FACULTEIT LETTEREN
EN WIJSBEGEERTE

NON-CANONICAL SUBJECT MARKING IN ROMANIAN

Status and evolution of the MIHI EST construction

Mihaela Ilioaia

Proefschrift voorgelegd tot het behalen van de graad van Doctor in de Taalkunde

UNIVERSITEIT
GENT

To my beloved teenagers, Akash, Astha, and Ştefan.
Always remember your dreams!

Promotor	Prof. dr. Marleen Van Peteghem
	Vakgroep Taalkunde
Copromotor	Prof. dr. Jóhanna Barðdal
	Vakgroep Taalkunde

Non-canonical subject marking in Romanian

Status and evolution of the MIHI EST construction

Mihaela ILIOAIA

UNIVERSITEIT
GENT

Acknowledgements

The completion of this dissertation represents, in many ways, far more than the manuscript enclosed here; it represents the broad linguistic education I received by working with great linguists and, at the same time, the valuable extralinguistic training (or, better said, life training) I received working with great people. It is my pleasure to express my gratitude to those who have been around me on this journey toward PhD-hood.

First, I would like to thank my supervisor, Professor Marleen Van Peteghem, who taught me how to conduct research with passion, but also with responsibility and integrity. From our very first meeting, she accepted enthusiastically to become my supervisor and to guide me through this journey that now approaches its end. Her passion for linguistics and her particular interest in Romanian made me pay closer attention to my native language and realize how beautiful and special this language is. It was always a pleasure to work with her and every meeting we had was encouraging and very stimulating. I was often impressed by her abundant linguistic knowledge. Besides the working meetings, we also had several informal meetings, especially during the lunch, when I had the privilege to listen to her and take valuable lessons from her, beyond the sphere of linguistics. Later, during the process of writing, her patience to read and re-read my chapters often inspired me to carry on. Professor Marleen Van Peteghem has always considered my ideas carefully and responded to them insightfully. Her relentless insistence on clarity of thought and precision of expression has helped me clarify my ideas and, certainly, improve my writing style. I found it very special that, no matter how badly a section was written, she still found something nice and positive to say! Thank you, Professor Marleen Van Peteghem, for believing in me, even in those moments when I, myself, did not believe anymore! I do not have sufficient words to express how grateful I am to you for being part of my life!

Second, I would like to thank my co-supervisor, Professor Jóhanna Barðdal, who always inspired me with her "Golden Hand". Her valuable writing suggestions have been of great use to me, especially at the beginning, when I was afraid to start writing. It was a privilege for me to receive direct guidance from such a great linguist! During our meetings, she

List of Abbreviations

✓	relevant and conclusive
✗	not relevant
=	indicates clitics
?	the example may be rejected by certain native speakers, while others totally accept it
*	the example is ungrammatical and may be rejected by any native speaker
1SG	first person singular
1PL	first person plural
2SG	second person singular
2PL	second person plural
3SG	third person singular
3PL	third person plural
A argument	the subject of a transitive structure
ACC	accusative
ADJ, Adj	adjective
ADV	adverb
Aux	auxiliary
C	the category under scrutiny, i.e. the MIHI EST construction
CL	clitic
COMP	complementizer
COND	conditional
CxG	Construction Grammar
DAT	dative
DO	direct object
DOM	Differential Object Marker
EXP	experiencer
FEM	feminine
F/N	feminine / neuter

FUT	future
GEN	genitive
GER	gerund
IMP	imperative
INF	infinitive
IO	indirect object
Lit.	literally
LOC	locative
MASC	masculine
M/N	masculine / neuter
N	noun
NA	not applicable
NC	non-conclusive
NEG	negation (particle)
NOM	nominative
NP	noun phrase
O	object
O argument	the object of a transitive structure
P	potential productivity
P*	global productivity
PASS	passive
PL	plural
POSS	possessive
PP	prepositional phrase
pro	(i.e. pro-drop) zero or null anafora
PRO	the unrealized subject of a non-finite verb in a control structure
PST	past tense
RECP	reciprocal
REFL	reflexive
S	Subject
S argument	the only argument of an intransitive structure
SAE	Standard Average European
SE	anaphoric reflexive-reciprocal clitic, or middle-passive
SC	small clause
SG	Singular
STIM	Stimulus
SUBJ	subjunctive
SUP	Supine
V	verb
VOC	Vocative

List of Tables

List of Figures

Table of Contents

Chapter 1 Introduction

This research deals with the Romanian dative experiencer construction illustrated in (1). Also called "MIHI EST construction", after its Latin ancestor, this construction contains a dative experiencer, the verb *fi* 'be' in the 3rd person singular, and a bare noun denoting a physiological or psychological state (1a). The dative experiencer is always encoded as a clitic pronoun, which may double a dative NP (1b), since Romanian is a clitic doubling language. As for the bare noun, it may select an additional argument, which mostly denotes the stimulus of the state and can occur as a prepositional phrase (1c), or as a clausal complement (1d–e).

(1) a. *Mi-*   *e*   *foame /*   *sete /*   *frică*
         me.DAT=   is   hunger /   thirst /   fear
         'I am hungry/ thirsty/ scared

     b. *Fetiței*   *i-*   *era*   *foame /*   *sete /*   *frică*
         girl.the.DAT   her.DAT=   was   hunger /   thirst /   fear
         'The little girl was hungry/ thirsty/ scared

     c. *Băieților*   *le*   *era*   *milă*   *de*   *acei*   *copii*
         boys.the.DAT   them.DAT   was   pity   of   those   children
         'The boys felt pity for those kids'

     d. *Mi-*   *a*   *fost*   *frică* *să*   *pun*   *întrebarea*
         me.DAT=   has   been   fear   SUBJ   put.1SG   question.the
         'I was afraid to ask the question'

     e. *Mi-*   *era*   *groază*   *văzând*   *atâtea*   *insecte*
         me.DAT=   was   terror   seeing   so_many   insects
         'I was terrified seeing so many insects'

Dedicated to the MIHI EST construction, this dissertation constitutes a part of a larger research project entitled *Non-canonical subject marking in Romanian, a synchronic and diachronic account*, granted by the FWO.[1] This project aims to study oblique subject-like

---

arguments in present-day Romanian, as well as their evolution in Romanian, from the 16th century until now. More particularly, it investigates to what extent dative and accusative experiencers behave like subjects in Romanian and whether dative experiencers exhibit a stronger propensity to subject behavior than accusative experiencers. The ultimate goal of the project is to investigate whether decline, stability or increase in non-canonical subject marking can be observed in the evolution of Romanian since the first attested texts of the 16th century, bringing it closer to or further away from the Standard Average European type.

The project was inspired by Bossong's (1998) and Haspelmath's (1998; 2001a; 2001b) claims that Romanian is closer to peripheral Standard Average European language families, such as Baltic and Slavic languages, than to other Romance languages with respect to the way it marks its core arguments. Standard Average European (SAE) is a linguistic area that comprises European languages from various families, even non-Indo-European ones, which share a number of specific features spread through language contact. Romance and Germanic languages are considered to be core languages, whereas Baltic, Slavic, as well as Balkan and Finno-Ugrian language groups are considered peripheral in that they show fewer of these specific features. One of those features is the tendency toward canonical marking of core arguments (cf. Haspelmath 1998, 2001a, 2001b, Seržant 2013, among many others), which, according to Bossong (1998) and Haspelmath (1998; 2001a; 2001b) isolates Romanian from the group of central SAE languages, and hence from the Romance family.

(2) *I like books*   (English)
(3) *Îmi*   *plac*   *cărţile*   (Romanian)
    me.DAT   like.3PL   books.the
    'I like books'

Bossong's study is based on ten common experiential predicates in 40 European languages and highlights their way of encoding experiencer arguments. Languages with what Bossong calls *generalizing* predicates (i.e. transitive) encode the experiencer argument in the nominative (2), whereas languages with *inverting* predicates encode these arguments in an oblique case (i.e. accusative, genitive or dative), as in (3). The gathered data are computed in such way, that the score of 0.0 is given to exclusively generalizing languages (cf. English) and 5.0 to exclusively inverting languages (cf. Lezgian). Based on this study, Romanian positions itself on the fourth place, with a score of 2.25, among the most inverting languages such as Icelandic (2.29) and Russian (2.11). As it is well known, Icelandic is argued to have a genuine dative subject since the study by Andrews (1976) and Thráinsson (1979) (cf. also Zaenen, Maling, & Thráinsson 1985; Sigurðsson 2004; Barðdal 2002), whereas Russian is also known to be prone to oblique subject marking (Moore & Perlmutter 2000). Taking the outcomes of Bossong's study as a starting point, Haspelmath (1998) draws the map in Figure 1.1.

Ice(2.29)   Fin(0.87)   Lap(0.81)

Nor(0.12)   Swd(0.12)   Est(0.83)

Ir(2.21)   Ltv(1.64)   Mar(0.79)

Wel(0.92)   Lit(0.83)   Mrd(1.16)

Eng(0.0)   Dut(0.64)   Pol(0.88)   Rus(2.11)

Brt(0.24)   Grm(0.74)   Cz(0.76)   Udm(1.09)

Fr(0.12)   Hng(0.22)

Bsq(0.10)   It(0.48)

SCr(0.75)

Spn(0.43)   Rom(2.25)   Lzg(5.0)

Prt(0.14)   Alb(1.02)   Blg(0.48)   Grg(3.08)

Mlt(0.69)   Grk(0.27)   Trk(0.46)
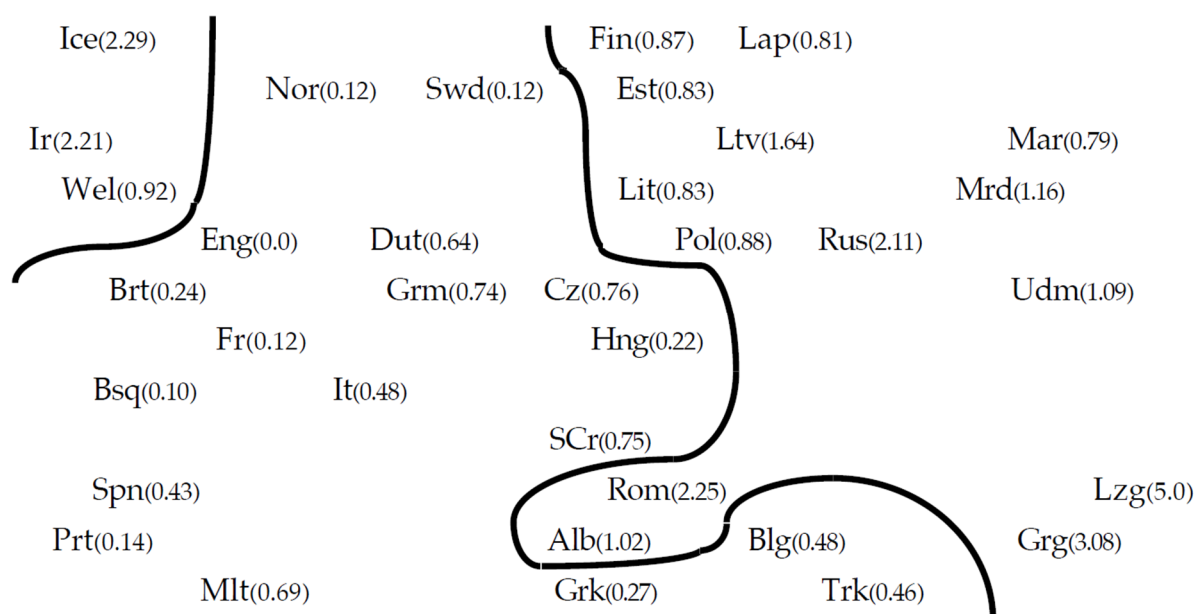
Figure 1.1 Map 2: Predominant generalization (center) vs. inversion (periphery) (Haspelmath 1998: 7)

As pictured in this map, Romanian is classified as being closer to peripheral language groups such as East-Slavic and Baltic languages than to other Romance languages. However, the results presented in Figure 1.1 are intriguing for several reasons. First, Romanian is isolated from the other Romance languages situated in central SAE. Second, in its history, Romanian has undergone substantial contact-induced influence by Old-Slavic and Balkan languages, but not by East-Slavic or Baltic languages, with which it is grouped together here. Furthermore, Haspelmath (1998, 2001b) himself underlines the fact that the results of Bossong's (1998) study are surprising, since, when several SAE features are considered, Romanian shares the same number of features with highly generalizing SAE languages such as Greek (0.27), or even English, which is considered exclusively generalizing (0.0). Figure 1.2 presents a visual representation of the classification of the most typical SAE languages, based on a selection of nine SAE features (cf. Map 107.13 from Haspelmath (2001b: 1505).[2]

---

[2] These nine SAE features are: definite and indefinite articles, relative clauses with relative pronouns, 'have'-perfect, participial passive, dative external possessors, relative-based equative constructions, intensifier-reflexive differentiation, subject person affixes as strict agreement markers (non-pro-drop languages), and negative pronouns and lack of verbal negation.
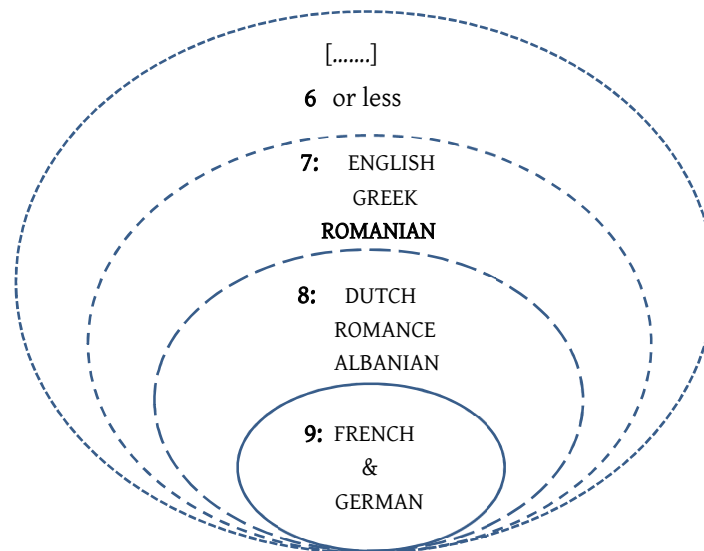
Figure 1.2   Different layers of SAE linguistic *Sprachbund* based on nine selected features (the numbers represent how many SAE feature these languages have in common)

It becomes evident that depending on the SAE features taken into consideration, Haspelmath comes to contradicting conclusions: with respect to a greater set of SAE features, Romanian clusters with highly transitive languages such as Greek and English, whereas the feature related to the encoding of the experiencer isolates Romanian from the central SAE languages. The question arises as to how to explain this contradiction in Haspelmath's conclusions. Van Peteghem & Ilioaia (2017) point out that the outcome of Bossong's investigation may be due to the choice of predicates under consideration. These are three cognition verbs ('see', 'forget', 'remember'), three emotion verbs ('rejoice', 'regret', 'love/ please'), and four sensation predicates ('hunger', 'thirst', 'cold', 'headache'). The problem is that, in Romanian, three out of the total of ten predicates, namely 'hunger', 'thirst', and 'cold', are of the MIHI EST type (4), where Romanian deviates from the other Romance languages (5).

(4) *Mi-*   *e*   *foame /*   *sete /*   *frig*
      me.dat=   is   hunger /   thirst /   cold
      'I am hungry/ thirsty/ cold'

(5) a. French: *J'*   *ai*   *faim*
                    I=   have.1SG   hunger   'I'm hungry'

      b. Italian: *Ho*   *fame*
                    have.1SG   hunger   'I'm hungry'

      c. Spanish: *Tengo*   *hambre*
                    have.1SG   hunger   'I'm hungry'

Hence, the fundamental aim of the project is to measure Romanian's propensity to non-canonical subject marking, based on a wider and more balanced inventory of predicates, namely nominal (cf. 4, above), verbal (6), adjectival (7), and adverbial (8) predicates.

As a part of this comprehensive project, my dissertation concentrates on nominal predicates, contributing with an in-depth analysis of the MIHI EST construction. The present research does not adhere to any particular linguistic framework but borrows concepts and terminologies from both constructional (e.g. concepts of construction, coercion, productivity) and more formal language theories (e.g. notions of raising, control). The MIHI EST structure is considered throughout this dissertation a construction in the constructionist interpretation of the term, defined as associating a particular form with a particular meaning.

(6) *Îmi*   *plac*   *cărţile*
    me. DAT   like.3PL   books.the
    'I like books'

(7) *Mi-*   *e*   *cald*
    me.DAT=   is   warm
    'I feel warm'

(8) *Mi-*   *e*   *bine*
    me.DAT=   is   good
    'I feel good'

The focus of this dissertation on the MIHI EST construction is motivated by its frequency in Romanian. It, indeed, embodies the most natural way of expressing psychological and physiological states in this language. Its prevailing presence in Bossong's list of experiential predicates represents an additional motivation. In spite of its interesting characteristics and behavior, the MIHI EST construction, as well as oblique subject constructions in general, have not yet been studied in detail neither for present-day Romanian, nor for pre-21st century Romanian. The objective of this dissertation is to fill this gap.

1.1   Research questions

The aim of the present research is twofold. In the first place, it aims to investigate whether the MIHI EST construction is an instance of non-canonical subject marking of the experiencer. This analysis has been proposed by Cornilescu (2009) but goes against traditional analyses of the MIHI EST structure, which analyze the construction either as subjectless, or as an impersonal structure with postverbal subject (Pană Dindelegan 2013a: 107).

The second objective is to study the evolution of the MIHI EST construction in Romanian. As mentioned *supra*, it is generally assumed that languages belonging to the SAE linguistic area show a tendency toward canonical encoding of core arguments and that this

tendency causes the regression of non-canonical subject marking (cf. Haspelmath 1998, 2001a, 2001b; and Seržant 2013). However, within the same linguistic area, certain languages are argued to show an opposite tendency, i.e. toward increasing non-canonical marking of core arguments. A case in point is represented by Spanish, as argued by Melis & Flores (2013), as well as certain peripheral SAE language groups such as East-Slavic and Baltic, among which Romanian has been classified (Bossong 1998, Haspelmath 1998, 2001a, and 2001b).

Two central research questions have led my way through the present research:

(i)   To which extent is the dative experiencer in the MIHI EST construction a genuine syntactic subject?

(ii)   Is the MIHI EST construction expanding or regressing in Romanian?

The first research question is meant to solve the dilemma of the syntactic analysis of the dative experiencer in the Romanian MIHI EST construction. To which extent does the dative experiencer in this construction show subject properties? Can the dative experiencer be analyzed as a genuine syntactic subject, as in Icelandic, or is it rather an instantiation of an I-nominal, i.e. an inversion construction in which the dative argument is an initial subject but a final indirect object, as claimed by Moore & Perlmutter (2000) for certain dative experiencers in Russian?[3]

An answer to the second research question will also address the much broader issue of the classification of Romanian among peripheral SAE language families (Haspelmath 1998, 2001a, 2001b, based on Bossong 1998), although it may not completely solve it. In order to answer this question, it is necessary to investigate the evolution of the set of state nouns occurring in the MIHI EST construction throughout the centuries, and also the competition between the MIHI EST construction and its competitor, the HABEO construction (that is, the structure in which state nouns occur with the verb *avea* 'have', conveying a similar interpretation, that of a psychological or a physiological state).

In order to find an answer to these questions, I will adopt a usage-based approach, built on a quantitative analysis of a corpus containing examples from the 16th to the 21st century-Romanian. For present-day Romanian I rely on a web corpus provided by the Sketch Engine platform. As for the pre-21st-century Romanian, I have created my own corpus, tagged and parsed with the tools offered by the same platform.

---

[3] As commented by one of my supervisors, this last question, - whether the Romanian dative experiencer can be analyzed as a genuine syntactic subject, as in Icelandic, or is it rather an instantiation of an I-nominal – is somehow a theoretical question and not an empirical one, as this type of analysis proposed by Moore & Perlmutter (2000) would never be an option in mono-stratal theories like Construction Grammar. An analysis of the non-canonical subject as an I-nominal would definitely need a two-stratal framework.

1.2   Structure of the thesis

This dissertation is structured as follows. Chapter two provides an overview of the literature on the notion of subject. In spite of several difficulties encountered, scholars pertaining to different theoretical frameworks have always attempted to define the concept of subject. Some define the subject based on the coding and behavioral properties put forward by Keenan (1976) and arrive at a scalar concept (Seržant 2013). Under this approach, different *degrees* of subjecthood are distinguished so that an argument may be analyzed in terms of "half" a subject (Malchukov 2018: 8). Other scholars describe the subject in terms of its position in the argument structure, which, in turn, is defined by the internal dynamics of the event (Eythórsson & Barðdal 2005; Barðdal & Eythórsson 2018). Chapter two also addresses other related issues, such as the universality of the concept of subject. Notions like *language-specific* or even *construction-specific* are introduced, since they are argued to be more suitable in describing the subject or the properties that characterize it (Eythórsson & Barðdal 2005: 827, (Barðdal 2006). Typologically, this chapter gives a better insight into the position of Romanian among the European languages with respect to non-canonical case marking of subjects, comparing it to several languages, which are argued to have non-canonical subjects.

Chapter three discusses the subject tests proposed by Keenan (1976) with reference to Romanian. The relevance of these tests is first examined for canonical subjects in order to determine the language-specific properties of Romanian subjects. Then the relevant tests are applied to accusative and dative experiencers of psychological verbs and to several complex-predicate constructions in Romanian. Complex-predicate constructions are defined as experiencer constructions that get their experiential meaning from the combination of an experiencer with a light verb or a copula and a state. I show that, while these accusative and dative experiencers lack Keenan's (1976) coding properties, they pattern with nominative subjects in canonical structures with regard to Keenan's behavioral properties.

