predict shell vs. non-shell uses, are Construction (CXN: *N que*, *N de que*, *N de* + infinitive, *N é* + clause) and Typicality (typical, less typical, untypical, based on the formal values). The Centuries (1200–1900) and the nouns (LEX) are considered as random effects, by which the importance of the other two variables is allowed to vary (see Winter 2020: Ch. 14).

The analysis was performed in R, using the **glmer** function. The numerical results of the regression model are presented in Table 7.¹¹ As the figures in bold face in the rightmost column of Table 7 indicate, the model as a whole is statistically significant (Intercept p value of 0,004) and four of the five independent variables values also stand out as significant. The Odds Ratio values in the second column indicate the direction of the impact, with values below 1 indicating a tendency towards no identification and values above 1 a tendency towards an identity relation. The positive Odds Ratio value (24.85) of the Intercept means that the two variable values which are not shown in the table, namely the Typical usage context and the *N de que* construction format, strongly favor the Identity relation. With an OR value of 14.27, the *N é que* construction format is strongly associated with Identity, i.e. encapsulation. The untypical uses as well as the *N de* + infinitive and the *N que* construction formats, on the other hand, do not favor there being an identity relation. Figure 3 illustrates these tendencies graphically, with the variable values marked in red favoring the absence of an identity relation and the blue ones, especially the one corresponding to the *N é que* construction format, showing predictors for the encapsulating function.

Table 7: Summary of the mixed-effect logistic regression model.

Predictors	Effects on Identity relation		
	Odds Ratios	CI	p
(Intercept)	24.85	2.74 – 225.09	0.004
Less typical	1.13	0.61 – 2.08	0.701
Untypical	0.15	0.08 – 0.27	<0.001
<i>N é que</i>	14.27	5.27 – 38.61	<0.001
<i>N de</i> infinitive	0.24	0.12 – 0.46	<0.001
<i>N que</i>	0.31	0.16 – 0.61	0.001

¹¹ The formula used to create the model was: ID_REL_NUM ~ TYPICAL_NUM + CXN3 + (1 | LEX) + (1 | SEC). The AIC value of the model is 785.3, which can be considered reasonable. The full R script and dataset used to create the tables and graphs presented here can be obtained from the author upon request. Table 7 was created by using the `tab_model()` function and Figures 3 and 4a +b by the `plot_model()` function found in the `sjPlot` package (see http://www.strengejacke.de/sjPlot/reference/plot_model.html).

Table 7 (continued)

Predictors	Effects on Identity relation		
	Odds Ratios	CI	p
Random Effects			
σ^2			3.29
τ_{00} LEX			9.82
ICC			0.75
N_{LEX}			9
Observations			1446
Marginal R^2 / Conditional R^2			0.191 / 0.797

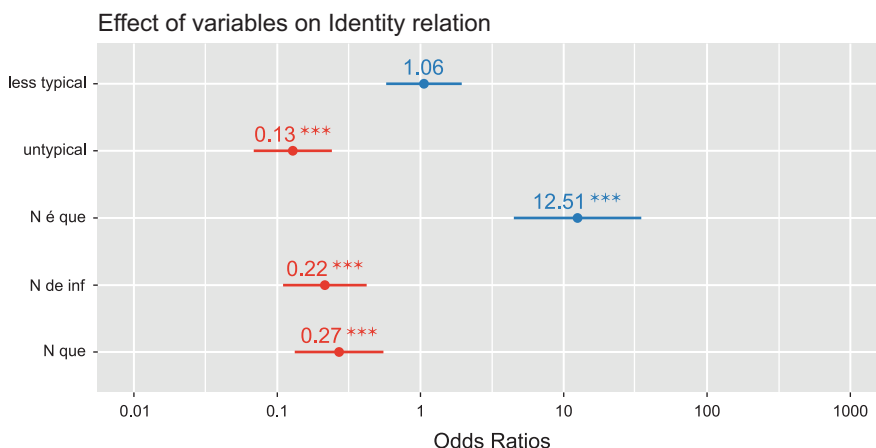


Figure 3: Odds Ratio values indicating the effect of the variables on the Identity relation, i.e. the encapsulating function.