Chapter four focuses on the Romanian MIHI EST construction as a descendant of the Latin MIHI EST pattern, categorizing it as a complex predicate construction. I distinguish between two main types of complex predicate constructions: the *determined state noun type* and the *bare state noun type*. The *determined state noun type* is represented by the *CAPIO inchoative* construction (cf. *Mă apucă foamea* lit. me.ACC seizes hunger.the 'I start feeling hungry') and by its aspectual opposite, the *cessative* construction (*Mi-a trecut foamea* lit. me.DAT has passed hunger.the 'I don't feel hungry anymore'). As for the *bare state noun type*, it is represented by the MIHI EST construction (cf. *Mi-e dor* lit. me.DAT is longing 'I miss') and by the *VENIO inchoative* construction, its inchoative variant (cf. *Îmi vine somn* lit. me.DAT comes sleep 'I begin feeling sleepy').

Chapter five provides details about the corpus and the methodology used for this study. It explains the methodology used in extracting relevant examples from the gathered corpora, highlighting, at the same time, the phases of the compilation of the dataset.

The methodology is further described in Chapter six, which explains the process of collecting the set of nouns that occur in the MIHI EST construction. This inventory, which seems to be quite dynamic throughout the centuries, counts 29 different nouns for the pre-21st century Romanian, and 95 nouns for the present-day language. Interestingly, the same nouns were found to occur also with the verb *avea* 'have', in the HABEO construction. A closer look at the dominant construction of the selected nouns reveals a tendency for these nouns to occur first in the HABEO construction before being recruited by the MIHI EST construction, tendency specific for the first three historical periods.

Chapter seven is entirely dedicated to the analysis of the MIHI EST construction, more particularly to the status of the dative and of the state noun. The tests considered relevant with respect to canonical subjects in chapter three are now applied to the MIHI EST construction. It is shown that, although not encoded in the nominative and not triggering verb agreement, the dative experiencer in this construction patterns with canonical subjects rather than with objects, whereas the state noun, which is traditionally analyzed as the subject of the structure, does not behave as a subject.

Chapter eight aims to answer the question whether this construction is expanding or regressing in Romanian. This issue is investigated by means of two distinct studies, based on two different types of data: a survey among Romanian native speakers presented in the first section, and a corpus study described in the second section of the chapter. The survey verifies whether a series of nouns, not attested in the MIHI EST construction – although semantically close to the nouns that do occur in it – are accepted in this construction by native speakers of Romanian. As for the corpus study, it estimates, by means of a combination of quantitative and qualitative assessment tools, the productivity degree reached by the MIHI EST construction at different stages of the language. This study aims to detect possible changes in the degree of productivity of the MIHI EST construction throughout the centuries, and to evaluate whether the usage of this construction is expanding or retracting in the present-day language.

Before closing this introductory chapter, I would like to highlight the importance of this dissertation for the project to which it belongs, but also for linguistics and typology in general and for Romanian linguistics in particular. This dissertation is novel and significant in the following ways: (i) it is the first in-depth synchronic and diachronic study of subject-like obliques in Romanian, with special focus on the nominal predicates; (ii) it is the first linguistic study of oblique-subject constructions based on a reliable electronic corpus of Old Romanian; (iii) it answers crucial questions regarding the status of subject-like obliques on the one hand, and the evolution of non-canonical subject marking in Romanian on the other hand, giving a first indication toward a better insight

into the typological position of Romanian between Romance, Balkan and Slavic languages.

Chapter 2   The subject: toward a universal definition

Subject is both one of the most controversial notions in linguistics and the one most often taken for granted.
(Bakker & Siewierska 2007: 141)

For almost any language, linguists will agree on what the subject is in a specific sentence, yet few would agree on a way to define it. The term "subject" has been used for over two millennia in the study of language, logic and philosophy, yet the term is still highly debated, when it is not taken for granted, as pointed out by Bakker & Siewierska (2007).

This chapter gives an overview of the abundant literature on the concept of subject. The first part presents the notion of subject as it has been perceived in different frameworks. In the second part, more light is shed on the concepts of canonical and non-canonical subject. The third part of this chapter surveys non-canonical subjects along the past five decades of linguistic research.

2.1   Canonical subjects

The notion of *subject* is fundamental in Aristotelian logic and in almost all Western traditions of thinking about philology and grammar. The term *subject*, as it is used in philosophy today, goes back to the Latin translation ('subjectum') of a Greek term coined by Aristotle, *hypokeimenon*, which literally means 'the underlying thing'.

For Aristotle, the term *subject* literally meant 'that which underlies an existing thing, its material substratum' (for more on Aristotelian notion of subject, see Blunden 2005). In the philosophical sense, the subject is the fundamental *substance* which makes the thing what it is rather than something else, and to which attributes (also called, in classical times, *accidents*) may be contingently attached. Aristotle's concept of subject has also been understood in the modern grammatical sense, in the context of simple assertive sentences, in which *a predicate 'is said of' a subject.*

Several linguists have attempted to make this notion clearer and more intelligible, whereas some scholars such as Martinet (1972) have considered abandoning it. In his book *Eléments de la linguistique générale* (1960), Martinet describes the subject in French as having as a main characteristic *non-omissibility* (Martinet 1960: 125). Concerned with the

difficulty to define the subject cross-linguistically, but also with the possible complications occurring if this notion is abandoned, he admits that the criterion of non-omissibility may not apply cross-linguistically, but it certainly works for the western Indo-European languages. However, the rejection of the notion of subject altogether would create *terminological complications* (cf. Martinet 1972: 175-179). In a later study, he points out that the subject has no specific, pre-defined semantic role, yet he also observes that the subject frequently coincides with the actor (cf. Martinet 1985: 178).

The interest in this concept, more precisely in its grammatical sense, has increased during the last few decades, starting with Keenan (1976) among others (Anderson 1976; Perlmutter 1982; Zaenen, Maling & Thráinsson 1985; Dixon 1994; Eythórsson & Barðdal 2005) who attempted to pinpoint universal properties of subjects. In his seminal article, Keenan discusses the behavior of arguments in a variety of languages and suggests several properties as characteristics of the universal subject. These properties are divided into *coding*, *behavioral*, *semantic* and *pragmatic* properties. Among these, morphological case, subject-verb agreement, and position are identified as coding properties, whereas a wide array of properties including control of reflexives, raising, and omission on identity in second conjuncts and in controlled infinitives are considered as behavioral properties. Additionally, he defines the thematic role as a crucial semantic property, whereas topichood, empathy, and definiteness are considered as pragmatic properties of the subject.

Keenan himself was aware of the difficulty to provide a universal definition of subject, as well as the difficulties in providing universal criteria. He acknowledges that there is no combination of subject properties which is both necessary and sufficient for an NP in any sentence in any L(anguage) to be the subject of that sentence. The scholar, nevertheless attempts to formulate a definition for this notion, based on the properties he identified: "an NP in a b[asic]-sentence (in any L) is a subject of that sentence to the extent that it has the properties in the properties list below. If one NP in the sentence has a clear preponderance of the subject properties, then it will be called the subject of the sentence." (Keenan 1976: 312)

Crucially, besides the traditional coding properties, only Keenan's concept of behavioral properties, has had any lasting value, since the other types of properties (semantic and pragmatic) cannot be used to distinguish between subjects and objects. This means, in effect, that in later literature, the coding and behavioral properties are the most relevant in identifying the subject.

Another definition of the concept of subject that has been repeatedly proposed in the literature is the one suggested by Anderson (1976), and developed by Dixon (1994) and Andrews (2007). Under this definition, the subject is a grammatical relation involving A (= the subject of a transitive structure) and S arguments (= the only argument of an intransitive structure), as opposed to the O argument (= the object of a transitive structure) (cf. Dixon 1994: 113–119). A basic idea in Dixon's work is that "A, S and O are

universal core categories, and that syntactic rules in every grammar are framed in terms of them" (cf. Dixon 1994: 113). Reflecting on the universality of the subject, Dixon observes that the relation of subject may have non-overlapping properties cross-linguistically, in that it will not be manifested in the same range of constructions, and different constructions within the same language may not all define this relation. However, he argues that the subject is a universal grammatical relation since, for all languages, there is at least some construction that defines the argument class taken as distinctive for this relation, even though this specific relation may be manifested differently in the grammar (cf. Dixon 1994: 113-119).

In the same vein as Dixon (1994), Bresnan et al. (2001, 2015) observe that grammatical relations such as subject and object play a role in all languages. The scholars suggest that these relations should be kept distinct from the forms of expression found in each language, such as particular phrase structure configurations or particular case inflections. Rather, grammatical relations are best viewed as classes encompassing varying forms of expression that are mapped in the same way onto argument structure from one language to another. These classes are universal and ensure that the varying forms of expression found in different languages are always mapped onto the same arguments. For instance, particular noun phrase configurations in English and noun phrases marked ergatively in Warlpiri are both marked onto agent arguments because they belong to the same class of subjects (Bresnan et al. 2015: 8-10). Bresnan et al. argue that in non-configurational languages such as Warlpiri grammatical relations cannot be defined in the same way as in English, because the language lacks the relevant phrase structure configurations.[4] This means that the information usually associated with phrases is conveyed at the morphological level, through case and number inflection for different nominal elements, as shown in Figure 2.1, where the relation subject is expressed in the same sentence by two different NPs, both encoded in the ergative case.

---

[4] Non-configurational languages are characterized by a flat phrase structure, allowing syntactically discontinuous expressions, and a relatively free word order (Golumbia 2004). This phenomenon has been explained as an instance of scrambling, in more recent studies (cf. Donohue 2011: 501).
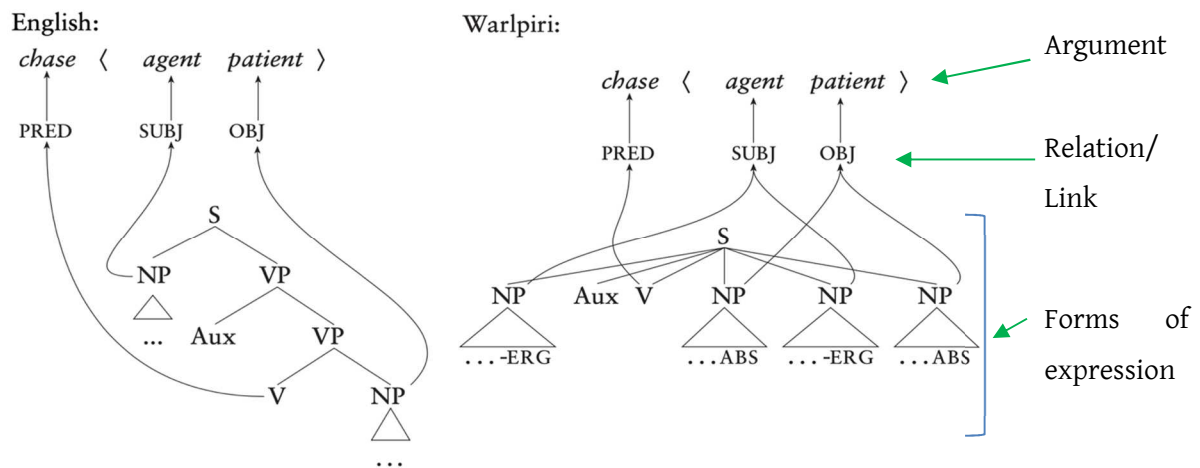
Figure 2.1   Grammatical structure of English and Warlpiri (original figures 10 and 11 merged from Bresnan et al. 2015: 8-9)

Nevertheless, in more recent studies, scholars take a new look at the concept of non-configurationality, explaining it as *scrambling.* Donohue (2011: 501) argues that the so-called *non-configurational NPs* are actually scrambled variants of more *basic* configurational structures, with contiguous NPs.

The concept of subject is fundamental in different paradigms within the broad tradition of generative grammar, especially in Lexical Functional Grammar and in Relational Grammar. For instance, in Lexical Functional Grammar, the subject is the noun phrase immediately below the sentence node or, in Minimalism, the noun phrase occupying the specifier position of IP (see Farrell 2005: chap. 5 for an extensive discussion of the details of these definitions).

As for Relational Grammar, Perlmutter proposes to discriminate between the universality of the *subject* and a universal definition of the subject. He takes *subject* to be a theoretical primitive of relational grammar – thus applicable to all languages –, but states that not the subject, nor any other relation can be defined in the same way in every language (Perlmutter 1982: 324). He suggests that different notions of subject may be identified in different languages (Perlmutter 1982: 324).

As opposed to these paradigms, McCloskey (1997: 197) points out that, in contrast with the original theory of Chomsky (1965), in the further approaches to generative grammar (the Standard Theory, the Extended Standard Theory, the Principles and Parameters Theory, and the Minimalist Program) the notion of subject does not play any formal role. Instead, the traditional category of subject is "progressively deconstructed" (McCloskey 1997: 197).

From a typologist's perspective, and highlighting the practice of defining the subject based on its properties, Lazard (2009) reiterates the impossibility of finding a cross-linguistic definition of the notion of subject based on a list of properties, since these properties tend to be different for each language (Lazard 2009: 151).

The same difficulty, related to typological differences between languages, has been pointed out by Eythórsson & Barðdal (2005: 827), who suggest that a universal concept of subject cannot be maintained due to the great variety of constructions containing it in each language. This may have caused the shift of the focus, observed in recent years, from universal properties of subjects to *language-specific* or even *construction-specific* properties of subjects. Hence, the subject itself is viewed as being construction-specific since it is characterized by sets of construction-specific properties, within one and the same language.

Nevertheless, comparing the distribution of subject properties in English, Icelandic, and German, Eythórsson & Barðdal (2005) suggest a theory-independent definition of subject. In their view, the subject is the leftmost argument of its subcategorization frame. The problem then araises as how to identify the internal order of the arguments in the argument structure. The same scholars specify that this order is determined by "the causal conceptual structure of the predicate and the force-dynamic relations between the participants of the event denoted by each predicate" (Eythórsson & Barðdal 2005: 831).

The definition proposed by Eythórsson & Barðdal (2005), which does not completely rely on the set of properties provided by Keenan, is basically, non-scalar. However, recent studies put forward a gradient view on the notion of subject. One of them is the work by Seržant (2013), in which a prototype theory is suggested. Under this approach, the prototypical subject is defined as the member with the maximal set of subject properties found in a specific language. Viewed as a radial category, the subject covers a variety of less prototypical instantiations grouped together around the prototype in a structured way. The problem with this approach is that it cannot be used to distinguish between subjects and objects. The idea of prototypical subject has also been developed in the cognitive linguistics paradigm (cf. Lakoff 1987; Janda 1993; Nesset, Endresen & Janda 2011, among others). However, Keenan's (1976) set of behavioral properties represent the only relevant prototypical approach able to distinguish between subjects and objects.

Before closing this section, it is worth observing that the notion of language-specific or construction-specific subject is not new. It was first elaborated by Dryer (1996; 1997), followed by several other scholars (cf. Van Valin et al. 1997: 250–274; Croft 2001; Barðdal 2006). Dryer (1996; 1997) proposes an approach to grammatical relations that radically differs from previous approaches, in that, it only posits language-specific, not cross-linguistically valid grammatical relations. In this view, individual constructions (such as e.g. case marking patterns) define different argument classes in nominative, ergative, and active languages, which, hence, have different grammatical relations. Similarly, particular grammatical relations like the subject can only be posited for

individual languages. Hence, the notion of grammatical relation cannot be cross-linguistically valid; and this, either in the sense that all languages have the same grammatical relations, or in the sense that the grammatical relations found in different languages should be expected to be instances of the same grammatical relation (even if that relation is not found in all languages).

For instance, the Philippine languages do not have such a grammatical relation encoding the actor or topic that could be called subject. Note that there is no difficulty within these languages to identify any grammatical relation; the difficulty occurs when it comes to decide how to apply the terminology that has been widely used for other languages. Hence, Dryer (1997) argues that using the same labels such as *subject* or *object* for grammatical relations in different languages is a terminological issue, since one is dealing with different grammatical relations in each case (Dryer 1997: 123-132).

In their paper, Barðdal & Gildea (2015: 30) contrast two views on these relations: a traditional approach labeled *Part-Part relation*, and a modern approach, namely the *Part-Whole relation.* The former entails that the subject and the predicate are in a specific relation at the syntactic level, labeled the "SYN field" in construction grammar, relation, which is supposed to be general and to apply to all subjects and their predicates. The latter, in contrast, yields a relation between each of the syntactic elements with the SYN field as a whole. In other words, in the Part-Whole relation, the subject is not in a syntactic relation with its predicate, but rather it has a specific role in the construction in which it occurs. From the point of view of Radical Construction Grammar (Croft 2001; Barðdal 2006; Cristofaro 2009), it is better to consider grammatical relations as a Part–Whole relation. This approach allows for variation in the behavior of subjects within a language, depending on the construction in which they occur. The two following graphs provided by the two scholars are meant to help the reader form a mental representation of these relations. The graph in Figure 2.2 is based on Croft's (2001) and Croft & Cruse's (2004) representation of a construction, to which the labels for the two fields of a construction, the SYN field and the SEM field have been added. The graph in Figure 2.3 is a visualization of the two relations, the Part-Whole (left) and the Part-Part (right) relations, presented above.
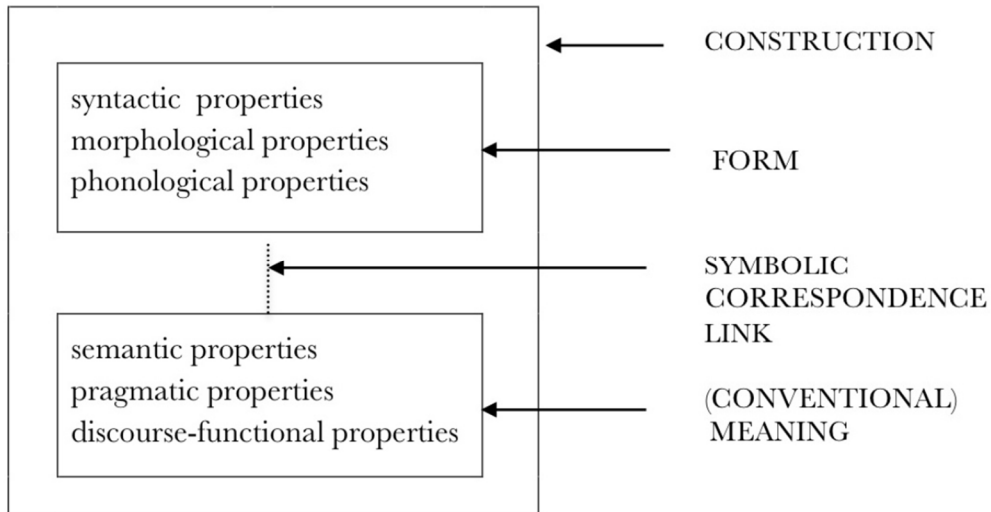
Figure 2.2   Form, meaning, and the correspondence between the two (Croft 2001: 18; Croft & Cruse 2004: 258)
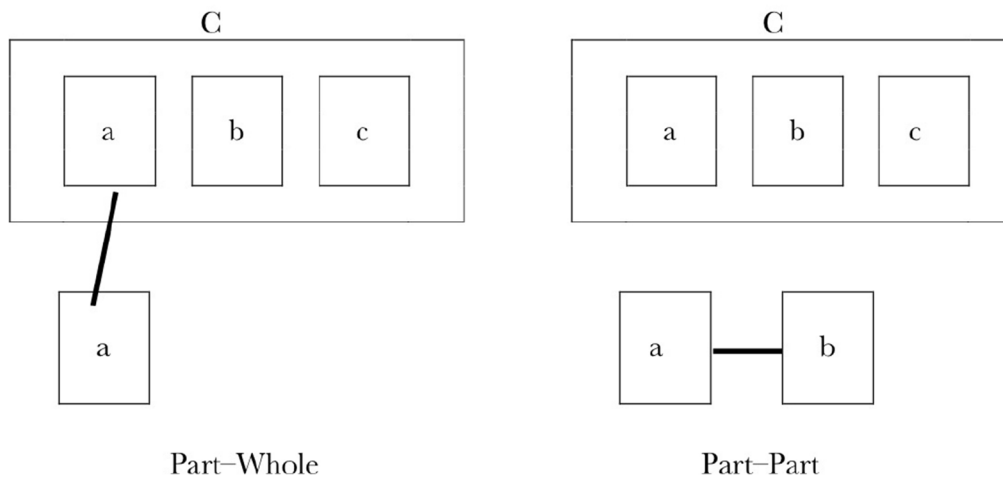
Figure 2.3   Part-Part vs. Part-Whole relations (based on Croft 2001: 24, given in Barðdal & Gildea 2015: 23)

Barðdal & Gildea (2015) highlight the invalidity of the Part-Part approach by suggesting that, diachronically, any change in the subject relation in a language should affect all the predicates in that language simultaneously and in the same manner. This, however, is not the case as it has been shown in several studies (Cole et al. 1980; Haspelmath 2001; Haspelmath & Caruana 2000; Fedriani 2009). In contrast, the Part-Whole approach appears to be more comprehensive, since it postulates that subjects of different constructions may behave differently at different times and that these changes implicitly affect the constructions in which these occur, which reinforces the construction-specific character of the subject relation.

Indeed, at the moment when this dissertation is written, the language-specific and construction-specific hypothesis on categories and relations is the most convenient since, as concluded by Cristofaro (2009), it does not require additional mechanisms to account

for this diversity (Cristofaro 2009: 33-34). Yet, I totally agree with Cinque (2007) in that, looking for universals is a sound linguistic enterprise, since grammatical representation in a speaker's mind may eventually reveal a unitary underlying plan, which, if not sought, could be missed. (Cinque 2007: 93).

2.2   Canonical vs. non-canonical subjects

Since several linguists agree on the hypothesis that the subject is not only language-specific, but construction-specific, it is not surprising to identify more than one type of subject. In the last decades, numerous articles have been devoted to the issue of so-called *non-canonical subjects*. As opposed to *canonical subjects*, the term *non-canonical subjects* is used to describe arguments that are not encoded in the nominative and do not show agreement, but do show other subject properties. It comes without saying that a universal definition of non-canonical subjects may be difficult to formulate, since the concept is genuinely related to the canonical subject, which, as we have seen in the previous section, has given rise to interminable discussions.

Arguing in favor of a language-specific concept of (canonical) subject, Perlmutter (1982) points to the necessity of considering different notions of subject (Perlmutter 1982: 324). Lazard (2009: 158) suggests distinguishing between two types of subjects, based on the type of properties that characterize each of them: *predication subject* and *referential subject*. The *predication subject* is the subject of an agentive clause, characterized by (a minimal) case marking and by agreement, whereas the *referential subject* is identified as the subject of clauses containing an experiencer, and has, as main properties, clause-initial position and reflexive binding. Seržant (2013) proposes to differentiate between a *prototypical subject*, which is the canonical subject, and several *non-prototypical subjects*, also called *non-canonical subjects*.

The prototype theory proposed by Seržant (2013) gives an interesting view on concepts such as *canonical* and *non-canonical*. As specified by the scholar, the *prototypical subject*, or the prototype, describes a *canonical subject*, that is, the member with the maximal set of subject properties found in a specific language. The maximal set of properties typically consists of the following types: (i) behavioral properties, such as control of PRO, raising, binding of reflexive anaphora, etc.; (ii) coding properties such as agreement and case-marking; and, to a lesser extent, (iii) semantic properties, such as thematic role (Keenan 1976). The maximal set of properties is subject to cross-linguistic variation, and can vary within one and the same language as well.

The less prototypical or *non-prototypical*, peripheral members of the category of subject, are defined by their lack of some of these properties (cf. Croft 2002: 162) and/or

the presence of properties of other prototypes. Thus, non-prototypical subjects may share some properties with other prototypes, such as the direct or indirect object (Seržant 2013).

Although there may be an infinite number of subtypes of non-prototypical subjects, Seržant (2013) distinguishes between two main subtypes: *non-canonical subjects* and *subject-like obliques.* The scholar suggests that, regarded from a diachronic perspective, these two notions represent two stages in the evolution of the subject toward canonicity. The latter represents an earlier phase in this evolution and is positioned further from the prototype, whereas the former is considered to be much closer to the prototype, and is an intermediate phase in the evolution toward a canonical subject. Hence, *non-canonical subjects* refer to oblique NPs that are not endowed with canonical subject case-marking and have no access to (canonical) verbal agreement; at the same time, they are characterized by several behavioral subject properties, what makes them syntactically full-fledged subjects (Sigurðsson 2002; Holvoet 2013). As for *subject-like obliques*, these are constituents that significantly deviate from the subject prototype, in that they lack not only morphological, but also most behavioral subject properties. These constituents only share a small subset of semantic or pragmatic properties with the subject prototype.

Seržant suggests that the labels *subject-like oblique* and *non-canonical subject* are to be understood as notions representing two opposite poles of sets of properties. The transition between them is gradual and primarily affects the behavioral properties (Seržant 2013: 321). This diachronic tendency for oblique experiencer arguments to acquire first behavioral properties, followed by coding properties, had been described much earlier for various languages by Cole et al. (1980), and is recalled by Haspelmath (2001: 75).

Seržant's (2013) prototype theory, along with the assumptions over the development of this prototype, proves to have several issues, as recent studies have shown (Barðdal & Eythórsson 2018). One issue is the fact that it admits the possibility that some subjects may share properties with direct or indirect objects (Seržant 2013), which makes more difficult the identification of each of them. Another issue is that it complicates even more the understanding of the notion of subject, by suggesting the existence of different degrees of subjecthood, at different stages of a language. As a consequence, a subject-like argument ends up by being analyzed as "one-third" of a subject, "one-fourth" of a subject, or "half " a subject, as it has been subtly proposed for Lithuanian (Holvoet 2013, 2016; Seržant 2013).

Instead, Barðdal & Eythórsson (2018) and Eythórsson & Barðdal (2005) propose an analysis of the subject in terms of argument structure and the relation between the arguments. Hence, following this approach, where the subject is seen as the leftmost element of the argument structure, the notion of non-canonical subject becomes superfluous, and the discussion involves either nominative subjects, or obliques. Oblique subjects may differ crosslinguistically, some languages having dative oblique subjects,

and other languages having accusative, or genitive, or all of them together. In order to make it easier for the reader, I continue to use the term *non-canonical subjects*, when referring to oblique subjects, in the remaining of this dissertation. The following section will elaborate on the record of oblique subjects in the literature.

2.3   Non-canonical subjects

Non-canonical subjects have been subject to considerable research in linguistics, both synchronic and diachronic. In the literature, the term is used in its broad sense, referring to all non-prototypical subjects in Seržant's terms, i.e. subject-like obliques, as well as to non-canonical subjects. In what follows, I continue using the term non-canonical subject in its broad sense, referring to any non-nominative subject.

Non-canonical subjects raise interesting problems, which have been tackled both in formal approaches such as Generative Grammar and in functionalist frameworks such as Construction Grammar. They are also of particular interest from a typological and diachronic point of view. It is therefore not surprising that numerous articles and books have been dedicated to this specific topic during the last two decades (cf. Aikhenvald, Dixon & Onishi 2001; Bhaskararao & Subbarao 2004; Serzant & Kulikov 2013; Barðdal, Pat-El & Carey 2018).

Non-canonical subjects have been given various names in the literature such as *non-prototypical subjects*, *subject-like obliques*, *quirky subjects*,[5] *oblique subjects*,[6] *derived subjects, dative/accusative/genitive subjects,* etc. These terms refer to dative or, more rarely, accusative or genitive arguments that display a certain number of behavioral subject properties identified by Keenan (1976). Hence, non-canonical marking of the subject results from a lack of clustering between certain coding properties of the semantic subject (case, verb agreement), on the one hand, and its syntactic properties, on the other hand. These arguments are most often experiencers, that is, they are non-agentive. Thus, they are not encoded as canonical subjects with respect to case and verb agreement. This means that non-canonical case-marking is mostly found in languages with a richer case-marking such as German, Icelandic, and Romanian, as opposed to languages such as English and French, which are endowed with a very poor case marking.

---

[5] The term *quirky* is defined in the Cambridge Dictionary as "unique, unusual in an attractive and interesting way" (https://dictionary.cambridge.org/dictionary/english/quirky Consultation date: 13/08/2019).

[6] The term *oblique* was originally used in opposition to "nominative", i.e. nominative case and oblique cases. Hence, in this context, "oblique" means 'non-nominative'.

A variety of behavioral properties of subjects put forward by Keenan (1976) have been utilized as syntactic tests for subjecthood in several languages. Certain properties relate to word order, which was initialy considered by Keenan as a coding property, but it is argued to be more suitable as a behavioral property (cf. Le Mair et al. 2017) (first position in declarative clauses, first position in subordinate clauses, subject-verb inversion). Other properties concern raising and control, two elements of the generative framework. Raising is defined as the movement through which the subject of the verb of the embedded complement clause becomes the subject or the object of the main verb (subject-to-subject raising and subject-to-object raising). Control occurs when a verb shares a semantic argument with its verbal complement, in most cases, an infinitive clause (control infinitives). Reflexivization is another element of the generative framework, also known as binding of reflexives (cf. Binding Theory), and refers to the relation of co-reference (i.e. binding) between an anaphor and a c-commanding antecedent. The anaphor is mostly situated in a local domain, i.e. in the same clause (clause-bound reflexivization) but it may also be situated outside the local domain, i.e. in another clause (long distance reflexivization). As for conjunction reduction, this entails that the subject may remain unrealized in a clause which is conjoined with the main clause.

From the abundant literature on this topic, several semantic predicate types that call for non-canonical marking of the subject can be identified. As observed by Shibatan (2001: 312), they center around the following semantic fields:[7]

(i)   Possession/Existence; Attitude: be indifferent, be OK for, be impossible,
(ii)   Psychological states; (emotion, cognition, social interaction ((not) get along)
(iii)   Physiological states;
(iv)   Visual/auditory perceptions, including the notion of 'appearance'/ 'seeming';
(v)   Modal states of necessity and wanting, including the notion of obligation ('must');
(vi)   Modal states of potentiality, including ability and the notion of permission ('may');
(vii)   Uncontrolled events; e.g. forgetting, finding, etc. + gain, benefit, fail, and get hindrance.

Note that, in most of the languages showing non-canonical marking of the subject, non-canonical constructions draw the relevant predicates from these semantic fields. The constructions formed are fairly productive and the predicates involved are mostly either simple verbs or compositional predicates consisting of the verb 'be' together with adjectives or nouns. What is common to all these predicate- and argument structure

---

[7] For a more detailed classification, see Barðdal et al. (2012).

constructions is that they express states rather than activities, which can differ aspectually, since inchoatives as well as cessatives can be found (cf. also Barðdal et al. 2012). This includes possessive constructions, as well as psychological and happenstance verbs.

The first studies on non-canonical subjects dealt with South-Asian languages (Japanese, Korean, Vedic Sanskrit, Hindi-Urdu, etc.), on the one hand, and Icelandic, on the other hand. In South-Asian languages, the non-canonical subject is expressed by an experiencer in the dative and shows subject honorification (i.e. a certain feature of a nominal is reflected on the verb morphologically; also regarded as an instance of agreement cf. Kishimoto 2012). Subject honorification is an important morpho-syntactic subject criterion specific to South-Asian languages (cf. Masica 1976; Verma 1976; Shibatani 1977; Klaiman 1980; Verma & Mohanan 1990; Yoon 2009). In these languages, besides subject honorification, other subject properties such as word order, binding of reflexives, and control have been shown to characterize the dative experiencer and to strengthen its analysis as subject (Shibatani 2001).

As for Korean, Yoon (1996, 2009) argues that control properties such as coreferential subject deletion in control infinitives and coordinate clauses, alongside with the above-mentioned subject honorification are reliable in diagnosing subject-like obliques as subjects. In these languages, the predicates allowing non-canonical marking cover most of the semantic domains given in the list above, except for the last one, the semantic field of uncontrolled events.

In Icelandic, oblique subjects have been found to show various syntactic properties, namely first position in declarative clauses, subject-verb inversion, first position in subordinate clauses, subject-to-object raising, subject-to-subject raising, long distance reflexivization, clause-bound reflexivization, control infinitives, and conjunction reduction (cf. Andrews 1976; Thráinsson 1979; Zaenen, Maling & Thráinsson 1985; Barðdal 2006). As for the predicates triggering non-canonical marking of the subject, Barðdal (2006b: 45) identified 13 different classes[8] covering approximately the same semantic fields as identified by Shibatani (2001), and listed above.

---

[8] The thirteen verb classes identified by Barðdal (2004, 2006a: 45) are the following:
  i)   Verbs of emotion: e.g., 'feel good/bad';
  ii)   Verbs of cognition: e.g., 'suspect', 'have in mind', 'remember';
  iii) Verbs of perception: e.g., 'taste', 'appear';
  iv)   Verbs expressing idiosyncratic attitudes, e.g., 'be indifferent', 'be (im)possible for sb.;
  v)   Verbs denoting bodily states: e.g., 'feel pain';
  vi)   Verbs denoting changes in bodily states: e.g., 'start to freeze';
  vii) Verbs denoting personal properties and innate tendencies: e.g., 'be natural for sb.', 'be typical for sb.';
  viii) Verbs of gain: e.g., 'benefit', 'receive';
  ix)   Verbs of success and/or performance: e.g., 'succeed';

Non-canonical marking of the subject also occurs in several native South-American languages, among others Imbabura Quechua (Hermon 2001). In this language, an accusative takes on behavioral subject properties, such as subject-to-subject raising, subject-to-object raising, coreferential subject deletion (EQUI), and control of subjects of infinitives in purpose clauses. Interestingly, it has been highlighted that, in Imbabura Quechua, the non-canonical marking of the subject divides across two patterns, each of them (most likely) representing a stage on the way to the canonical subject, as described by Seržant in his (2013) article.

The two patterns are: a) lexical experiencers and b) desiderative experiencers. It has to be noted that the latter type is not lexically selected, but receives a desiderative connotation when a desiderative affix is added to the subject argument. As shown by Hermon (2001), lexical experiencers cannot be the target of subject-to-subject raising, or of EQUI, while desiderative experiencers can. Nevertheless, lexical experiencers score as much as desiderative experiencers with regard to other behavioral properties such as subject-to-object raising, and hence, they both pattern with canonically marked subjects (Hermon 2001: 161-162). An explanation for this is that desiderative experiencers pass more subjecthood tests due to the higher degree of control and intentionality of the experiencer, as opposed to the lexical experiencers, which lack intentionality. This supports Seržant's (2013) claim that non-canonically marked subjects behave like canonical subjects when there is a higher degree of implication or intentionality of the experiencer, that is, when the experiencer takes, in some way, control over the event.

Turning to Indo-European (IE) languages, they are generally nominative-accusative languages with canonical subject marking. These features are considered to be typical for the Standard Average European languages, more particularly for the core members, i.e. West-Germanic (cf. English) and Gallo-Romance (cf. French), but not for peripheral IE language groups such as Baltic, Slavic and some of the Germanic languages (Haspelmath 2001). Indeed, in these peripheral language groups, non-canonically marked subjects are frequent and tend to take on at least a certain number of the subject properties. Some of the most studied languages with non-canonical subject marking are briefly discussed in what follows.

Within Germanic languages, Icelandic, Faroese, Old Norse, Old and Modern German (cf. among many others, Zaenen, Maling & Thráinsson 1985; Barðdal & Eythórsson 2003; Sigurðsson 2004; Barðdal 2011; Barðdal et al. 2012), Icelandic has been already mentioned in this section as the language in which non-canonically marked subjects are shown to be the closest to canonical subjects. As for the ancient languages, it is more difficult to claim

---

x)   Verbs of failure or mistake: e.g., 'fail', 'get hindrance';

xi)   Verbs of decline: e.g., 'deteriorate';

xii)   Verbs of ontological existence: e.g., 'be in a particular manner';

xiii)   Verbs of social interaction: e.g., 'be friends', 'not get along'.

the subject status of oblique constructions, mostly due to lack of data. However, for languages such as Old Norse, Old Swedish, as well as for Early Middle English and Gothic, several instances of control constructions have been documented. The available data for these languages lead to a subject analysis of the obliques, rather than to an object analysis (Barðdal et al. 2012: 515).

The situation in Modern German is more controversial. Analogous non-canonically-marked arguments have been argued to be objects in German (cf. Zaenen, Maling & Thráinsson 1985: 478; Andrews 2001; Sigurðsson 2004). However, Barðdal & Eythórsson (2006), and Barðdal, Eythórsson & Dewey (2019) show that out of the nine criteria found relevant in diagnosing obliques as subjects in Icelandic, six work for the non-canonical subjects in German as well. As for the other three, they require ellipsis of the oblique argument. It has been shown, however, that in German, there are special restrictions on contexts with ellipsis (Barðdal 2006: 90). Having in mind the construction-specific character of subjects, Barðdal (2006a: 90) suggests that the oblique construction in German is not as entrenched as in Icelandic, but it is certainly more entrenched than in English. The scholar shows that the oblique construction has a lower type and token frequency in German than in Icelandic, being, hence, less fixed in the speaker's mind. As she explains, this affects the oblique argument's ability to be left unexpressed in German, limitation enhanced also by the special restrictions on elliptic contexts. Nevertheless, the oblique arguments in German are as much oblique subjects as in Icelandic (Barðdal 2006: 90).

Among Slavic languages, Russian, Slovenian and Bulgarian have been studied so far (cf. Greenberg & Franks 1991; Moore & Perlmutter 2000; Sigurðsson 2002; Fleisher 2006; Madariaga 2011; Paykin & Van Peteghem 2017; for other Slavic languages Rivero 2003, 2009). Russian, for instance, presents three types of dative subject constructions, according to Moore & Perlmutter (2000): i) the *governed I-construction*, where the dative is governed by particular predicates, such as *žal'* 'pity', *nužno* 'necessary', or *nravit'sja* 'please'; ii) the *infinitive dative subject construction*, and iii) the DIRC, or the *Dative Impersonal Reflexive Construction* as labeled by Benedicto (1995), also called the *productive I-construction.*

The main arguments for the subject analysis of the dative in the above-mentioned patterns are the following: the dative NP controls the anaphoric reflexive *sebja* '~self' or the possessive reflexive *svoj* '~own', the dative can control the covert PRO-subject of gerund and of infinitive in purpose clauses, of the type 'in order to + inf'. However, both Greenberg & Franks (1991) and Moore & Perlmutter (2000) argue that the infinitive dative subject construction (cf. (ii) above) is the only one that contains a real dative subject in Russian. In the other two constructions, the dative nominals are not surface subjects but surface indirect objects, called Inversion (I)-Nominals by Moore and Perlmutter (2000). More recently, Paykin & Van Peteghem (2017) argue that the Russian dative reflexive construction (cf. (iii) above), which enters in case variation with a nominative

construction, subsumes two subtypes: NOM-V, taking one argument verbs such as *rabotat'* 'work', and DAT-V.REFL-STIM$_{NOM/PP/CL}$ occurring with two-argument verbs such as *dumat'* 'think'. In both subtypes, the shift from nominative to dative corresponds with a decrease in subject agentivity.[9]

Among Baltic languages, non-canonically marked subjects have been studied in Latvian and Lithuanian (cf. Seržant 2013, 2015; Holvoet 2015). Holvoet (2015) shows that, in Latvian, most of the oblique candidates to subject position pass only the behavioral test of control of reflexives, out of the four tests he considers to be relevant in diagnosing a subject: case, agreement as coding tests, and control of reflexives and pivot/ controller in coordination as behavioral tests. The scholar identifies in Latvian three possible structures with a non-canonical subject in the dative: the *patikt* 'please' construction DAT-V-NOM, the *vajadzēt* 'be needed' construction DAT-V-DAT/GEN, and the *debitive*[10] construction DAT-V-NOM/ACC. Notably, these obliques co-occur with NPs in the nominative, but also with NPs in other cases such as dative, genitive, and accusative. The scholar introduces the concept of *recoverable subjecthood*, that is, the existence of some relevant subject (coding) properties encoded in the second argument (Holvoet 2015). The first pattern, DAT-V-NOM, shows *recoverable subjecthood* through the nominative encoding of the second argument, the second pattern, DAT-V-DAT/GEN,[11] lacks totally this possibility, whereas in the *debitive* pattern, the second argument can only partially recover the subjecthood. Holvoet concludes that the best way of accounting for the three non-canonically marked patterns is through the notion of *diffuse subjecthood*, that is, subject properties that are spread over several NPs (Holvoet 2015: 318).

As for Lithuanian, Seržant (2013, 2015) has shown that non-canonically marked subjects are datives occurring in at least two oblique constructions: the *verbs of pain* construction EXP-V-STIM and the *predicative* construction EXP-BE-NOUN, the latter being,

---

[9] It has to be noted that similar constructions with case variation are analyzed in Icelandic (Barðdal 2004) from a pragmatic perspective. She argues that the issue is not "agent" vs. "experiencer", but pragmatic, since it has to do with the subjective stance of the speaker.

[10] In the literature, the debitive is considered a mood or a modal affix (De Haan 2006: 36–37). Nau (1998: 39–40) suggests to consider it as a voice, alongside passive. Acknowledging this similarity with passive, Holvoet (2015) considers that the debitive construction, which comprises both A and B arguments, is more similar to the active one, in that, the nominative-marked NP (the original object) is normally the second-ranking argument, whereas the dative NP (the original subject) is least oblique and normally assumes the position of unmarked topic (Holvoet 2015: 318).

[11] With an animate referent, the genitive seems to be replaced by the accusative in Modern Latvian, although they both can still be found in the same sentence (Holvoet 2015: 308). As Holvoet (2015) suggests, this process could be seen as one of transitivization, following the example of the English *like*. This verb selected in Old English a dative experiencer, but has a nominative one in Modern English (Allen 1986: 390), being still far from a canonical subject, which would require the structure NOM-V-ACC. Note, however, that Barðdal (2008: 168) had explained this case change as being the result of synonymy at a specific language stage. She advocates that the most entrenched construction, i.e. the one highest in type frequency, will resist the change longer.

possibly, an instantiation of the MIHI EST pattern, in my opinion. The scholar emphasizes that they take on a limited number of subject properties, such as control of reflexives and the first position in an unmarked word order, whereas such subjecthood tests as control of PRO in infinitival complements and conjunction reduction are considered not to hold due to difficulty of finding examples (Seržant 2015: 337).