Turning to the random effects, Lexemes and Century, Figures 4a and 4b illustrate similar plots of the effects of the individual lexemes and the centuries on the Identity relation. As Figure 4a shows, there are clear differences between two groups of nouns: those that favor encapsulation, *mercê*, *ideia*, *facto* and *caso*, and those that don't: *vontade*, *temor*, *sinal*, *razão* and *questão*. In agreement with the three-way distinction presented above (§ 4.1), the multi-use nouns, *questão*, *razão* and *sinal* stand out as presenting much higher OR values as compared to non-encapsulating *vontade* and *temor* (i.e. they are much less strongly associated with the non-encapsulating function).

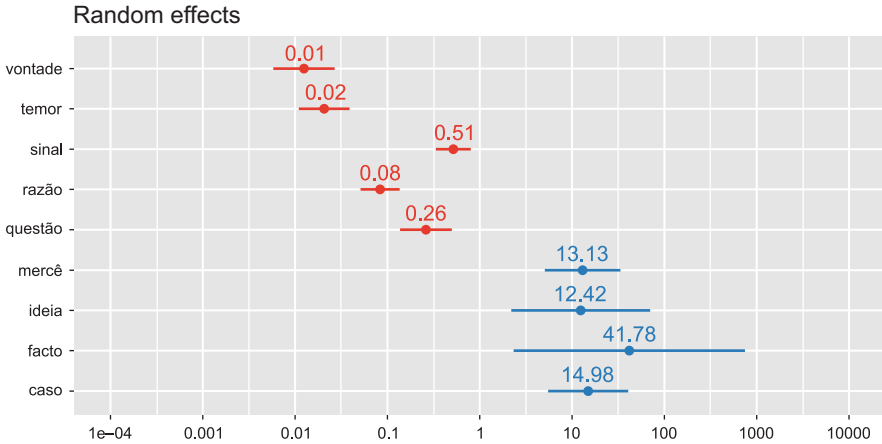


Figure 4a: Odds Ratio values indicating the effect of the different nouns on the Identity relation, i.e. the encapsulating function.

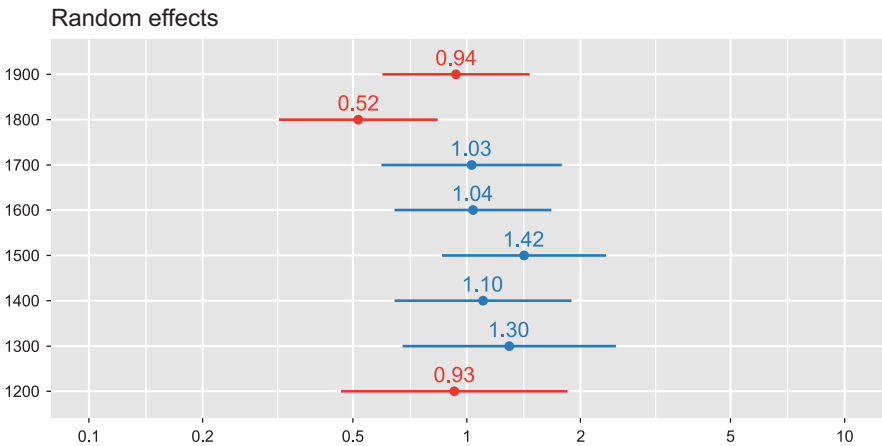


Figure 4b: Odds Ratio values indicating the effect of the different centuries on the Identity relation, i.e. the encapsulating function.