In Finnish, two constructions are considered to have non-canonical subjects: *Existential* clauses with partitive *subjects* and *Necessive* clauses with genitive *subjects*. For the partitives, it has been argued by Hakulinen (1983) that they are subjects, just not prototypical ones. Karlsson (1982: 109) argues that the subject-object distinction is simply neutralized, whereas Vilkuna (1989) and Sands (2000) argue that partitives are rather objects than subjects. With regard to genitives, these structures had been analyzed first as subjectless or impersonal, lacking a prototypical subject (Laitinen 1997: 111). The most common analysis of this genitive is as a *real genitive subject* (see, for example, Hakulinen & Karlsson 1979: 158, 172; Sulkala & Karjalainen 1992: 210–211). Sands & Campbell (2001) consider that these genitives enter in two different types of constructions and can thus be analyzed in two different ways: (i) either as *genitive dative-adverbials* (oblique NPs) occurring in a clause without an overt subject, or (ii) as the *subject of the infinitive* and not of the main verb, in which case they represent the canonical marking for non-finite verb forms (Sands & Campbell 2001: 269-279).

When it comes to Romance languages, non-canonically marked subjects have been discussed mostly for Spanish by several scholars,[12] but also for Modern and Old French,[13] for Italian,[14] and for Romanian.[15] Certain dative experiencers in Romance indeed show several of the subject properties, such as binding of reflexives, control into gerunds and raising. They also meet additional subject tests relying on quantifiers and scope, proposed in generative approaches, such as Cuervo (2010), based on work by Belletti & Rizzi (1988), McCloskey (1997), Cuervo (1999), Fernández-Soriano (1999).

Yet, Romance non-canonical subjects, instantiated in almost all languages by dative experiencers and rarely by accusatives, show important differences with the exemplary Icelandic oblique subject in that they do not meet certain crucial subject criteria, such as conjunction reduction and dative PRO. Therefore, controversy exists in the literature as to whether these arguments may be analyzed as syntactic subjects, or rather as I-Nominals, i.e. underlying subjects that surface as indirect objects, as argued within the framework of Relational Grammar (cf. González 1988, cf. also Moore & Perlmutter 2000

---

[12] Among others, Masullo (1993), Fernández-Soriano (1999) Rivero (2004), Diaconescu & Rivero (2007), Gutiérrez-Bravo (2006), Vázquez-Rosas (2006), Cuervo (2010), and Melis & Flores (2013).

[13] Cf. Legendre (1990) for Modern French, and Mathieu (2006) for Old French.

[14] Cf. Belletti & Rizzi (1988) and (Benedetti 2013a).

[15] Cf. Cornilescu (2009), Caluianu (2013), (Van Peteghem 2016), Van Peteghem (2017), Van Peteghem & Ilioaia (2017), and Ilioaia 2020.

for Russian). According to certain linguists, several of the properties taken on by these non-canonically marked arguments are, in fact, to be explained by the topicality of the oblique argument, rather than by its subject status and are therefore not appropriate subject tests (cf. Haspelmath & Caruana 2000; Onishi 2001; Gutiérrez-Bravo 2006). This claim demonstrates to be very controversial, as shown in recent papers (for a discussion of this issue, see Smitherman & Barðdal 2009; Barðdal & Eythórsson 2018).

Spanish has the so-called *stative construction* (cf. Elvira 2011: 6), with a dative nominal in clause initial position and a post-verbal nominative: DAT-V-NOM, as in (9). This construction has been shown to expand in Spanish (Melis & Flores 2013), contrary to the tendency to canonical marking witnessed in the core SAE languages.

(9) *A*   *Pedro*   *le*   *gustan*   *los*   *helados*
     DAT   Pedro   him.DAT-   please.3PL   the   ice_cream.PL
     'Pedro likes ice cream'

The first argument in this construction usually refers to a human and is marked in the dative. The second argument is frequently inanimate, is marked with a nominative, and always triggers verb agreement. Throughout the centuries the dative nominal has been used in combination with a coreferential pronoun that has ended up being mandatory in this construction in present-day Spanish (cf. the phenomenon of *clitic doubling*, see Elvira 2011: 189). The dative argument shows subject properties, such as raising, initial position in interrogatives, secondary predication, quantification, and nominalization, and has been argued to instantiate a typical case of non-canonical subject (Masullo 1993; Campos 1999; Fernández-Soriano 1999; Cuervo 2010; Elvira 2011). Elvira (2011) points out that, in fact, both the dative and the nominative arguments are non-canonical. Usually described as the subject because it agrees with the verb, the nominative argument shows at least two object properties: it is post-verbal in unmarked sentences, and it can be used without a definite article. Observing the non-canonical behavior of these two arguments in one and the same construction, Elvira (2011) underlines that the definition of the term *non-canonical argument* must be related to the construction it belongs to.

Modern French, one of the core SAE languages, displays structures with the verbs *manquer* 'lack' and *rester* 'remain' of the type *il*-DAT-V-NOM (cf. Legendre 1989). This impersonal construction systematically codes situations, in which the deficiency of a given entity, with respect to its ideal counterpart, is quantified or evaluated (Achard 2015: 131). Mathieu (2006: 288) considers them rather I-nominals in the sense of Moore & Perlmutter (2000) and not genuine non-canonical subjects. It is worth mentioning the existence of another structure, with an inchoative meaning, in which a dative experiencer co-occurs with the verb *venir* 'come' or *prendre* 'take' and a state noun (10).

(10) *Il*   *me*   *vient/*   *prend*   *l'*   *envie*   *de*   *cuisiner*

     it   me.DAT   comes /   seizes   the   craving   INF   cook

     'I begin feeling like cooking'

Not surprisingly, the impersonal construction in Modern French in (10) is a remnant of a real quirky subject construction from Old French, studied by Mathieu (2006). Labeled as impersonal constructions in the traditional literature, such structures typically involve an empty subject position filled by a dative or an accusative (11)-(12). Verbs like *plaire* 'please' are used with a dative-marked subject, whereas *convenir* 'to suit/to be convenient' selects for a quirky argument in the accusative, which, eventually is replaced by a dative, through the well-known process of *Dative Sickness*, that is, the systematic replacement of accusative experiencers by dative experiencers driven by thematic/semantic considerations (Eythórsson 2000, 2002; Barðdal 2011).

(11) *Plest*   *vos*   *oïr*   *de*   *une*   *corneile ...*

     please.3SG   you.DAT   hear.INF   of   a   carrion-crow

     'Do you want to hear the story about a carrion crow?'

                     (Guillaume d'Angleterre, 12th c., 40, 1, ex. 22 in Mathieu 2006: 289)

(12) *Talent*   *li*   *prist*   *d'*   *aler*   *chacier*

     eagerness   him.ACC   took.3SG   COMP-   go.inf   hunt.INF

     'He felt like going hunting'   (Lais, year 1160, line 76, ex. 36b in Mathieu 2006: 293)

Mathieu shows that, in Old French, the non-canonically marked arguments are genuine quirky subjects, just as the ones in Icelandic. He bases his claim on the following diagnostics: first position in main and in embedded clauses, control of reflexives, subject-verb inversion, control (i.e. being the controlled argument), conjunction reduction, and subject-to-subject raising (Mathieu 2006: 293).

For Italian, Benedetti (2013a) has shown that non-canonically marked subjects occur in clauses with noun predicates. She investigates constructions of the type DAT-V-NOM, as in (13), bearing an inchoative meaning. This construction is characterized by the combination of the light verb *prendere* 'to take' with a dative experiencer which displays syntactic subject behavior, and with a postverbal bare noun. Subject properties, such as preverbal position in unmarked contexts and a variety of syntactic control phenomena, show that the experiencers encoded in the dative take on the function of subject.

(13) *Gli*   *prese*   *paura*

     him.DAT   took.3SG   fear

     'He got scared'

Benedetti's (2013a) analysis leads to the postulation that these datives are impersonal "inversion" structures, or I-Nominals, as described by Moore & Perlmutter (2000). Benedetti (2013a) suggests that the inchoative '*gli prese paura*-construction' traces back to

the Latin *capere* 'seize, take' used with an experiencer and a state noun.[16] Van Peteghem (2017: 4) also signals the existence of a pain-construction of the type DAT-V-NOM_BP occurring with a dative experiencer, a verb of pain and a noun expressing a body part, in the nominative in Italian. This construction witnesses serious rivalry from a competing construction, which occurs with a nominative experiencer and the verb *avere* 'have'.

Romanian, unlike modern French, shows a particular fondness for non-canonically marked arguments showing subject behavior (cf. Bossong 1998 and Haspelmath 2001). Structures of the type DAT-V-NOM, (14), as well as complex predicate constructions containing the verb *fi* 'be', DAT-*fi*-NOM (15), DAT-V-PP (16), or ACC-V-NOM (17) exist in present-day Romanian, and they existed in old Romanian too.

(14) *Îmi*   *place*   *muzica*
     me.DAT   pleases   music.the
     'I like music'

(15) *Mi-*   *e*   *frig*
     me.DAT=   is   cold
     'I feel cold'

(16) *Îmi*   *place*   *de*   *Victor*
     me.DAT   pleases   of   Victor
     'I like Victor'

(17) *Mă*   *interesează*   *muzica*
     me.ACC   inerests   music.the
     'I'm interested in music'

Depending on the framework, scholars analyze them either as subjectless structures, as structures with a postverbal subject (Pană Dindelegan 2013a), or as constructions with non-canonical subjects (Dumitrescu & Masullo 1996 cited in Rivero 2009; Cornilescu 2009; Dobrovie-Sorin 2013). The main arguments put forward in the literature in favor of the subjecthood of the oblique arguments are first position in declarative clauses, raising, binding, conjunction reduction, weak cross-over, and bare quantifiers in clause-initial position.

---

[16] The Latin *capere* structure is considered to be the inchoative version of the Latin non-inchoative constructions "*esse* + dative" by Benedetti (2013b: 134).

2.4   Diachronic evolution of non-canonical subjects

As oblique subjects seem to be more frequent in old Indo-European languages than in many contemporary languages (Barðdal & Eythórsson 2012), some attention has been paid to the diachronic evolution of these constructions. A much-debated issue is whether these oblique experiencers have acquired subject properties during later stages, as suggested by Seržant (2013), or held subject properties in early stages, as argued in research works from the last decades (see Barðdal & Eythórsson 2003; Barðdal & Eythórsson 2012; Eythórsson & Barðdal 2005; Danesi, Johnson & Barðdal 2017; Danesi & Barðdal 2018; Pooth et al. 2019).

In Icelandic and other Germanic languages, three evolutions have been identified. The first one concerns the change of non-canonical subject marking into canonical subject marking, a phenomenon often called *Nominative Sickness.* Hence, dative or accusative experiencers become nominative experiencers and trigger verb agreement. As shown by Eythórsson (2000), and by Barðdal & Eythórsson (2003), this evolution is clearly observed in all Germanic languages, although to a varying degree. It is generally considered to be motivated by syntax in that it involves a change from inherent case on subjects to structural nominative case. The general underlying mechanism is that experiencers tend to be gradually assimilated to subjects with respect to their morpho-syntactic behavior because they refer to definite human participants and are topical (cf. Haspelmath 2001; Melis & Flores 2013).

A second evolution frequently observed is the substitution of the accusative case with the dative, a change, which is generally called *Dative Sickness* (cf. Eythórsson 2000; Jónsson 2003; Jónsson & Eythórsson 2005). Contrary to *Nominative Sickness*, this change is not motivated by syntax, but by semantics. *Dative sickness* is found in Icelandic and has been explained by an evolution of idiosyncratic into thematic case. The accusative, which is less frequent and completely unpredictable, is considered as idiosyncratic, whereas the dative is linked to particular thematic roles, namely goals, recipients, experiencers.

Barðdal (2011) argues against the assumed underlying mechanism for both evolutions. She presents them both as a result of the productivity of the lasting construction, due to high type frequency.

As for the third scenario, the same scholar suggests that alternating predicates stop being alternating and that one of the alternants survives - sometime different ones in different languages or dialects. Such development has been documented for the Scandinavian languages (Barðdal 1998). A clear example is the evolution of *seem* in English, where only *this seems to me*, has survived until the present-day English, whereas the earlier *me seems this* and *this seems me* have disappeared. Similar developments are documented in Icelandic and German (Barðdal, Eythórsson & Dewey 2019).

In Romance languages, oblique subjects are considered to have their roots in Latin, which is claimed by Bossong (1998) and Haspelmath (2001) to belong to the SAE type. Indeed, Latin had a rather substantial group of impersonal verbs (76 types are documented in the *NonCanCase* database),[17] among which some select an accusative, whereas other a dative human participant, as in (18a-b) (see Fedriani 2013).

(18) a. *Me   pudet*
        me.ACC   shames
        'I feel ashamed'
     b. *Ita   nobis   decet*
        so   us.DAT   befits
        'So befits us'

As shown by Mathieu (2006), Old French still had non-canonical marking, which is nearly lost in Modern French, an exception being the archaic structure in (19). This evolution is consistent with the tendency to canonical subject marking claimed to be typical for SAE languages (cf. Seržant 2013 and Haspelmath 2001).

(19) *Il   me   souvient   de   ces   jours....*
     it   me.DAT   remembers   of   these   days
     'I remember these days'

However, Elvira (2011) and Melis & Flores (2013: 168) argue that Spanish evolves into the opposite direction, and shows a gradual extension and an increasing productivity of oblique subjects. This inverse process in Spanish is attributed by these authors to certain typological properties that play an important role in the expansion of this pattern, such as the Animacy Hierarchy, the position of the object clitics, flexible word order rules, and *external possessor* strategy. Interestingly, all these properties are also found in Romanian.

A double tendency can be observed among the SAE languages: toward canonical subject marking (in core SAE languages), and toward dative marking of the subject (in Spanish and Romanian). This dual evolution has made linguists dig deeper until Proto-Indo European (PIE) in order to understand the origins of non-canonical marking of core arguments.

The traditional position that PIE was a language of the accusative alignment type (Drinka 1999) has been progressively losing ground, in favor of an analysis involving a split-S system in more recent studies (cf. Kortlandt 2001; Matasović 2012; Willi 2018 cited in Pooth et al. 2019). However, Pooth et al. (2019: 5) argue that the later accusative system developed from a semantic alignment system. Furthermore, they claim that this semantic

---

[17] The *NonCanCase* database is one of the products of the ERC Research project EVALIZA (The Evolution Of Case, Alignment And Argument Structure in Indo-European) 2013–2018, funded by the European Research Council, and directed by Jóhanna Barðdal.

alignment system gave rise to non-canonical subject marking in the IE daughter languages. Since this matter is beyond the scope of this study, I refer the reader to the references provided above for further details.

The following chapter focuses on the ways the subject has been perceived in Romanian, on the one hand, and on the way linguists have dealt with non-canonical subject marking in this language, on the other hand.

Chapter 3   Subject properties in Romanian

In order to identify the characteristics of the subject in Romanian, I will rely, on the one hand, on the set of properties put forward by Keenan (1976), and, on the other hand, on the theory-independent approach suggested by (Barðdal & Eythórsson 2003, 2012), and Eythórsson & Barðdal (2005). First, I will discuss the coding properties of canonical subjects in Romanian, and then I will resort to syntactic properties in order to explore the behavior of canonical and non-canonical subjects in Romanian.

The outcomes of the approach based on the subject properties will be verified through the definition proposed by Eythórsson & Barðdal (2005). In their search for a theory-independent definition of subjects, the two scholars propose a non-scalar notion of subject defined as the leftmost argument of the argument structure (cf. Eythórsson & Barðdal 2005: 829, already referred to in Section 2.1, in Chapter 1). Note that, although in certain languages such as German the leftmost argument of the argument structure mostly coincides with the first argument in a regular active declarative clause with neutral word order (Eythórsson & Barðdal 2005: 829), this is not always the case. In terms of event structure, from which the argument structure derives (cf. Croft 1998, 2012; Barðdal 2001, among others), the first and leftmost argument is considered as the originator of the event (see also Barðdal et al. forthcoming). With certain non-agentive verbs, as is the case with mental state verbs, the originator of the event is not easy to identify, given that it depends on the degree of control each participant has over the event. In spite of this difficulty, I credit this definition of subject, because it is not dependent on any theoretical framework, nor on language-specific or construction-specific properties of subjects. Yet, in my study, I will complement it with a close investigation of a set of the subject properties peroposed by Keenan (1976).

This chapter is organized as follows. Section 3.1 focuses on the coding properties of canonical subjects in Romanian, more particularly on their morpho-syntactic properties, i.e. case marking and verb agreement. Section 3.2 deals with the syntactic behavior of subjects in Romanian and examines to what extent accusative and dative experiencer arguments behave like subjects from a syntactic point of view and may, hence, be analyzed as non-canonical subjects.

3.1 Coding properties of the subject in Romanian

Among the three coding subject properties distinguished by (Keenan 1976), two are considered as basic subject properties in the grammatical tradition of Romanian, namely nominative case and verb agreement. As for the third coding property, i.e. word order, I will consider it as a behavioral property and discuss it in Section 3.2 (cf. Le Mair et al. 2017).

3.1.1 Nominative case

The nominative case is traditionally considered the most basic property of the subjects in Romanian. In the examples in (20) and (21), *băiatul* 'the boy' and respectively *copilul* 'the child' are the subject and are encoded in the unmarked nominative case.

(20) ***Băiatul***   *citește*   *o*   *carte*
     boy.the   reads   a   book
     'The boy is reading a book'

(21) *S-*   *a*   *îmbolnăvit*   ***copilul***
     CL.REFL.ACC   has   become-sick   child.the
     'The child got sick'

It is generally assumed that Romanian has five cases: nominative, accusative, genitive, dative, and vocative. In more recent studies, Romanian inflectional case marking is said to distinguish between Vocative: VOC *băiete*! 'boy!', Genitive-Dative (or Oblique): GEN-DAT *băiatu-lui* 'of/to the boy', and an unmarked form (Nominative-Accusative): NOM-ACC *băiat* 'boy', used for subjects, objects (other than indirect objects, and prepositional direct objects) and complements of (most) prepositions. (Dobrovie-Sorin & Giurgea 2013: 11).

Morphological case is marked on NPs, more particularly on the determiner of the NP, and on all types of pronouns, i.e. strong, weak and clitic pronouns.[18] However, morphological marking of case is limited, especially on NPs. Indeed, for NPs the Romanian case system shows syncretism between the nominative and the accusative, on the one hand, and the genitive and the dative, on the other hand. In other words, the nominative NP *cartea* 'the book', in (22a) below, is homonymous with the accusative in (22b). The accusative is only marked with definite NPs and strong pronouns denoting humans through the presence of the preposition *pe*, as is shown in (22c). Like Spanish,

---

[18] The terminological distinction between weak pronoun forms and pronominal clitics becomes less important and is replaced by the term *clitics* (cf. GLR 2005, Dobrovie-Sorin & Giurgea 2013: 233-234).

Romanian has a prepositional accusative, and, hence, shows differential object marking (cf. Mardale 2008; Dobrovie-Sorin & Giurgea 2013: 11, and Hill & Mardale 2017).

(22) a. ***Cartea***   *a*   *căzut*
        book.the   has   fallen
        'The book fell'

vs.   b. *Am*   *citit*   ***cartea***
        have.1SG   read   book.the
        'I read the book'

      c. ***L-***   *am*   *văzut*   ****(pe)***   ***băiat***
        him.ACC   have.1SG   seen   ACC   boy
        'I saw the boy'

The same syncretisms (i.e. nominative/accusative vs. genitive/dative) hold for most strong pronouns, such as demonstratives, relatives, and indefinites. Hence, a pronoun such as *cineva* 'somebody' has only two forms: NOM/ACC *(pe) cineva* 'somebody' vs. GEN/DAT *cuiva* 'somebody's / to somebody'). However, with personal pronouns, case marking is different from NPs and other types of strong pronouns. Strong personal pronouns have a different form for the nominative and the accusative in the 1st and 2nd person singular (NOM *eu* vs. ACC *mine*; NOM *tu* vs. ACC *tine*), as shown in Table 3.1, whereas weak forms and clitics lack a nominative form, and have different forms for the accusative vs. the genitive[19]/dative as in Table 3.2. Hence, like in many other languages, personal pronouns have the most explicit case marking, especially weak pronouns and clitics.[20]

---

[19] Clitics may also be hosted by nouns or prepositions, but this is rather marginal, and considered archaic by by Dobrovie-Sorin & Giurgea (2013). Confusion may then arise between the genitive and the dative form, since these forms are identical in Romanian (i)-(ii). Nevertheless, the genitive clitic still occurs in the present-day language with the adjective *propriu* 'own' as in (iii) (examples (14a–c) from Dobrovie-Sorin & Giurgea 2013: 237):

(i)   *împotriva*   *vieții/*   *împotriva*   *-i*
      against   life.the.GEN / against   =her.GEN
      'against life/ against it'

(ii)   *sufletul*   *omului/*   *sufletu*   *-mi*
      soul.the   man.the.GEN /   soul.the   =mine.GEN
       'the man's soul/ my soul'

(iii)   *propria*   *-mi*   *mamă*
      own.the   =mine.GEN   mother
      'my own mother'

[20] The GLR (2005) choses to utilize the same term, *clitic*, for both the *weak* and the *clitic* form of pronouns (cf. also Footnote 18), and distinguishes between *free clitics* (the weak forms) and *conjunct clitics* (the clitic forms). Since this distinction is not relevant in the present study, I will use, in the following sections, the term *clitics* for both the free and the conjunct forms.

Table 3.1   Declension personal pronouns – strong forms

Case	Singular				Plural			
	1st Pers	2nd Pers	3rd Pers		1st Pers	2nd Pers	3rd Pers	
			M/N	F			M	F/N
NOM	eu	tu	el	ea	noi	voi	ei	ele
ACC	(pe) mine	(pe) tine	(pe) el	(pe) ea	(pe) noi	(pe) voi	(pe) ei	(pe) ele
DAT	mie	ție	lui	ei	nouă	vouă	lor	lor
GEN	–	–	(al) lui (a) lui (ai) lui (ale) lui	(al) ei (a) ei (ai) ei (ale) ei	–	–	(al) lor (a) lor (ai) lor (ale) lor	(al) lor (a) lor (ai) lor (ale) lor
VOC	–	tu!	–	–	–	voi!	–	–

Table 3.2   Declension personal pronouns – weak forms, followed by clitics

Case	Singular				Plural			
	1st Pers	2nd Pers	3rd Pers		1st Pers	2nd Pers	3rd Pers	
			M/N	F			M	F/N
ACC	mă, m-	te, te-	îl, l-	o, -o	ne, ne-	vă, v-	îi, i-	le, le-
(GEN)/ DAT	îmi, mi, mi-	îți, ți, ți-	îi, i, i-	îi, i, i-	ne, ni, ne-	vă, vi, v-	le, li, le-	le, li, le-

As mentioned in the beginning of this section, the nominative case is the case of the canonical subject, in Romanian. The absence of a nominative form for weak pronouns and clitics is related to the fact that Romanian is a pro-drop language, in which the subject is left unrealized, but is retrievable from the rich verb inflection and from the context.

(23) *Citeşte /*   *Citim*   *aceeaşi*   *carte*
     reads /   read.1PL   same   book
     '(S)he is/ We are reading the same book'

In the example in (23), for instance, the nominative subjects 'he/ she', and 'we', respectively, can be retrieved from the verb inflection for the 3rd person singular *-eşte*, in *citeşte* 'he/she reads' and for the 1st person plural *-m*, in *citim* 'we read'.

3.1.2 Verb agreement

Verb agreement, the second coding property of subjects put forward by Keenan (1976), constitutes in Romanian another basic subject criterion for canonical subjects. Indeed, all finite verbs agree in person and number with the nominative constituent, which may be preposed or postposed to the verb. In the examples in (24a-b), *băiatul* 'the boy' and *copiii* 'the children' are in the nominative case and trigger verb agreement in number and person.

(24) a. ***Băiatul***   *citeşte*   *o*   *carte*
      boy.the   reads   a   book
      'The boy is reading a book'
    b. *S-*   ***au***   *îmbolnăvit*   ***copiii***
      CL.REFL.ACC=   have.3PL   became-sick   children.the
      'The children got sick'

However, in Romanian, nominative nominals may also co-occur with non-finite forms of the verb, such as infinitives or gerunds, which do not agree with the nominative nominal (25)-(26). The only non-finite form showing agreement with its nominative subject is the past participle, which agrees with it in number and gender (27) (Pană Dindelegan 2013a: 101).

(25) *înainte*   *de*   ***a***   ***ajunge***   ***profesorul*** /   ***eu*** /   ***noi***
      before   of   INF   arrive   teacher.the /   I /   we
      'before the teacher / I / we arrived'
(26) *Ion*_i   *a*   *reuşit*   ***trudind***   ***alţii***_j /   ***alţii***_j   ***trudind***
      Ion   has   succeeded   working   others.NOM /   others.NOM   working
      pentru   *el*
      for   him
      'Ion succeeded by others toiling for him'   (example (7c), Pană Dindelegan 2013a: 103)
(27) *Odată*   ***plecat***   ***directorul*** /   ***plecaţi***   ***părinţii,***   *a*
      once   gone.MASC.SG   director.the /   gone.MASC.PL   parents.the.   has
      şi   *început*   *vacarmul*
      already   started   racket.the
      'The director / parents having left, the racket had already started'
                          (adapted from ex. 10a, Pană Dindelegan 2013a: 104)

The co-occurrence of non-finites with nominative subjects casts doubt on the claim that nominative assignment is dependent on agreement. Therefore, several scholars have argued that nominative assignment is dependent on tense rather than on agreement (cf. Haeberli 1999, 2002; Pesetsky & Torrego 2001; Rivero & Geber 2004; Alboiu 2005; Mathieu 2006).

3.2   Behavioral properties of subjects in Romanian

Turning to behavioral properties of subjects, I investigate in the present section to which extent these properties apply in Romanian to canonical as well as to non-canonical subjects. Therefore, the relevance of the following syntactic properties will be explored: first position in declarative clauses, raising, control, binding, conjunction reduction, deletion in imperatives, and deletion in telegraphic style, bare quantifiers in clause-initial position, and secondary predication.

Several scholars consider that syntactic properties are more basic than morphological encoding and, therefore, argue that dative or accusative subject-like arguments are non-canonical subjects.[21] Nevertheless, non-canonical subject marking typically occurs with pragmatically salient arguments (topical, human, definite), which are highly ranked on the thematic scale, such as experiencers. Indeed, experiencers are often not encoded as nominative subjects, and do not trigger verb agreement in that case. Experiencers typically occur with psychological verbs, which have been the topic of many studies (Postal 1970; Guéron 1986; Belletti & Rizzi 1988; Montrul & Chen 1999; Alexiadou & Iordăchioaia 2014). With these verbs, the experiencer may be encoded as a canonical subject in the nominative, but it is mostly encoded in the dative or in the accusative, thus lacking the coding properties of a canonical subject. Whereas the experiencer argument displays several syntactic subject properties, the nominative-marked stimulus does not behave in all regards like a syntactic subject.[22]

Therefore, in what follows, I will investigate how canonical and non-canonical subjects behave with respect to syntactic subject properties in Romanian. This chapter is structured as follows. Section 3.2.1 gives an overview of dative and accusative experiencer constructions in Romanian, whereas the following sections examine how nominative subjects and non-nominative experiencers behave with respect to word order (3.2.2), binding (3.2.3), control of implicit subjects of non-finite verb forms (3.2.4), subject raising (3.2.5), conjunction reduction (3.2.6), deletion in imperatives (3.2.7),

---

[21] Some early work involves Andrews (1976) and Thráinsson (1979) on Icelandic, Masica (1976), Kachru, Kachru & Bhatia (1976), and Klaiman (1980), on South Asian languages, Elmer (2011, 1981), (Seefranz-Montag 1983, 1984), and Allen (1986, 1995) for Old English.

[22] Building on the theta-frames proposed by Fillmore & Kay (1993: 8–16), Croft (1998) suggests to assign other thematic roles to the participants of the EXP-STIM event type, in order to distinguish between EXP-STIM (*fear*, *enjoy*) and the STIM-EXP (*frighten*, *please*) theta-frames. He suggests replacing the term Experiencer with Attender and the term Stimulus with Content, since the stimulus is seen as a content toward which the experiencer directs his/her attention. Nevertheless, Croft admits that this is not the best solution, since doing so does not allow one to capture the similarity between the two event types (the EXP-STIM and the STIM-EXP), which he sees actually as alternative readings of the same conceptual event type (Croft 1998: 30–32). For the purpose of clarity, I will continue to utilize, in what follows, the term *stimulus*.

deletion in telegraphic style (3.2.8), bare quantifiers in clause-initial position (3.2.9), and secondary predication (3.2.10).

3.2.1   Dative and accusative experiencers in Romanian

In Romanian, several verb classes select for an experiencer, as is the case in other Romance languages. Since the experiencer necessarily denotes an animate and mostly a human entity, it is always encoded as an NP or a pronoun and, hence, it is always case-marked.

Like in Italian (cf. Belletti & Rizzi 1988), three classes of psychological verbs may be distinguished in Romanian, depending on the case marking of the experiencer:

(i) The **temere** class (28), whose experiencer is encoded in the nominative (cf. *teme* 'fear', *iubi* 'love', *adora* 'adore, worship'). In terms of event structure, the experiencer coincides with the leftmost argument of the argument structure, since the experiencer has a certain degree of control over the event, of which it is the originator, even though this verb class contains non-agentive verbs (cf. Eythórsson and Barðdal 2005, Barðdal et al. forthcoming). The experiencer argument of these verbs is a canonical subject, hence, I will not further comment on this construction.

(28) *Mara   iubeşte   florile*
     Mara   loves   flowers.the
     'Mara loves the flowers'

(ii) The **preoccupare** class (29a–b), whose experiencer is in the accusative (cf. *preocupa* 'concern', *îngrijora* 'worry', *interesa* 'interest/concern', *mira* 'wonder', *speria* 'frighten', *plictisi* 'bore', *dezamăgi* 'disappoint'). In Romanian, this class also includes the so-called **pain** verbs (cf. *durea* 'ache', *ustura* 'burn, irritate', *arde* 'burn', etc.), which select for an experiencer encoded in the accusative (30a-b). In the context of the Romance languages, this construction is specific to Romanian since all other Romance languages encode the experiencer in the dative with these verbs (Van Peteghem 2017).[23]

(29) a. *Mă   îngrijorează   tristeţea   lui*
       me.ACC   worries   sadness.the   POSS.GEN
       'I worry about his sadness'

---

[23] In a broader context, accusative experiencers can be found also in other languages, among which Greek (cf. Anagnostopoulou 1999), German – which allows for both accusative and dative case marking of the experiencer with *schmerzen* 'to ache' (*mich/mir schmerzt der Kopf* 'I have a headache') - (cf. Seržant 2015: 341, Van Peteghem 2017), West-Slavic languages (Czech), Ukrainian and even Russian at some point in its history (Old Russian) (cf. Serzant 2015: 341).

b. *Mă*	*îngrijorează*	*să*	*te*	*văd*	*în*	*starea*	*asta*	
me.ACC	worries	SUBJ	you.ACC	see.1SG	in	state.the	this	

'I worry when I see you like this'

(30) a. *Mă*	*doare*	*capul*	
me.ACC	aches	head.the	

'I have a headache'

b. *Mă*	*doare*	*în*	*gât*	*rău*	*de*	*tot*
me.ACC	aches	in	neck	badly	of	all

'My throat hurts very badly'

Given that the theme in (29a), *tristeţea* 'sadness', is the *cause of the emotion* felt by the experiencer in Pesetsky's (1987) terms, it is expected to occupy the first position in the argument structure of the event. However, a closer look at the Romanian data reveals that the experiencer occupies a higher position than the theme. This is in line with what Belletti & Rizzi (1988) argue for Italian with regard to this class of psych verbs, namely that the leftmost argument of the event structure is the experiencer and not the theme. The two scholars show that the argument structure of the *preoccupare* verb class is in line with the ones of the other two verb classes. In their argumentation, they rely on a mapping principle, which is given in (31). Note that, in this principle, *higher* means *asymmetrically c-commanding*.

(31) *Given a θ-grid [Experiencer, Theme], the Experiencer is projected to a higher position than the Theme.*

Hence, the data for Romanian point toward the same positioning of the arguments in the event structure, with the experiencer as the leftmost argument, as described by Belletti & Rizzi (1988: 344).[24] Note that this is not the general view on Romanian structures illustrated in (29) and (30). In these examples, the stimulus is generally encoded as a nominative NP (cf. 29a and 30a), but it may also be encoded as a clause (29b). When it is a body part, it may also occur as a locative PP (30b). Most grammars of Romanian analyze the nominative NP in (29a) and (30a) as the subject, whereas constructions containing a clause (29b) or a locative PP (30b) are analyzed as subjectless (Pană Dindelegan 2013a: 111).

(iii) The third verb class is the ***piacere*** class, in which the experiencer occurs in the dative, and is clearly the first argument in the argument structure (32) (cf. *plăcea* 'please', *conveni* 'suit', *lipsi* 'lack'). The stimulus is mostly encoded as a nominative NP (32a), but it occurs also as a clause (32b) or as a PP (32c). Given the nature of the event, namely a

---

[24] Belletti & Rizzi (1988: 344) describe the syntactic structure of the three classes of psych-verbs as follows:

(i) *temere* 'fear' [Experiencer, Theme];
 preoccupare 'worry' [Experiencer, Theme];
 piacere 'please' [Experiencer, Theme].

mental state, it is difficult to establish a cleary defined argument structure, since none of the arguments seems to be the originator of the event. Hence, both the experiencer and the stimulus compete for the subject position. Belletti & Rizzi (1988) analyze the experiencer as the leftmost argument in the argument structure of this verb class. Indeed, the very fact that the stimulus may be encoded as a clause or a PP, thus losing the two coding subject properties, shows that it tends to occur in a lower position than the experiencer (32b) and (32c).

(32) a. *Îmi*   *plac*   *cărțile*
     me.DAT   please.3PL   books.the
     'I like books'
   b. *Îmi*   *place*   *când*   *plouă*
     me.DAT   pleases   when   rains
     'I like when it rains'
   c. *Nu*   *-mi*   *place*   *de*   *vecina*   *nouă*
     Not   =me.DAT   pleases   of   neighbor.the   new
     'I don't like my new neighbor'

Dative experiencers may also occur in constructions expressing bodily sensations, which are usually included in the pain verb class (33a-b). This is the case with verbs such as *țiui* 'ring', *vâjâi* 'howl', *amorți* 'numb', *bate* 'beat', *crăpa* 'split', *curge* 'run', *lacrima* 'tear', *pocni* 'hit', *tremura* 'shake', *trosni* 'crackle', etc., which, however, are not intrinsic pain verbs. According to Van Peteghem (2016: 17), the dative in this pattern is an argument of the body part noun, and not of the verb and must be analyzed as a possessive dative.

(33) a. *Îmi*   *lăcrimează*   **(ochii)*
     me.DAT   tear.3SG=3PL   eyes.the
     'My eyes tear up'
   b. *Îmi*   *țiuie*   *urechea /*   *urechile*
     me.DAT   ring.3SG=3PL   ear.the /   ears.the
     'My ear is ringing / My ears are ringing'

The above-mentioned verbs occurring with an experiencer carry themselves an experiential meaning. However, constructions as illustrated in (33a-b) are different in that they also contain an experiencer, although the verb does not belong to the class of psychological verbs. The experiential meaning is then conveyed by the specific combination of a dative experiencer with a complex predicate containing the verb and the nominative constituent. This type of structure is mostly known as a complex construction, since it contains a complex predicate.

By definition, a complex predicate comprises a light verb or a copula and a predicate (cf. Goldberg 1995; Michaelis & Lambrecht 1996: 242; Bickel 2004; Verhoeven 2007: 31–32; Le Mair et al. 2017, who use the term *compositional predicate)*. Another term which is used for these complex constructions is *non-congruent expressions* (34a-b) (Verhoeven 2007: 88). Verhoeven (2007) describes these complex constructions as

metaphorical expressions,[25] in which the verb refers to a different semantic domain than the whole expression does, and opposes them to *congruent expressions* (35a-b), as defined by Reh (1998: 11) as constructions containing psychological verbs such as *preoccupare* and *piacere* (cf. Reh 1998: 11, based on Halliday 1985).

(34) *Non-congruent expressions* or *complex constructions*

    a. *Ich   habe   Hunger*
       I   have.1SG   hunger
      'I am hungry'   (example 86a, Verhoeven 2007: 89)

    b. *Mi-   e   foame*
      me.DAT=   is   hunger
      'I am hungry'

(35) *Congruent expressions*

    a. *Pe   mama   o   preocupă   istoria   religiilor*
      ACC   mother.the   her.ACC   interests   history.the   religions.the.GEN
      'Mother is interested in the History of religions'

    b. *Îmi   plac   cărțile*
      me.DAT   please.3PL books.the
      'I like books'

In the following sections, the subject diagnostics are applied to nominative subjects and to non-nominative subject-like arguments. Among the non-nominative subject-like arguments, only the congruent constructions will be discussed, more particularly the ones pertaining to Belletti & Rizzi's classification, alongside with the pain construction in (30a-b). As for the non-congruent experiencer constructions, namely the ones occurring with a complex predicate, these are discussed in Chapter 4 of this dissertation.

3.2.2   Word order

As mentioned above, word order was put forward as a coding subject property by Keenan (1976). However, more recent studies tend to include this test among the behavioral properties of a subject (Le Mair et al. 2017). For the present study, I also consider word order among the behavioral properties of subject.

    As in all Romance languages, subjects tend to occupy the first position in most types of clauses in Romanian. This is indeed the case in the majority of declarative clauses and in several subordinate clauses, although not in all. However, the subject may also occur in postverbal position. In declarative clauses, the position of the subject depends on the agentivity of the verb and its argument structure (Pană Dindelegan 2013a: 121). For

---

[25] Note that, although the verb is used metaphorically, the construction instantiates the neutral and, sometimes, the only way of expressing that specific state/emotion in the language. Iorga (2014, Ch. 2) considers the experiencer constructions in general as the result of some spatial metaphore.

instance, it is more common for agentive verbs to occur with a preposed subject (36a) than for non-agentive verbs (36b). Furthermore, one-argument verbs tend to postpose their subject (36b), whereas verbs with two or three arguments tend to prepose it, as in (36a). Hence, the postverbal position of the subjects is generally related to the non-agentivity of one-argument verbs, whereas the preverbal position of the subject is linked to the agentivity of two or three argument verbs (Pană Dindelegan 2013a: 121).[26]

(36) a. *Artistul   pictează   un   tablou*
     artist.the   paints   a   painting
     'The artist paints a painting'
   b. *S-   a   îmbolnăvit   copilul*
     CL.REFL.ACC   has   become-sick   child.the
     'The child got sick'

Without arguing against this view, Giurgea (2017, cf. also Giurgea & Remberger 2012) shows that the postverbal position of the subject constitutes, in fact, a sign that the subject is not a topic. He suggests that postverbal subjects in Romanian are more conditioned than has been considered in the literature, many of them being actually marked elements. Furthermore, Giurgea identifies two possible conditions for subjects to occur in postverbal position: the so-called presentationals (37), and the structures with a topic different from the subject in the (immediately) preverbal position (38a), or with a narrow focus on the subject (38b). Presentational or thetic sentences are VS sentences without an overt topic,[27] or which take the spatio-temporal coordinates of the event as the topic. The structures with a topic different from the subject and the ones with narrow focus imply that either the subject or the direct object can fulfill the topic requirements.

(37) Presentationals
   *Plânge   **copilul***
   cries   baby.the
   '**The baby**'s crying.'   (Giurgea 2017: 283, ex. 9a)

---

[26] An exception to this are the unergatives in Romanian, which, although always agentive, behave like one-argument verbs (Pană Dindelegan 2013a: 121).

[27] Giurgea (2017: 292) points out that presentational sentences are also constrained, in that they only allow episodic predications, i.e. those anchored in the *here* and *now* of the current discourse, which constitutes the *stage topic*. Presentational sentences are disallowed in generic, iterative and I-level (Individual-level) predications, since they do not involve particular spatio-temporal locations, which can function as stage topics (see Soare 2009; Giurgea & Remberger 2009, 2012; Kiss 2002 for Hungarian). This can only contribute to a higher frequency of sentences with SV word order in Romanian.

(38) a. Preverbal topic, other than subject

Tabloul	*ăsta*	*l-*	*a*	*cumpărat*	***mama***	*mea*	*de_la*
painting.the	this	him.ACC=	has	bought	mother.the	mine	from

un *anticar*
an antiquarian

'This painting, my **mother** bought it from an antiquarian.'   (Giurgea 2017: 283, ex. 10)

   b. Narrow focus on the subject

Va	*vorbi*	***MARIA***	*cu*	*Ion* [Context: open issue = who will tell Ion?]
FUT.3SG	talk.INF	Maria	with	Ion

'**MARIA** will talk to Ion.'   (Giurgea 2017: 283, ex. 11a)

Pană Dindelegan (2013a) highlights other factors which may influence word order, such as the human vs. non-human character of the subject (39a-b), definiteness vs. indefiniteness (40a-b), and the rhematic nature of the subject (41).

(39) a. ***Copilul*** *a* *mâncat* *tot*
       child.the   has   eaten   all
       'The child has eaten everything'

   b. *A* *căzut* ***tavanul***
       has   fallen   ceiling.the
       'The ceiling has fallen'   (ex. 66b, in Pană Dindelegan 2013a)

(40) a. *Toamna* *cad* ***frunze***
       autumn.the   fall.3PL   leaves
       'In the autumn leaves fall'

   b. ***Frunzele*** *ruginii* *au* *căzut*
       leaves.the   rusty   have.3PL   fallen
       'The rusty leaves have fallen'

(41) A: *Cine* *ți-* *a* *cumpărat* *papucii?*
       who.NOM   you.DAT   has   bought   shoes.the

   B: *Mi* *i-* *a* *cumpărat* ***mama***
       me.DAT   them.ACC=   has   bought   mama.the
       'Who bought your shoes? Mother has bought them for me'

In subordinate clauses, supplementary restrictions occur both in finite and non-finite clauses. In non-finite clauses, when the subject is realized, it is obligatorily postposed with an infinitive (42) or a supine, but with a gerund or a participle it can be either pre- or postposed (43).

(42) *E* *important* *a* *decide* ***tu*** ***însuți***
     is   important   INF   decide   you   yourself
     'It is important for you to decide yourself'

(43) *Odată* ***(directorul)*** *plecat* ***(directorul),*** *a* *și*
     once   (director.the)   gone.MASC.SG   (director.the)   has   already
     început *vacarmul*
     started.3SG   racket.the
     'The director having left, the racket had already started'

In finite subordinate clauses, constraints are imposed by the type of clause. For instance, in a DP-internal relative clause, when the subject is not part of a *wh*-phrase, subject-verb inversion is possible, although not obligatory, as shown in (44).