Figure 4b, finally, shows small differences between the centuries and the ranges of the OR values overlap to such a degree that Century as a variable does not seem to be an important predictor for the encapsulating function. This may be explained by the fact that the nouns which are strong predictors are found across the centuries: *caso*, for example, is a strong predictor for the Identity relation and it is used as a shell noun in all centuries. Although *ideia* and *facto* are also strongly associated to uses with an Identity relation, and are nouns which are

only used in the 19th and 20th centuries, the third present-day noun *questão* is strongly associated with non-encapsulating uses. This highlights the fact that, overall, the individual nouns are stronger predictors of the encapsulating uses as compared to the diachronic dimension, at least from the perspective of the nine nouns analysed in this paper.

5 Concluding Remarks

The results of the analysis allow me to answer the research questions in the following fashion. Starting with the last one of them, we have seen that there is a clear correlation between the formal characterisation of the usage context and the shell uses of the nouns: In typical (and less typical) contexts, the nouns are more often used with encapsulating function as compared to untypical contexts. I thus conclude that, for Portuguese, the formal characterisation proposed by Granvik (2019) seems to be a quite reliable proxy for identifying shell-noun uses, understood as uses in which there is an identity relation between the noun and some other element. There is also a (statistically significant and reliable) relationship between some of the constructional formats and the encapsulating function, so that the [N *é* + clause] and the [N *de que* + clause] constructional formats are associated with shell-noun uses. In contrast, when the nouns are inserted in the other two formats, N *que* + clause and N *de* + infinitive the encapsulating function is not much more frequent than a predicative-like function. From this point of view, then, the main encapsulating constructions in Portuguese seem to be [N *é* + clause] and [N *de que* + clause].

Interestingly, although the analysis shows that the shell function becomes more frequent over time (see Table 4, above) the diachronic dimension (centuries) does not stand out as a significant predictor in the regression analysis. Instead, the nouns and the usage context are more important, and the main reason for the increase in frequency of the shell-uses in the 20th century are *facto* and *ideia*, which are almost exclusively used as shells. In fact, as part of the analysis, the nouns were divided into three groups: typical shell nouns (*caso*, *mercê*, *ideia* and *facto*), multi-use nouns: *questão*, *razão* and *sinal*, and the non-encapsulating nouns *temor* and *vontade*. However, it is worth remembering that even when an untypical shell noun such as *questão* is used in a context classified as non-encapsulating (as in (27)), there is often a nuance of encapsulation, marginal though it may be, which reminds us that the construction formats included in this analysis are, really, part of a broader shell-noun construction.

- (27) Ela **fizera questão de** manter sempre trancada a porta que separava os dois cômodos (CP, 19:Fic:Br:Penna:Menina)
 ‘She made a point of always keeping the door that separated the two rooms locked’

In (27), then, the relationship between *questão* and the complement clause is ambiguous. On the one hand, the *question* or *point* is ‘to keep the door locked’. On the other, keeping it locked is what is aimed at, it functions as a sort of direct object, the objective, of *his or her* action, where *fazer questão de* corresponds to ‘making a point’ of achieving something.

From a historical perspective, it is important to notice that the encapsulating uses appear already in the 13th and 14th centuries, with *mercê*, *caso*, *razão*, *sinal*. This is thus not a modern innovation which can be associated with informatically dense texts, as Borreguero (2018) argues. But future work will need to address in more detail which nouns are indeed used as encapsulators in the medieval period and in which text types. A broader understanding of the origins of the shell-noun construction in Portuguese would also need to consider other shell-noun constructions, i.e. demonstrative pronoun + N and (Schmid 2000; Rodríguez Espiñeira 2025), which were not considered here, as well as other salient nouns in order to complete the picture.