(44) *Fata*	*căreia*	**(Victor)**	*îi*	*va*	*da*	**(Victor)**	*cartea*
girl.the	which.DAT	(Victor)	her.DAT	FUT.3SG	give.INF	(Victor)	book.the

(Victor)
(Victor)
'The girl whom Victor will give the book'

Moreover, different constraints apply in subordinate clauses depending on the complementizers. In a *că* 'that' + indicative clause, both SV and VS word order are possible (45). The same holds for the *ca (… să)* + subjunctive clause expressing a goal, where the subject occurs either preverbally, placed between *ca* and *să* (46a), or postverbally (46b). By contrast, the *să* + subjunctive clause allows only a postverbal subject (46c).

(45) *Ştiu*	*că*	**(artistul)**	*pictează*	*?(artistul)*	*un*	*tablou*	*?(artistul)*
know.1sg	that	(artist.the)	paints	(artist.the)	a	painting	(artist.the)

'I know that the artist paints a painting'

(46) a. *Mă*	*zbat*	*ca*	*(Mara)*	*să*	*ajungă*	*(Mara)*
me.ACC	struggle.1SG	COMP.SUBJ	(Mara)	SUBJ	arrive.3SG	(Mara)

'I struggle so that Mara can arrive'

b. *Mă*	*zbat*	*ca_să*	**(Mara)*	*ajungă*	*(Mara)*
me.ACC	struggle.1SG	COMP.SUBJ	*(Mara)	arrive.3SG	(Mara)

'I struggle so that Mara can arrive'

c. *Mă*	*zbat*	*să*	**(Mara)*	*ajungă*	*(Mara) /*
me.ACC	struggle.1SG	SUBJ	*(Mara)	arrive.3SG	(Mara) /

să	*-i*	*dea*	*cartea*	*(Mara)*
SUBJ	=him / her. DAT	give.3SG	book.the	(Mara)

'I struggle so that Mara can arrive/ so that Mara gives him/her the book'

In old Romanian, word order is even more flexible. As emphasized by (Pană Dindelegan 2016: 112), old Romanian displays features of a VS language through its preference for postverbal subjects, as opposed to present-day Romanian, where a more constrained word order can be observed (Vasilescu 2013: 537).

Nonetheless, it has been argued, especially in the generative framework (cf. Dobrovie-Sorin 1987; Dobrovie-Sorin 1993; Cornilescu 1997; Cornilescu 2000; Alboiu 2002),[28] that present-day Romanian is also a VS language, like old Romanian.

---

[28] In fact, current studies on Romanian (see Dobrovie-Sorin 1993, Alboiu 2002, Cornilescu 2000, Hill 2002) argue that Romanian is a VSO language. In their argumentation, the lexical verb undergoes obligatory displacement

The arguments supporting the hypothesis that the VS word order is the basic word order in present-day Romanian are based on the following:

(i) In *să* + subjunctive clauses, nothing but clitics can occur between the subjunctive marker *să* and the verb, as in (46c). Therefore, in the 20th century the complex complementizer *ca … să* was created by prescriptivists,[29] as a split variant of *să*, specialized for focalization in subordinate clauses, including subject focalization (cf. Pană Dindelegan 2013a: 124).

(ii) Romanian has an infinitive marker (*a* = 'to') and a subjunctive marker (*să*), with an ambiguous status (inflectional marker and complementizer), which occur only in VS languages (cf. Dobrovie-Sorin 1993: 87, 106);

(iii) The adjective always occurs in postnominal position in Romanian, which is typical of VS languages.

The last argument is an indirect one and has been deducted from two of Greenberg's quasi-universal implications, i.e. "(i) languages with a dominant VSO word order prefer placing the adjective postnominally (N-Adj); (ii) languages with a dominant VSO word order place the auxiliary in front of the main verb (Aux-V)" (Greenberg 1963: 85, cited in Pană Dindelegan 2013a: 125).

Still, Pană Dindelegan (2013a: 125) does not fully endorse the above-presented argumentation because of the similar behavior of overt subjects in Romanian and in other pro-drop Romance languages, which I also consider a valid counter-argument. Moreover, not all the arguments provided by the generative studies mentioned above are reliable. For instance, infinitive markers are present in all the Germanic languages, yet these are not known as VS languages. The uncertainty regarding the basic word order in Romanian stems from the fact that each argumentation is highly dependent on the theoretical framework it emerges from. In order to shed more light on this matter, an independent, usage-based study of the word order in Romanian would be necessary. However, such an in-depth study considering all types of structures in Romanian in all historical periods goes beyond the scope of this research.

Nevertheless, recent usage-based studies partially cover the problem of word order in Romanian. Pană Dindelegan (2016: 112–123) presents a study of the word order of subjects in old Romanian (16th–18th centuries), confirming the well-known freedom in word order in this period. In Table 3.3, I compiled the outcomes presented in Pană Dindelegan (2016: 112–114, Tables 2.4 to 2.6). The obtained table displays the preferred word order of subjects in different syntactic configurations, which show word order

---

into the T domain, and Case is valued via Agree without dislocation to SpecTP for classical EPP purposes (cf. Alboiu 2007: 13).

[29] As pointed out by Gheorghe (2013: 470), the complex complementizer *ca … să* was rarely used until the beginning of the 20th century. Its use in contemporary Romanian is the result of prescriptive rules (see also Stan 2007).

constraints: (i) impersonal expressions with clausal subjects (47), (ii) structures with inverted auxiliaries or clitics (48), and (iii) in main and subordinate clauses in general.

(47) *tâmplă* *-se* [*de* *se* *duseră* *în* *oaste*]ₛ (Moxa 1620: 61ʳ)
  happened.3SG =CL.REFL.ACC that CL.REFL.ACC went.3PL in army
  'and it happened that they joined the army'   (Pană Dindelegan 2016: 112, ex. 216a)

(48) *întru* *focu* *tremite* *-i*   *-va* *Dumnezeu*
  in fire send.INF =them.ACC.3PL -FUT.3SG God.
  'and God will send fire upon them'   (Pană Dindelegan 2016: 113, ex. 218b)

Table 3.3 Old Romanian word order in constrained structures (cf. Tables 2.4–2.6, Pană Dindelegan 2016: 112–114)

Structures with constraints	Preverbal subject (SV)	Postverbal subject (VS)
Impersonal construction with clausal subj	6,5 %	93,5 %
Inverted auxiliary	42,1 %	57,9 %
Clauses	Preverbal subject (SV)	Postverbal subject (VS)
Subj position in main clause	81 %	49,8 %
Subj position in subordinate clause	19 %	50,2 %

Based on these data, Pană Dindelegan (2016: 112–116) concludes that, in complex contexts, preverbal position of the subject is mostly found in the main clause, whereas the postverbal position is preferred in subordinate clauses and in impersonal constructions with clausal subjects. In structures with an inverted auxiliary, the figures do not show a significant difference between preverbal and postverbal position of the subject. Pană Dindelegan underlines that translated texts show a higher percentage of subject postposition in this period, which, in most cases, can be explained by imitation of the source text.

In a study based on texts from the 20ᵗʰ century, Giurgea (2017: 280) claims that, statistically, Romanian is not a VS language, but is of a mixed type VS/SV. His analysis is based on the assumption that Romanian is, at least to a certain extent, a *topic-oriented* language,[30] which means that word order marks the *topic-comment* partition (49a). Nevertheless, Giurgea also mentions that there are SV contexts where the subject is not the topic (49b). In other words, in a *topic-oriented* language, *the topic* should be positioned either before or after *the comment*, i.e. the rest of the clause, in all clauses containing a topic.

---

[30] In Dobrovie-Sorin (1987), preverbal subjects are analyzed as *themes*, a notion that, in current studies on information structure, corresponds to *topic*, as highlighted by Giurgea (2017).

(49) a. | *Maria* | *îşi* | *ţine* | *maşina* | *la* | *taică* | *-su* |
|---|---|---|---|---|---|---|
| Maria | CL.REFL.DAT | keeps | car.the | at | father | =hers |

'Maria keeps her car at her father's.'   (Giurgea 2017: 282, ex. 6a)

b. | *Cineva îi* | *aduce* | *flori* |
|---|---|---|
| somebody her.DAT | brings | flowers |

'Somebody brings her flowers'   (Giurgea 2017: 288, ex. 23a)

In his study, Giurgea shifts the focus from the conditions in which a subject occurs in preverbal position, to the restrictions that apply to postverbal subjects. He shows that the choice for a postverbal subject mostly has a pragmatic motivation, that is, it signals that the subject is not a topic (Giurgea 2017: 283).

For the purpose of my own research, I have carried out a case study on the word order in canonical structures of the Nom-Acc type occurring with an experiencer in the nominative and a state noun in the unmarked accusative (50a-b). The aim of this case study is to determine the predominant word order in structures with a nominative subject. The study is based on examples from my dataset, as described in Chapter 6, *infra*, containing data from the 16th century until today. In my dataset, nominative experiencer subjects occur either preverbally (50a) or postverbally (50b).

(50) a. | *Pe* | *timpul* | *lui,* | *românii* | *au* | *făcut* | *foame* | *şi* | *frig* |
|---|---|---|---|---|---|---|---|---|
| in | time.the | his.POSS.GEN | Romanians.the | have.3PL | made | hunger | and | cold |

'In his time, Romanians forbore hunger and cold'   (elenaudrea.ro)

b. | *Fiindcă* | *aşa* | *are* | *chef* | *guvernul* |
|---|---|---|---|---|
| because | so | has | will | government.the |

'Because this is the will of the government'   (hotnews.ro)

Recall, however, that the subject is not always realized in Romanian as a result of its pro-drop character. Therefore, I investigate the word order preference only for realized subjects. In my data, 72,5 % of the examples for all the historical periods contain a *pro* nominative subject (2 368 out of 3 266 examples). This implies that my conclusions must rely on the remaining 27,5 % of the examples (898). Note that the state noun is always present in the selected examples. Table 3.4 shows that the same ratio between a realized subject and *pro* is observed in each historical period. These facts are visualized in Figure 3.1.

Table 3.4   Realization of the nominative experiencer argument (16th–21st) – overt vs. *pro*

Nom-Acc	16th–18th	19th	20th	21st
NP experiencer	80 (27,3 %)	69 (33,2 %)	87 (28,3 %)	662 (26,9 %)
Pro	213 (72,7 %)	139 (66,8 %)	220 (71,7 %)	1 796 (73,1 %)
Total	293 (100 %)	208 (100 %)	307 (100 %)	2 458 (100 %)
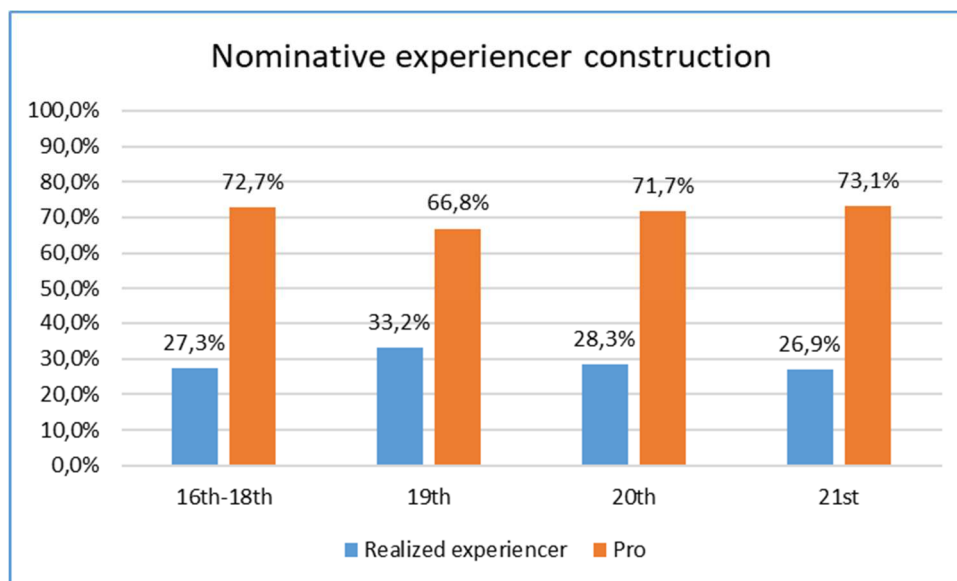
Figure 3.1   Realization of the nominative experiencer (16th–21st) – overt vs. *pro*

Regarding the remaining 898 examples in which the subject is realized (27,5 % of the total number of examples with nominative experiencers), the data show that a great part, namely 846 examples (24,86 % of all instances), contain a preposed overt subject (SV), and only 52 examples (2,63 % of all instances) postpose their overt subject (VS), as is illustrated in Table 3.5. As for the state noun, which in this construction is realized as the accusative object, it mostly occurs in postverbal position (VO), as shown in Table 3.5. In view of these data, the prevailing word order in the nominative experiencer construction is SVO.

Table 3.5   Preferred order of the arguments in the nominative experiencer construction

Nominative subject		Accusative object	
SV	24,86 %	VO	95,07 %
VS	2,63 %	OV	4,93 %
pro	72,50 %	Covert	0 %
Total	100 %	Total	100 %

From a diachronic perspective, the tendency toward preverbal position of the subject increases with every period, from 82,5 % in the 16th–18th centuries, to 95,8 % in the 21st century. At the same time, the preference for the postverbal position, by the same token, decreases significantly, from 17,5 % in the 16th–18th centuries, to only 4,2 % in the 21st century, as shown in Table 3.6, and displayed in the graph in Figure 3.2.

Table 3.6   The nominative experiencer (16th–21st) – preverbal vs. postverbal

	16th–18th	19th	20th	21st
Preverbal	66 (82,5 %)	64 (92,8 %)	82 (94,3 %)	634 (95,8 %)
Postverbal	14 (17,5 %)	5 (7,2 %)	5 (5,7 %)	28 (4,2 %)
Total	80 (100 %)	69 (100 %)	87 (100 %)	662 (100 %)
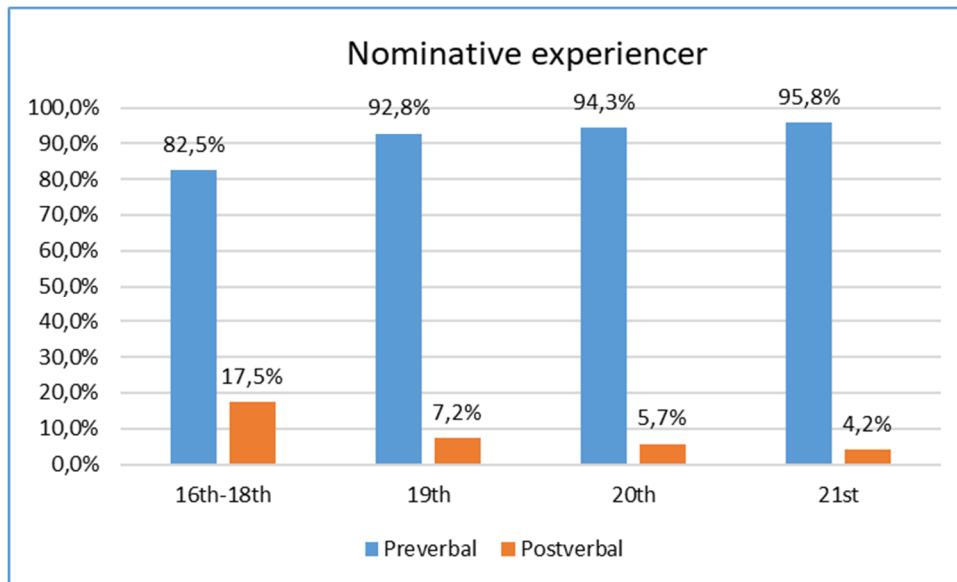
Figure 3.2   The nominative experiencer (16th–21st) – preverbal vs. postverbal

As mentioned at the beginning of the section, this case study focuses only on canonical constructions with a subject experiencer in the nominative. Nevertheless, its outcomes may challenge the generativist argumentation with regard to the VS dominant word order of subjects in Romanian. To confirm the validity of these results, a much larger, usage-based study is necessary.

For what concerns word order as a test for the subjecthood of dative and accusative experiencers, it raises an additional problem. In Romanian, both dative and accusative arguments are encoded either as NPs or as strong pronouns, always doubled by a clitic, or they are encoded only as clitics (51)-(52). This is because Romanian is a clitic doubling language, which implies that accusative and dative NPs or strong pronouns must always be doubled by a clitic.

(51) *Pe   Roxana /   O   îngrijorează   tristeţea   lui*
    ACC   Roxana   her.ACC   worries   sadness.the   his.POSS.GEN
    'Roxana / She is worried about his sadness'

(52) *Mie /   Îmi   plac   cărţile*
    me.DAT / me.DAT   please.3PL   books.the
    'I like books'

However, clitics and weak pronouns are usually attached to the left of the verb, in a fixed order. One exception to this is with imperatives and with gerunds, where the clitic is postposed to the verb. Hence, the word order of clitics is not free and does not provide a relevant subjecthood test. Free word order may only be observed with experiencer NPs, which are, however, less frequent because the experiencer is generally topical and mostly expressed through a pronoun, and hence dropped. Therefore, in order to study the position of the doubled experiencer NP, a large corpus is necessary.

Van Peteghem (2016: 8–9), in her research on pain verbs in present-day Romanian, examines the position of the nominative NP denoting the body part. Her study shows that, in the 21st century, the nominative NP tends to occur post-verbally in 85,99 % of the examples, although it is always definite (53a-b). Given the nominative NP's preference for the postverbal position, Van Peteghem argues that this NP does not behave like a subject. Drawing on other arguments, such as binding, she claims that accusative experiencers show more syntactic subject properties than the nominative NPs in these constructions.

(53) a. *Mă*   *doare*   *capul*
     me.ACC   aches   head.the
     'I have a headache'
   b. *Capul*   *mă*   *doare*
     head.the   me.ACC   aches
     'I have a headache'

The results obtained by Van Peteghem (2016) reveal a striking similarity with the results of my case study on canonical structures. Table 3.7 compares the position of the nominative subject in the canonical structure, for the 21st century, with the position of the nominative body part in the pain verb construction, for the same period.

Table 3.7   Position of the subject in the Nom-Acc structure in relation to the body part

	S-V	V-S	Total
Nom-Acc (nominative subj)	95,8 %	4,20 %	100 %
Acc-Nom (body part)	14,01 %	85,99 %	100 %

These data corroborate the analysis proposed by Van Peteghem (2016). Indeed, the nominative NP denoting the body part does not pattern with the canonical subject with respect to its dominant word order. Moreover, both Van Peteghem's study and the present study reveal a striking similarity between the positions of nominative NPs in the pain construction and the accusative object in the canonical structure, which both occur predominantly in postverbal position, as evident from Table 3.8. Indeed, based on these figures, the nominative NP denoting the body part in pain construction patterns, in 21st-century Romanian, with the object in the canonical structure, rather than with

the subject. This evidence provides further support to Van Peteghem's claim that the accusative experiencer is more subject-like than the nominative NP in these constructions.

Table 3.8   Position of the object in the Nom-Acc structure in relation to the body part

	O-V	V-O	Total
Nom-Acc (accusative object)	3,6 %	96,4 %	100 %
Acc-Nom (body part)	14,01 %	85,99 %	100 %

To sum up, in this section, I have shown, based on a dataset containing 3 266 examples, that in the nominative experiencer construction the basic word order situates the canonical nominative subject in preverbal position, whereas the object occupies the postverbal position in the majority of the examples. Furthermore, comparing my study with the one carried out by Van Peteghem (2016), I validate that the nominative body part-NP of pain constructions patterns with the accusative object of nominative experiencer constructions, contrary to traditional analyses. This evidence implies that the subject-like argument of the pain construction patterns with the canonical subject of nominative experiencer constructions in Romanian, in that they both occur predominantly in preverbal position.

Word order provides insightful information on the syntactic behavior of nominative subjects in Romanian, which, in canonical structures, clearly behave differently from objects. Therefore, I consider word order a reliable subjecthood diagnostic in Romanian, and the preverbal position of the accusative and dative experiencers under scrutiny an indication of their subjecthood.

3.2.3   Binding

One of the most used subject tests is binding, more particularly binding of reflexives, which are necessarily bound by a syntactic subject in most languages. In this section, I examine to what extent binding of reflexives provides a reliable subject test in Romanian.

Romanian has specialized reflexive markers only for the 3rd person: the weak/clitic pronouns ACC *se /s-,* DAT *îşi/ şi-* and the strong pronouns ACC *sine* / DAT *sieşi*.[31] The strong forms may be reinforced by the intensifier *însuşi* 'himself' (54).[32] This marker may also

---

[31] Henceforth I will not make a distinction between *clitics and weak forms* (cf. Footnote 20, in Section 3.1.1).

[32] This pronoun is composed of two morphemes, each of which has a different type of declension: *îns-* in gender and number, and *-şi* in person and number. Its etymology is controversial. Three hypotheses have been

transform a strong non-reflexive pronoun into a reflexive pronoun, as evident from the same example. Just like *himself* in English, this combination [personal pronoun + *însuşi*] tends to grammaticalize in Romanian as a reflexive pronoun (cf. Vasilescu 2017: 11; Dobrovie-Sorin & Giurgea 2013: 288–294).[33]

(54) Ştefan$_i$ **se$_i$** / ***îl$_i$** iubeşte **pe** **sine$_i$** (*însuşi*) / **pe** **el$_i$** *însuşi*[34]
    Ştefan CL.REFL.ACC / him.ACC loves ACC himself himself / ACC him himself
    'Ştefan loves himself.'

The concept of binding stems from the well-known generative 'Binding theory'. The principles of the Binding theory have been claimed to be able to account for binding relations crosslinguistically. This theory distinguishes three types of expressions: (i) A-expressions (A = anaphor), which must have an antecedent in a local governing domain (55a); (ii) P-expressions (P = pronouns), which cannot have an antecedent in a local domain but can have a c-commanding antecedent outside their local domain (55b); and (iii) R-expressions (R = referential), i.e. non-pronominal elements, which cannot have a c-commanding antecedent, and are, hence, totally free (55c).

(55) a. *Victor$_i$* **se$_{i/*j}$** *spală*
      Victor CL.REFL washes
      'Victor washes himself'
   b. *Victor$_i$* **îl$_{*i/j}$** *spală*
      Victor him.ACC washes
      'Victor washes him'
   c. *Victor$_i$* **îl$_{*i/j}$** *spală.* ***Acest_tip$_{i/j/k}$*** *ştie* *ce* *face*
      Victor him.ACC washes this_type knows what does
      'Victor washes him. This guy knows what he has to do'

The examples in (55a-c) show contexts in which the Binding theory principles apply in Romanian. Nevertheless, the distinction (i) vs. (ii) is not found in all forms in Romanian:

---

proposed regarding the analysis of its components: (i) *îns* + the dative reflexive *şi* (< Lat. *sibi);* (ii) *îns* + the deictic adverbial -*şi* (< Lat. *sic);* (iii) *îns* + -*şi,* the clitic form of Lat. *ipse* (for details see Manoliu-Manea 1987; Zafiu 2012, cited by Vasilescu 2017). As for *îns/ ins,* it is a direct descendant of the Latin *ipse,* and it means 'guy, person, individual'.

[33] Note that the reflexive [personal pronoun + *însuşi*] represents only one use of *însuşi*. *Însuşi* has indeed multiple uses and combinatorial possibilities (see Vasilescu 2017). Depending on the context, the combination [personal pronoun + *însuşi*] can also be used as an intensified non-reflexive pronoun as in (i), *însuşi* being in the first place an intensifier (cf. Vasilescu 2017).

(i) *El* *însuşi* *a* *venit*
    he himself has arrived
    'He himself came'

[34] Note that *pe* is analyzed in Romanian in terms of DOM (Differential Object Marking) (Mardale 2008). Nevertheless, since the exact value of *pe* is not relevant for the present research, I choose to gloss it ACC in order not to complicate the glosses unnecessarily.

in spite of the existence of a specialized strong reflexive *sine*, strong personal non-reflexive pronouns of the 3rd person can occur instead of strong reflexives (cf. example 89a-b from Dobrovie-Sorin & Giurgea 2013: 269, given here under (56a-b).

(56) a. Ion_i   se_i /   *îl_i   iubeşte   pe   sine_i /   pe   el_i   însuşi
      Ion   CL.REFL.ACC / him.ACC   loves   ACC   himself /   ACC   him   himself
      'Ion loves himself.'

    b. Ion_i   se_i   iubeste   numai   pe   el_i.
      Ion   CL.REFL.ACC   loves   only   ACC   him
      'Ion loves only himself.'

Note, however, that, as opposed to the strong reflexives, the clitic reflexive forms are always obligatory in Romanian when the co-referring forms are in the same binding domain, i.e. in the same clause. This is evident from (56a), where the non-reflexive accusative clitic *îl* 'him' may not occur instead of the reflexive accusative clitic *se* 'himself'.

Two types of reflexive binding can be distinguished: *clause-bound* and *long-distance reflexivization.* In most languages, reflexive pronouns seek their antecedent close by, in the same binding domain, which is the clause. This is known as *clause-bound reflexivization*, as in the examples in (56) above. However, in certain languages, the reflexive pronoun and its antecedent do not have to be located within the same clause; the antecedent can occur in the main clause and the reflexive pronoun in the subordinate clause. This relation is called *long-distance reflexivization.* Long-distance reflexivization exists in Icelandic, Faroese, Norwegian dialects (Andrews 1976; Sigurðsson 1989; Barðdal 2001), in Dutch (Koster & Reuland 1991; Zribi-Hertz 1996: 199–200), and in Latin and Ancient Greek (Barðdal et al. forthcoming), but not in German (Barðdal 2006).

In Romanian, long-distance reflexivization is possible with the reflexive *sine* 'himself', as well as with the reinforced personal pronoun *el însuşi* 'himself', which, as mentioned above, grammaticalizes as a reflexive pronoun. Indeed, in (57a-b), the reflexive ACC *sine*, and the reinforced personal pronoun *el însuşi* 'him himself' may refer either to the subject of the main clause (*tatăl*), or to the subject of the subordinate clause (*Marius*).

(57) a. **Tatăl_i**   vrea   ca   Marius_j   să   vină   **la**   **sine_i/j** /
      father.the   wants   COMP   Marius   SUBJ   come.3SG   at   self /
    el   **însuşi_i/j** /   **el_i/j**
    him   himself /   him
    'The father wants that Marius trusts himself / him'

    b. **Tatăl_i**   vrea   ca   Marius_j   să   reuşească   **datorită**   **sieşi_i/j** /
      father.the   wants   COMP   Marius   SUBJ   succeed.3SG   thanks_to   self.DAT /
    lui   **însuşi_i/j** /   **lui_i/j**
    him.DAT   himself /   him.DAT
    'The father wants that Marius succeeds thanks only to himself / him'

These Romanian examples are in line with the remark made by Zribi-Hertz (1996), according to which, in several languages, the reflexives occurring in long distance reflexivization alternate with non-reflexive personal pronouns, violating the principle B (cf. (ii) above) of the standard Binding theory.

As for the clitic reflexives (cf. ACC *se / s-,* DAT *îşi / şi-*), they do not occur in long distance reflexivization contexts, and can be bound only in their local domain. This may be a consequence of their maximal underspecification, which is in conformity with Burzio's (1989, 2010) *Morphological / Referential Economy principle.* This principle stipulates that, in a language with many types of pronouns, the ones which are obligatorily bound in their domain are the most weak morphologically, and the most underspecified.

The relevance of reflexive binding as a subject test is crucially dependent on the assumption that reflexives may only be bound by subjects. However, in several languages, instances are found in which reflexives may be bound by objects (for Icelandic see Hyams & Sigurjónsdóttir 1990, for Latin and Greek see Barðdal et al. forthcoming). This also holds for Romanian, where both the reflexives ACC *sine, se /* DAT *sie(şi), î(şi)* and the reinforced personal pronoun *el însuşi* 'he himself' may be bound by accusative objects and by dative recipients (58a-b).

(58) a. *Tatăl$_i$*   *îl$_j$*   *ajută*   **pe**   **Mihai$_j$**   *cu*   *un*   *text*   *despre*
  father.the   him.ACC   helps   ACC   Mihai   with   a   text   about

  sine$_{i/j}$/   *el*   *însuşi$_{i/j}$/*   *el$_{i/j}$*
  himself /   him   himself /   him

  'The father helps Mihai with a text about himself / him'

  b. *Tatăl$_i$*   *i$_j$-*   *a*   *dat*   **lui**   **Mihai$_j$**   *o*   *carte*   *utilă*
  father.the   him.DAT   has   given   DAT   Mihai   a   book   useful

  sieşi$_{i/j}$/   *lui*   *însuşi$_{i/j}$/*   *lui$_{i/j}$*
  himself.DAT /   him.DAT   himself /   him.DAT

  'The father gave Mihai a book useful to himself / him'

Concerning languages such as Icelandic, and also Latin and Greek, Barðdal et al. forthcoming) argue that, in spite of the existence of object binding, the test is relevant because, in these languages, there is a choice between reflexives and pronouns in object binding, while subjects have no such choice. However, in Romanian binding rules are less strict in that the choice between reflexives and pronouns is available not only for objects (58a-b), but also for subjects (56b) above. Indeed, in Romanian, the strong reflexive forms and the non-reflexive forms alternate with each other. The only context in which the distributional complementarity of reflexives and pronouns is preserved, is when the reflexive is a clitic (56a) above.

Table 3.9 summarizes the situation for Romanian with respect to binding of strong and clitic forms of reflexives and of personal pronouns, by canonical subjects and by objects, in both clause-bound and long distance reflexivization.

Table 3.9   Binding by canonical subjects and by objects

		Clitics		Strong forms		
		Reflexives *se/ [î]şi*	Pers. pron. *îl/îi*	Reflexives *sine/ sieşi*	Reinforced pers. pron. *el însuşi*	Anaphors *el/ ea*
Subjects	Clause-bound	✓	✗	✓	✓	✓
Objects		✓	✗	✓	✓	✓
Subjects	Long distance	✗	✓	✓	✓	✓
Objects		✗	✓	✓	✓	✓

As expected, binding of reflexives is also possible with the accusative and dative experiencers under study, either in clause-bound (59a-b) or in long-distance reflexivization (60a-b). As in the case of canonical structures, both the strong reflexives and the personal pronouns may be bound within the same clause, as well as in long-distance reflexivization, whereas the clitic forms of the reflexives can be bound only within their own clause.

(59) a. *Îl$_i$*   *întristează*   *zvonurile*   *despre*   *sine$_{i/*j}$ /*   *el*   *însuşi$_{i/*j}$ /*   *el$_{i/j}$*
      him.ACC   saddens   rumors.the   about   himself /   him   himself /   him
      'The rumors about himself saddened him'

   b. *Lui   Victor$_i$*   *îi$_i$*   *place*   *poza*   *cu*   *sine$_{i/*j}$ /*
      DAT   Victor   him.DAT   pleases   picture.the   with   himself /
      el   *însuşi$_{i/*j}$ /*   *el$_{i/j}$*
      him   himself /   him
      'Victor likes the picture with himself'

(60) a. *Pe   Roxana$_i$*   *o$_i$*   *preocupă*   *că*   *mama$_j$*   *o*   *iubeşte*
      ACC   Roxana   her.ACC   worries   that   mother.the   her.ACC   loves
      mai   mult   *pe   Clara*   *decât*   *pe*   *sine$_{i/j}$ /*   *ea*   *însăşi$_{i/j}$ /*   *ea$_{i/j}$*
      more   much   ACC   Clara   than   ACC   herself /   her   herself /   her
      'Roxana is worried that mama loves more Clara than herself'

   b. *Anei$_i$*   *îi$_i$*   *displace*   *că*   *mama$_j$*   *o*   *iubeşte*   *mai*
      Ana.DAT   her.DAT   displeases   that   mother.the   her.ACC   loves   more
      mult   pe   Mara   *decât*   *pe*   *sine$_{i/j}$ /*   *ea*   *însăşi$_{i/j}$ /*   *ea$_{i/j}$*
      much   ACC   Mara   than   ACC   herself /   her   herself /   her
      'Ana dislikes that mother loves more Mara than herself'

These examples show that accusative and dative experiencers can be binders of reflexives when they are situated within the same clause or in a subordinate clause. However, this does not really validate their subject status, since in Romanian binding is unable to distinguish subjects from objects.

    Moreover, Haspelmath (2001a: 71) shows that in certain languages, such as Italian, Polish, or German, the nominative stimulus may also bind the reflexive dative

experiencer when it occurs in preverbal position (cf. Haspelmath's example 38a for Italian, given here under 61). However, Barðdal, Eythórsson & Dewey (2019) claim that these structures refer to the alternating pair of the verb *please*, which has a different order of the arguments, namely Nom-Dat.[35] The specific of these alternating predicates is that the participants keep their case marking while changing only their grammatical relation. With other words, a dative experiencer will continue to be the dative-marked experiencer whether it encodes the subject or the indirect object of the alternating predicate. Such alternating verbs are argued to exist in Icelandic and in German. Haspelmath's analysis of the example in (61) points toward the existence of such alternating predicates in the above-mentioned languages too. However, this claim needs to be verified by means of a thorough study for each language.

(61) Italian

Anna	*piace*	*a*	***sé***	***stessa***
Anna	pleases	to	REFL	self

'Anna likes herself'

In Romanian, an example such as (62a), which translates literally the Italian example in (61), is less natural to speakers of Romanian. However, at a closer examination, I found examples such as (62b and c), which show that the stimulus may also bind the dative experiencer when it occurs in preverbal position, since it can bind the possessive *propriu* 'own' (62b) and can participate in reciprocal binding (62c) (cf. (Chomsky 1980: 9–12).

(62) a. **Ana$_i$* *(îşi)$_i$* *place* *sieşi$_i$/* *ei* *înseşi$_i$*
   Ana herself.DAT pleases herself.DAT / her.DAT herself.DAT

   b. *Ana$_i$* *(le)* *place* *propriilor* *săi$_i$* */* *ei$_i$* *părinţi*
   Ana them.DAT pleases own.the.DAT hers.POSS/ hers.POSS parents
   'Her own parents like Ana' / 'Ana is to the liking of her parents'

   c. *Ana* *şi* *Mihai$_i$* *(îşi)* *plac* *unul altuia$_i$*
   Ana and Mihai themselves.DAT please.3PL one another.DAT
   'Ana and Mihai like each other'

These examples show that Romanian too has alternating verbs of the type Nom-Dat and Dat-Nom, as it has been argued for Icelandic and for German (Barðdal, Eythórsson & Dewey 2019). Hence, if *plăcea* 'please' is an alternating verb in Romanian, the ability of the nominative to bind a possessive or a reciprocal pronoun in the examples in (62b and c) does not exclude a subject analysis of the dative, since binding involves this time Nom-Dat structures and not Dat-Nom structures.

---

[35] This type of verb with two opposite configurations is also called *alternating predicate,* a concept first introduced by Bernódusson (1982) for Icelandic, and further described by several scholars (Barðdal 2001; Eythórsson & Barðdal 2005; Sigurðsson 2006; Thráinsson 2007; Rott 2013, and Barðdal, Eythórsson & Dewey 2018).

In spite of these difficulties, binding is used as a subject test for non-canonical subjects in Romanian by Geber (2011). However, she uses yet another kind of binding, i.e. quantifier binding (cf. Culicover 1992), which is also used by (Fernández-Soriano 1999) for Spanish dative experiencers with subject behavior. This phenomenon is illustrated in Geber's (2011) example 41, given in (63a), in which the quantifier *fiecare* 'every' binds the possessive *sa* 'his/her', giving rise to a distributive reading of the possessive. Note that binding of these anaphors is possible when the binder occurs within a dative NP denoting an experiencer, but not when it occurs within a nominative NP denoting a stimulus (63b).

(63) a. **Fiecărui**   **copil$_i$**   **îi**   *place*   *jucăria*   **sa$_i$.**
    each.DAT   child.DAT   him.DAT   pleases   toy.the   his.POSS
    'Each child likes his toy.'
  b. ***Fiecare**   **jucărie$_i$**   **îi**   place   copilului   **său$_i$.**
    each   toy   him.DAT   pleases   child.the.DAT   hers.POSS

However, in this kind of binding the binding quantifier may also occur in object position (64a), or with a dative recipient (64b). Based on this evidence, quantifier binding too is problematic in distinguishing between subjects and objects.

(64) a. *A*   *convins*   *pe*   **fiecare$_i$**   **student**   *de*   *greşeala*   **sa$_i$**
    has   convinced   ACC   each.ACC   student   of   mistake.the   his.POSS
    'S/He convinced each student of his mistake'
  b. *A*   *dat*   **fiecărui$_i$**   **student**   **propriul**   **său$_i$**   *exemplar*
    has   given   each.DAT   student   own.the   his.POSS   copy
    'S/He gave each student his own copy'

Therefore, we may conclude that in Romanian accusative and dative experiencers may bind anaphors. However, the binding test is not fully reliable, since object accusatives and recipient datives may also bind, to the same extent, reflexives and personal pronouns.

3.2.4 Control of implicit subjects of non-finite verbs

As mentioned in Chapter 1, in many languages, another crucial subject test comes from phenomena related to control of unexpressed subjects of non-finite verb forms.[36] In this section, I will examine whether and how such control phenomena can be used as subject diagnostics in Romanian.

---

[36] Cf. Falk (1995: 203), cited by Eythórsson & Barðdal (2005: 828).

The term *control* refers to the relation between the covert subject of a non-finite subordinate clause and its antecedent in the main clause. This antecedent can be the subject, a direct object, or an indirect object. With other words, control occurs when a verb shares a semantic argument with its verbal complement. This argument may then receive two different semantic roles assigned by the two different verbs (cf. Barbu 2015: 319; Borsley 1996). In the example in (65), I illustrate the phenomenon of control by providing Barbu's (2015: 319) example (7), in which I preserve her own notation (sc = small clause; *e* = PRO).

(65) John^i$_{Subject.experiencer}$ hopes [$_{sc}$ *e*^{i}$_{Subject.agent}$ to leave]$_{Complement}$

The covert subject, often referred to as PRO, must be distinguished from the null category *pro*, also called *zero* or *null anaphora*, which is a null position with referential properties co-occurring with finite verb forms in pro-drop languages. Example (66a) illustrates control in Romanian, contrasting it with the null category *pro* in (66b).

(66) a. *Clara$_i$*   *a*   *încercat*   ___$_i$ *a*   *învăţa japoneza*   *când*   *era*
        Clara   has   tried   PRO   INF   learn Japanese.the   when   was
        (deja)   *prea*   *târziu*   *pentru*   *ea*
        already   too   late   for   her
        'Clara tried to learn Japanese when it was already too late for her'
     b. *Dimineaţa*   ___   *bea*   *două*   *cafele*
        morning.the   *pro*   drinks   two   coffees
        'In the morning, s/he drinks two cups of coffee'

In most languages, PRO can occur with infinitives (cf. Icelandic, see Sigurðsson 1992, 2004; Barðdal 2002, 2006a; Barðdal & Eythórsson 2003, 2012; Eythórsson & Barðdal 2005) or gerunds (cf. Russian, see Moore & Perlmutter 2000). In Romanian, both non-finite verb forms can occur in control structures, plus an additional one, called *supine* (67), which is a non-finite verb form homonymous with the past participle, but used differently.[37]

(67) *Marius$_i$*   *se*   *apucă*   ___$_i$ *de*   *cântat*
     Marius   SE   start   PRO   SUP   sung
     'Marius starts to sing'

---

[37] The supine is a non-finite verb form that expresses purpose, but also obligation and necessity. Specific to Romanian in the Romance context, it has a controversial history. Some scholars consider it directly inherited from Latin (Diaconescu 1971: 151; Lombard 1974: 302), whereas others believe it developed in Romanian, out of the past participle, as a consequence of the loss of the verbal value of the infinitive (Caragiu-Marioţeanu 1962; Vasiliu & Ionescu-Ruxăndoiu 1986: 196–198). In other studies (see Brâncuş 2007: 168), the supine is considered a Balkan Sprachbund feature.

Several researchers have argued that in Balkan languages, among which Romanian, the subjunctive may also be accounted for in terms of control, a phenomenon known as *finite control*.[38] In line with this claim, the subjunctive will be accounted for in terms of control too in the following sections. For a better understanding, a few explanations are needed on the history of the Romanian infinitive and its relationship with the subjunctive.

3.2.4.1   Infinitive vs. subjunctive in Romanian

In Romanian, the infinitive has lost its Latin ending *-re*, and corresponds to the lexical root of the verb without any inflection (e.g. Lat. *cantare > cânta* 'sing'). It is mostly headed by the morpheme *a* 'to', which is both an inflectional marker and a complementizer (cf. Dobrovie-Sorin 1993: 82-87).[39] However, like other Balkan languages, Romanian has gradually replaced the infinitive with the *Balkan subjunctive* (cf. Hill 2014; Hill & Mišeska-Tomić 2009). This explains why the infinitive is very rare in my dataset, at least in complement clauses.

As for the Balkan subjunctive, it is a finite verb form, inflected for person and number, which is always preceded by the morpheme *să*. This morpheme derives from the Latin complementizer *si* 'if, whether', and has evolved in Romanian into a clitic subjunctive marker. In Romanian, it functions both as an inflectional marker and as a complementizer, just like the infinitive marker *a*. *Să* differs from other complementizers, such as *că* 'that' (68a), in that, due to its nature as a clitic, nothing but clitics can intervene between *să* and the verb (68b). Hence, when realized as an NP, the subject must occur postverbally (68c).[40] However, when the subject NP is topical or focal, it may be fronted, but in that case the complex complementizer *ca … să* is used and the topicalized or focalized NP occurs between *ca* and *să* (68d) (cf. Gheorghe 2013: 470).

(68) a. *Mara*   *spune*   *că*   *Marius*   *a*   *venit*
        Mara   says   that Marius   has   come
        'Mara says that Marius has come'
    vs.

---

[38] The phenomenon of *finite control* is a well-known characteristic of Balkan languages. It has been documented and studied extensively in quite a few languages. For Greek, see Iatridou (1988), Terzi (1992, 1997), Varlokosta & Hornstein (1993), Philippaki-Warburton & Catsimali (1999), Roussou (2001), Krapova (2001); for Albanian, see Terzi (1992), Turano (1994), Dobrovie-Sorin (2001); for Romanian, see Comorovski (1985), Farkas (1985), (Dobrovie-Sorin 1993, 2001); for Bulgarian, see (Krapova 1999, 2001), and for Serbo-Croatian, see (Zec 1987).

[39] The only context in which the inflectional marker is not present is when the infinitive occurs with an auxiliary (cf. *voi cânta* 'I will sing').