Bibliography

- Abad Serna, Silvia (2015): *Estudio contrastivo del funcionamiento semántico de los encapsuladores nominales en la prensa española y alemana: de la anáfora a la catáfora conceptual*. Doctoral dissertation. Universidad Autónoma de Madrid. <<https://repositorio.uam.es/handle/10486/669678>>.
- Bogard, Sergio and Concepción Company Company (1989): “Estructura y evolución de las oraciones completivas de sustantivo en el español”, in *Romance Philology*, XLIII (2), pp. 258–273.
- Borreguero Zuloaga, Margarita (2018): “Los encapsuladores anafóricos: una propuesta de clasificación”, in *Caplletra*, 64, pp. 179–203. DOI: 10.7203/Caplletra.64.11380.
- Borreguero Zuloaga, Margarita and Álvaro S. Octavio de Toledo y Huerta (2007): “Presencia y función de los encapsuladores en las crónicas periodísticas del s. XVII”, in *Philologia Hispalensis*, 21, pp. 125–159.
- Croft, William and Alan D. Cruse (2004): *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Davies, Mark and Michael Ferreira (2006-): *Corpus do Português: 45 million words, 1300s-1900s*. Online at <<https://www.corpusdoportugues.org/hist-gen/>>.
- Flowerdew, John and Richard W. Forest (2015): *Signalling nouns in English. A Corpus-Based Discourse Approach*. Cambridge: Cambridge University Press. <<https://doi.org/10.1017/CBO9781139135405>>.

- González-García, Francisco (2012): “La(s) gramática(s) de construcciones”, in Iraide Ibarretxe-Antuñano and Javier Valenzuela (dirs.). *Lingsüística cognitiva*. Barcelona: Anthropos, pp. 249–280.
- Granvik, Anton (2014): “Hablando *de, sobre y acerca de* la gramaticalización y la lexicalización: panorama diacrónico de las relaciones entre preposiciones y locuciones prepositivas dentro del campo semántico de tema/asunto”, in José Luis Girón Alconchel and Daniel M. Sáez de Rivera (eds.). *Procesos de gramaticalización en la historia del español*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert, pp. 77–117. <<https://doi.org/10.31819/9783954871988-006>>.
- Granvik, Anton (2015): “Oraciones completivas de sustantivo: un análisis contrastivo entre español y portugués”, in *Verba*, 42, pp. 347–401.
- Granvik, Anton (2017a): “Oraciones completivas de sustantivo en español y portugués: ¿infinitivo y oración finita?”, in *Cuadernos de Lingüística de El Colegio de México*, 4 (1), pp. 103–180. DOI: <<http://dx.doi.org/10.24201/clecm.v4i1.54>>.
- Granvik, Anton (2017b): “Análisis histórico-comparativo de las oraciones completivas de sustantivo en español y portugués: nacimiento y evolución de una alternancia sintáctica”, in *Neuphilologische Mitteilungen*, I, CXVIII (2017), pp. 31–63.
- Granvik, Anton (2019): “Sobre los orígenes de la construcción encapsuladora en español”, in Marta Blanco, Helia Olbertz and Victoria Vázquez Rozas (eds.). *Corpus y construcciones. Perspectivas hispánicas*. Anexo 79 de *Verba. Anuario galego de filoloxía*, 2019, pp. 41–79. DOI: <<https://dx.doi.org/10.15304/9788417595876>>.
- Gries, Stefan Th. (2013): *Statistics for Linguistics with R. A practical introduction*. Berlin/New York: De Gruyter Mouton.
- Guimarães, Igor Caixeta Trindade (2011): *A formação nominal em português: um estudo sintáticosemântico de bases enunciativas*. MA thesis (Dissertação de mestría), UFMG. Belo Horizonte.
- Hilpert, Martin (2019): *Construction Grammar and its application to English*. 2nd ed. Edinburgh: Edinburgh University Press.
- Leonetti, Manuel (1993): “Dos tipos de completivas en sintagmas nominales”, in *Lingüística (ALFAL)*, 5, pp. 5–40.
- Leonetti, Manuel (1999): “La subordinación sustantiva: las subordinadas enunciativas en los complementos nominales”, in Ignacio Bosque and Violeta Demonte (coords.). *Gramática descriptiva de la lengua española*. 3 vols. Madrid: Espasa Calpe, pp. 2083–2104.
- López Samaniego, Anna (2011): *La categorización de entidades del discurso en la escritura profesional: las etiquetas discursivas como mecanismo de cohesión léxica*. Doctoral dissertation. Universidad de Barcelona. <<http://hdl.handle.net/10803/48757>>.
- Paredes Pereida da Silva, Vera Lúcia and Gabrieli Pereira Bezerra (2013): “SNS como rótulos em livros didáticos de história do Brasil: simples ou complexos?”, in *Revista diacrítica*, 27 (1), pp. 217–237.
- Pountain, Christopher (2014): “Preposición + que en español”, in *Cuadernos de Lingüística del Colegio de México*, 2. Estudios de cambio y variación, pp. 9–54.
- Ribeiro, Dayhane Alves Escobar (2010): “Levantamento lexical dos encapsuladores utilizados nas redações de alunos do pré-vestibular”, in *Cadernos do CNLF*, XIV (4, tomo 3), pp. 2781–2808.
- Rodríguez Espiñeira, María José (2015): “El sustantivo *hecho* como ejemplar de nombre encapsulador factual”, in *Studium grammaticae: homenaje al profesor José A. Martínez*. Oviedo: Universidad de Oviedo, pp. 655–674.
- Sanches Duran, Magali and Maria das Graças Volpe Nunes (2023): “Aposições anafóricas e catafóricas no português e sua anotação no esquema Universal Dependencies”, in *Anais do*

- Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL)*. DOI: 10.5753/stil.2023.233699.
- Schmid, Hans-Jörg (1997): “Constant and ephemeral hypostatization: *thing*, *problem* and other “shell nouns”, in Bernhard Caron (ed.). *Proceedings of the 16th International Conference of Linguistics* (Paris, July 20–25 1997). Elsevier.
- Schmid, Hans-Jörg (1999): “Cognitive effect of shell nouns”, in Karen Van Hoek, Andrej A. Kibrik and Leo Noordman (eds.). *Discourse Studies in Cognitive Linguistics: Selected papers from the 5th International Cognitive Linguistics Conference, Amsterdam, July 1997*. Amsterdam/Philadelphia: John Benjamins, pp. 111–132. <<https://doi.org/10.1075/cilt.176.09sch>>.
- Schmid, Hans-Jörg (2000): *English Abstract Nouns as Conceptual Shells. From Corpus to Cognition*. Berlin/ New York: Mouton de Gruyter. <<https://doi.org/10.1515/9783110808704>>.
- Schmid, Hans-Jörg (2018): “Shell nouns in English – a personal round-up”, in *Caplletra*, 64, pp. 109–128. DOI: 10.7203/Caplletra.64.11368.
- Schmid, Hans-Jörg and Annette Mantlik (2015): “Entrenchment in Historical Corpora? Reconstructing Dead Authors’ Minds from their Usage Profiles”, in *Anglia*, 133(4), pp. 583–623. DOI 10.1515/anglia-2015-0056.
- Schmid, Hans-Jörg and Helmut Küchenhoff (2013): “Collostructional analysis and other ways of measuring lexicogrammatical attraction: Theoretical premises, practical problems and cognitive underpinnings”, in *Cognitive Linguistics*, 24(3), pp. 531–577.
- Stefanowitsch, Anatol and Stefan Gries (2003): “Collostructions: Investigating the interactions of words and constructions”, in *International Journal of Corpus Linguistics*, 8:2, pp. 209–243.
- Traugott, Elisabeth Closs and Graeme Trousdale (2013): *Constructionalization and constructional changes*. Oxford: Oxford University Press. <<https://doi.org/10.1093/acprof:oso/9780199679898.001.0001>>.
- Winter, Bodo (2020): *Statistics for linguists. An introduction using R*. New York/London: Routledge.