[40] This is one of the arguments in favor of the underlying VSO word order in Romanian, put forward by Dobrovie-Sorin (2011: 10), Hill (2002) (cf. also Section 3.2.2).

b. *Mara*   *speră*   *să*   *-l*   *vadă*
    Mara   hopes   SUBJ   =him.ACC   see.3SG
    'Mara hopes to see him'

c. *Mara*   *speră*   *să*   *(*Marius)*   *vină*   *(Marius)*
    Mara   hopes   SUBJ   Marius   come.3SG   Marius
    'Mara hopes that Marius will come'

d. *Mara*   *speră*   *ca*   ***Marius***   *să*   *vină*
    Mara   hopes   COMP   Marius   SUBJ   come.3SG
    'Mara hopes that Marius will come'

In certain structures, the subjunctive behaves in Romanian like the infinitive in languages with control infinitives (e.g. English).[41] The subjunctive marker *să* seems to be similar to the English infinitive marker *to,* just like the Greek marker *na,* in that it is at the same time a complementizer and an inflection marker introducing a verb form which shows limited temporal properties (Beys 2009: 110; Alexiadou & Anagnostopoulou 2002; Iatridou 1993). Hence, in certain structures, i.e. in obligatory control structures (69a), the null subject of the subjunctive may be analyzed as PRO (Comorovski 1985; Farkas 1985; Dobrovie-Sorin 1993, 2001; Alboiu 2007), while in non-obligatory control structures, as in (69a), the null subject of the subjunctive behaves like *pro*, i.e. pro-drop (Farkas 1988; Motapanyane 1995; Dobrovie-Sorin 2011; Cotfas 2016). The *pro* analysis is supported by the fact that it alternates with NPs and nominative strong personal pronouns (69b).

(69) a. *Marius$_i$*   *vrea*   *să*   *plece*   *___$_{i/j}$*
      Marius   wants   SUBJ   leave.3SG=PL   PRO / *pro*
      'Marius wants to leave'

    b. *Marius*   *vrea*   *să*   *plece*   ***ea* /**   ***părinţii***
      Marius   wants   SUBJ   leave.3PL=SG   she /   parents.the
      'Marius wants her / the parents to leave'

However, in Romanian, the null subject of infinitives (70) and gerunds (71) can also alternate with nominative NPs and strong pronouns, more particularly when the subject of the non-finite forms is not coreferential with the subject of the main verb. The fact that nominal and pronominal nominative subjects are possible in these contexts has led

---

[41] Besides the Balkan languages, subjunctives with similar non-finite behavior are signaled also in certain Southern Italian dialects (Salentino and Southern Calabrese, cf. Groothuis 2019). Following Bianchi (2003) the subjunctive in these languages shows non-finite behavior, in that, when it occurs in control structures, it is able to express tense only with respect to the event of the main clause, and is not connected (not anchored) to the speech (external) event (Bianchi 2003: 13), whereas it still preserves person agreement. Therefore, the Romanian data represent important evidence supporting the hypothesis that finiteness may be a scalar rather than a binary notion (Groothuis 2019).

scholars to reject the PRO-analysis for Romanian infinitive, gerund or subjunctive clauses (Manzini & Roussou 2000; Hornstein 1999, 2001, 2003; Dobrovie-Sorin 2001, 2011)

(70) *El$_i$*   *plecase*   *înainte*   *de*   *a*   *ajunge*   **profesorul$_j$/**   **eu$_j$/**   **noi$_j$**
     He   leaved.3SG   before   of   INF   arrive   teacher.the   I   we
     'He left before the teacher / I / we arrived'

(71) *Ion$_i$*   *a*   *reuşit*   *trudind*   **alţii$_j$**   *pentru*   *el*
     Ion   has   succeeded.3SG   working   others.NOM   for   him
     'Ion succeeded by others toiling for him'   (adapted example 7c, Pană Dindelegan 2013a: 103)

Nevertheless, other linguists such as Landau (2001, 2004), or Jordan (2009) defend a PRO analysis of the unexpressed subject of the subjunctives in obligatory control structures. Their arguments are the following (cf. also Varlokosta & Hornstein 1993; Krapova 2001; Landau 2004, for the Balkan languages):

(i)   With verbs showing obligatory control in other languages, the subject must be unexpressed in Romanian also. Lexical subjects are then excluded (72a-b);

(72) a. *Clara$_i$*   *a*   *încercat*   *să*   *facă*   ___$_{i/*k}$ *un*   *chec*
     Clara   has   tried.3SG   SUBJ   make.3SG   PRO   a   cake
     'Clara tried to bake a cake'
    b.*Clara$_i$*   *a*   *încercat*   *să*   *facă*   **Mara$_j$ /**   ___$_j$ *un*   *chec //*
     Clara   has   tried.3SG   SUBJ   make.3SG   Mara /   pro   a   cake //
     să   *faci*   **tu$_k$ /**   ___$_k$ *un*   *chec*
     SUBJ   make.2sg   you /   pro   a   cake

(ii) With obligatory control verbs, PRO lacks independent reference: it must be coreferential with an antecedent in the same clause. This is shown in (73), where the reflexive *ea însăşi* 'herself', controlled by PRO, can only refer to Clara, which is the antecedent of the PRO.

(73) *Clara$_i$*   *a*   *încercat*   *să*   *facă*   ___$_i$ *un*   *chec*   *pentru*   **ea**   **însăşi$_{i/*j}$**
     Clara   has   tried   SUBJ   make.3SG   PRO   a   cake   for   her   herself
     'Clara tried to bake herself a cake'

(iii) PRO manifests sloppy reading under ellipsis (cf. Landau 2001), as illustrated in Jordan's example (93a), given here under (74);

(74) *Maria$_1$*   *încearcă*   *[să*   *plece*   ___$_1$ *devreme]*   *şi*   *Ana –*   *de_asemenea*
     Maria   tries   SUBJ   leave.3SG   PRO   early   and   Ana –   too
     'Maria is trying to leave early and Ana is too.'

(iv) PRO supports only a *de se* interpretation (cf. Varlokosta & Hornstein 1993: 508), which means that it patterns with emphatic or reflexive forms like *himself*, and not with personal pronouns, which, in contrast, would allow a *de re* (referring to somebody else)

interpretation (cf. Jordan's 2009 example 98b, given here under 75).[42] In Jordan's example given in (75), e_1 represents PRO, and cannot have another referent than *uitucul* 'the forgetful', the subject of the main clause.

(75) *Uitucul₁*	*și-*	*a*	*amintit*	*[e₁/*₂*	*să*	*ia*	*trenul]*
forgetful.the	CL.REFL.DAT=	has	remembered	PRO	SUBJ	take.3SG	train.the

'The forgetful man remembered to take the train.'

Though the PRO analysis of the subject of the subjunctive is controversial, I will include subjunctives in my discussion on the relevance of subject tests related to control for non-nominative subject-like arguments in Romanian. The main reason for this choice is that subjunctives occur as substitutes for the infinitive in the same contexts where other languages have PRO with an infinitive. In addition, the examples of control subjunctives as well as the cases of raising out of subjunctive clauses are as natural as the embedded finite subjunctive clauses in Romanian.

As for infinitives in Romanian, due to their competition with subjunctives, they are very rare in control structures, but they can still be found in certain configurations. In present-day Romanian infinitives never occur in non-obligatory control structures (76a) and are rare with obligatory control verbs (76b) (cf. Nedelcu 2016: 243). Nevertheless, controlled infinitives are still frequent in prepositional adjunct clauses, expressing purpose or time (76c).[43]

(76) a. *Marius_i*	*vrea*	*să*	*plece*	___$_{i/j}$ /	**a*	*pleca*	___$_{i/j}$
Marius	wants	SUBJ	leave.3SG	PRO /	INF	leave-	PRO

'Marius wants to leave'

b. ___$_i$	*A*	*încercat*	*a*	*-l*	*convinge*	___$_i$
pro	has	tried.3SG	INF	=him.ACC	convince	PRO

'S/he tried to convince him'

c. ___$_i$	*A*	*mâncat*	*înainte*	*de*	*a*	*pleca*	___$_i$
pro	has	eaten	before	of	INF	leave	PRO

'S/he ate before leaving'

It is worth mentioning that, in old Romanian, the infinitive occurs more frequently than in present-day Romanian in structures with obligatory control, whereas the subjunctive is possible but very rare with obligatory control. According to Frâncu (1969: 82, 93) and

---

[42] Cf. Lat. *de se*, 'of oneself', as opposed to *de re* 'of the thing' (and to *de dicto* 'of the word') (Jordan 2009: 138).

[43] As mentioned in Nedelcu (2016: 245), in the 16th century, the infinitive, preceded by the infinitive marker *a* 'to', is used with only a few prepositions (*de* 'of, for', *în loc (de)* 'instead of', *până* 'until'). In the following centuries, a wider range of prepositions starts selecting an infinitive. From the 19th century until today, the infinitive occurs frequently after prepositions, probably as a result of French influence (cf. Schulte 2007: 303–304, and Spătaru-Pralea 2013: 146–157, cited by Nedelcu 2016: 245).

Nedelcu (2013: 130–131), the higher frequency of controlled infinitives in old Romanian is due to their lack of ambiguity.

3.2.4.2   Control phenomena in Romanian as a subjecthood diagnostic

Let us turn now to control phenomena as a subject test. As shown in many studies, control phenomena provide a reliable subject test for Icelandic and German (cf. Zaenen, Maling & Thráinsson 1985; Sigurðsson 1989; Fanselow 2002; Haider 2005, 2010; Barðdal 2006), but also for French (cf. Legendre 1989) and for Russian (cf. Moore & Perlmutter 2000: 407), among others. However, control phenomena are used as subject diagnostics in two different ways, giving rise to the following distinct control tests:

(i) The first test focuses on the controlled PRO and hinges on the hypothesis that only subjects of non-finite forms can be controlled by an argument of the main clause. This is the way it has been shown to work in Icelandic (Zaenen, Maling & Thráinsson 1985; Barðdal 2006) and German (Haider 2005, 2010).

(ii) The second test focuses on the controller and relies on contexts in which only subjects may control PRO, more specifically, adjunct clauses (cf. Legendre 1989; Moore & Perlmutter 2000: 407; (Haspelmath 2001a): 73-74).

In what follows, I will discuss the relevance of both tests with respect to Romanian, starting with the first one.

3.2.4.2.1   Only subjects may be controlled PRO

Andrews (1976), Thráinsson (1979), and Zaenen, Maling & Thráinsson (1985) have argued for Icelandic that, in control contexts, the unexpressed argument of the embedded infinitive corresponds to the subject of the structure, had this been a finite clause, whereas the object is never affected by the process of embedding (Barðdal 2006: 163-164).

(77) a. *Ég$_i$*   *vonast*   *til*   ___$_i$ *að*   *fara*   *heim*
      I   hope   for   PRO   INF   go   home
      'I hope to go home'   (ex. 28a in Zaenen, Maling & Thráinsson 1985: 454)

    b. *Ég$_i$*   *vonast*   *til*   ___$_i$ *að*   *vanta*   *ekki*   *peninga*
      I   hope   for   PRO.ACC   INF   lack   not   money.ACC
      'I hope not to lack money'   (ex. 29b in Zaenen, Maling & Thráinsson 1985: 454)

    c. *Ég$_i$*   *vonast*   *til*   ___$_i$ *að*   *vera*   *hjálpa*
      I   hope   for   PRO.DAT   to   be.INF   helped
      'I hope to be helped'   (ex. 21a in Barðdal 2002: 71)

The examples in (77a-c) illustrate that, in Icelandic, PRO can correspond to an accusative (77b) or to a dative argument (77c). Given that only the subject of the embedded infinitive may be left unexpressed on coreference with an argument of the main clause,

examples (77b) and (77c) show that PRO may receive oblique case and that the accusative and dative arguments are syntactic subjects.

Before applying this test to non-canonical subjects, I will first check whether in Romanian too only subjects may be PRO. The following examples show that this is the case. Indeed, in the attested examples in (78) and (79), only the subject of *acționa în judecată* 'sue' and of *invita* 'invite' is left unrealized and not the object, although it is an argument of the main verb.

(78) **El$_i$** **ne$_j$** *obligă* *a* *-l$_{i/k}$* *acționa* *în* *judecată* ___$_j$/
    he   us.ACC   obliges   INF   =him.ACC   proceed   in   trial   PRO /
    *a *actiona* *în* *judecată* **noi$_j$**
    INF   proceed   in   trial   we
    'He obliges us to sue him'   (forumgratuit.ro)

(79) ___$_i$ **L$_j$** *a* *încurajat* *pe* *Walt$_j$* *să* *-l$_{i/k}$* *invite*
    pro   him.ACC=   has   encouraged   ACC   Walt   SUBJ   =him.ACC   invite.3SG
    ___$_j$ [...] / **să* *invite* **el$_j$**
    PRO   [...] /   SUBJ   invite.3SG   he
    'He encouraged Walt to invite him'   (books.google.be)

Additionally, observe that in these examples, the unrealized subject cannot be a *pro*, since it cannot be realized by an NP or a strong pronoun.

The same holds for the obligatory control structures in complement position (80), assigned by obligatory control verbs (*reuși* 'manage', *îndrăzni* 'dare' or *încerca* 'try').[44] As explained in Section 3.2.4.1, in these structures, the unrealized subject of the embedded clause is obligatorily controlled by the subject of the main clause (cf. Comrie 1984).[45]

(80) **Un** *simplu* **ajutor$_i$** *de* *tipograf* *a* *reușit* *a* *imita*
    a   simple   apprentice   of   pressman   has   managed   INF   reproduce
    ___$_i$/ *(*el)* *la* *perfecție* *stilul* *marelui* *autor*
    PRO /   (*he)   at   perfection   style.the   big.the.GEN   author
    'A simple apprentice has managed to reproduce flawlessly the style of the great author'
    (nemira.ro)

We may conclude with respect to Romanian that, in control structures with infinitives or subjunctives, the subject must remain unexpressed under identity with an argument in the main clause, whereas the object must be expressed.

---

[44] With respect to *încerca* 'try' (and to volitional verbs such as *dori* 'wish', in the pre-21st century periods), it has to be noted that when it occurs with the infinitive, it generates an obligatory-control structure. When it occurs with the subjunctive, two readings are possible: the coreferential and the non-coreferential reading.

[45] Comrie (1984) has shown (for English and for German) that in complement clauses control of the unexpressed subject (PRO) is lexically determined. This holds for Romanian too, where, depending on the verb, PRO may be controlled by the subject (e.g. *reuși* 'manage', *îndrăzni* 'dare, *încerca* 'try'), the direct object (e.g. *ajuta* 'help', *convinge* 'persuade'), or the indirect object (e.g. *(re)aminti* 'remind', *impune* 'impose').

Let us now apply this subject test to accusative and dative experiencers of embedded infinitives or subjunctives in Romanian. The examples in (81)-(82) show that the accusative or dative pronoun referring to the experiencer of the embedded infinitive (81a-b) or subjunctive (82a-b) must be expressed, and hence seem to argue against the subject status of these experiencer arguments.

(81) a. ? **$Ştefan_i$**   *a*   *reuşit*   *a*   *nu*   **(-l$_i$)*   *surprinde*   *ceea*   *ce*

Ştefan   has   managed   INF   not   =him.ACC   surprise   that   what

se   *întâmplă*

SE   happens

'Ştefan managed not to be surprised about what happens'

    b. ? **$Ştefan_i$**   *a*   *reuşit*   *a*   *(-i$_i$)*   *plăcea*   *de*   *colega*   *lui*

Ştefan   has   managed   INF   =him.DAT   please   of   colleague.the   his

'Ştefan managed to like his colleague'

(82) a. **$Ştefan_i$**   *a*   *reuşit*   *să*   *nu*   *-l$_i$*   *surprindă*   *ceea*   *ce*

Ştefan   has   managed   SUBJ   not   =him.ACC   surprise.3SG   that   what

se   *întâmplă*

SE   happens

'Ştefan managed not to be surprised about what happens'

    b. ? **$Ştefan_i$**   *a*   *reuşit*   *să*   *-i$_i$*   *placă*   *de*   *colega*   *lui*[46]

Ştefan   has   managed   SUBJ   =him.DAT   please.3SG   of   colleague.the   his

'Ştefan managed to like his colleague'

The ungrammaticality of these examples when used without the clitic can be explained by the obligatoriness of clitics in experiencer constructions. Indeed, in Romanian, experiencer accusative and dative clitics must be present, even when they co-occur with fronted NPs or strong pronouns in clitic-doubling constructions (cf. Geber 2011: 72).[47]

---

[46] A similar example was completely rejected by Alboiu (2007: 25). The scholar prefers a structure where the first subject of the main clause is encoded in the dative. However, certain Romanian native speakers, including myself, consider examples of the type illustrated in (82b) acceptable.

[47] Note that, in Romanian, examples can be found in which the clitic experiencer is not obligatory when the dative experiencer occurs in postverbal position (cf. examples (62b-c) in Section 3.2.3). Indeed, (Rivero 2009: 2) shows that in the 18th century postverbal NP and strong pronoun experiencers may occur without clitic doubling (cf. Rivero's example 30, given here in (i)).

(i)   *Că*   *aşa*   *place*   *lui*   *Dumnezeu*

    Because   so   pleases   DAT   God

    'Because in this way it pleases God'   (A. Ivireanul, Didahii, 1710)

Geber (2011: 147) also mentions the existence of such examples, and postulates a semantic difference between the constructions with preposed, clitic-doubled experiencers and the construction with postposed, not doubled experiencers, the former being more specific, whereas the latter conveying a generic meaning.

Therefore, clitics have been analyzed as case markers (cf. Manoliu-Manea 1987), or the 'spell-out' of certain morpho-syntactic features of the verb (Monachesi 2000).[48]

Nevertheless, the examples in (81) and (82) above may be seen as evidence in favor of a subject analysis of the accusative and dative pronoun contained in the embedded clause. Indeed, the verb in the main clause is *reuși* 'manage', which obligatorily controls the subject of the embedded clause. In the examples in (81) and (82), the accusative (a) and dative (b) experiencers are clearly controlled by the subject of the main clause since they are coreferential with it.[49] Although such examples are not attested, they are accepted by certain speakers of Romanian when the clitic denoting the experiencer is expressed, which argues in favor of its subject status. This is also confirmed by the fact that, in the (b) examples above, the stimulus is encoded as a PP, and not as a nominative NP.

Based on the evidence presented above, I conclude that dative and accusative experiencers behave like subjects of the embedded clause in that they may be controlled by the subject of verbs such as *reuși* 'manage', *îndrăzni* 'dare, *încerca* 'try', which are obligatory subject control verbs. Therefore, the control test under scrutiny, which focuses on the controlled subject, is relevant and conclusive in Romanian.

3.2.4.2.2   Only subjects may control PRO

The second use of control into non-finite clauses as a subject test concerns configurations in which only a subject may be a controller. These contexts are limited since, in complement clauses, control of the unexpressed subject is lexically determined (cf. Comrie 1984), as already mentioned in the previous section. Hence, PRO may be controlled by a subject, a direct or an indirect object, depending on the verb of the main clause.

As argued for certain languages such as French (cf. Legendre 1989), German (cf. Haspelmath 2001a: 73-74), and Russian (cf. Moore & Perlmutter 2000: 407), a configuration allowing only control by the main subject is the adjunct clause. Indeed, in

---

[48] A similar analysis has been proposed for Spanish (Enrique-Arias 2000, Ormazabal & Romero 2007), Macedonian (Franks 2007), and Semitic languages (Khan 1984).

[49] It has to be mentioned that *placea* 'please' and *suprinde* 'surprise' may also occur in regular control infinitive clauses, with nominative PRO.

(i)	*Marius*ᵢ	*a*	*reușit*	*a*	*-i*	*plăcea*	___ᵢ	*[lui Ion]*
	Marius	has	managed	INF	=him.ACC	please	PRO	[DAT Ion]
	'Marius managed to please Ion'							
(ii)	*Marius*ᵢ	*a*	*reușit*	*a*	*-l*	*surprinde*	___ᵢ	*[pe Ion]*
	Marius	has	managed	INF	=him.ACC	surprise	PRO	[ACC Ion]
	'Marius managed to surprise Ion'							

Hence, the control test shows that these verbs have two argument structures: Dat-Nom and Nom-Dat and are, thus, alternating predicates (Barðdal, Eythórsson & Dewey 2019).

these languages, the subject of adjunct infinitives and gerunds is obligatorily controlled by the subject of the main clause.

However, in Romanian this constraint is less clear with regard to adjunct infinitive clauses. In most infinitive adjunct clauses the unexpressed subject is indeed controlled by the subject of the main clause (83a), but examples can be found in which this is not the case (83b-c). For instance, in (83b and c) the PRO in the adjunct infinitive clause may be controlled respectively by the direct object or the dative recipient situated in the main clause, or by the subject. This shows that control into adjunct infinitives cannot distinguish between subjects and direct objects, or goal datives.

(83) a. ***Clara*$_i$** *i*$_j$- *a* *pregătit* *o* *prăjitură* *pentru* *a* *-l*
Clara him.DAT= has prepared a cake in_order INF =him.ACC
încuraja ___$_{i/^*j}$
encourage PRO
'Clara baked him a cake in order to encourage him

b. ***Mama*$_i$** *m*$_j$- *a* *simţit* *înainte* *de* *a* *intra* ___$_{i/j}$
mother.the me.ACC= has felt before of INF enter PRO
'Mother sensed me before entering in the house (before she / I entered in the house)'

c. ***Mama*$_i$** *mi*$_j$- *a* *dat* *un* *sandwich* *înainte* *de* *a*
mother.the me.DAT= has given a sandwich before of INF
pleca ___$_{i/j}$
leave PRO
'Mother gave me a sandwich before leaving (before she left / I left)'

A different picture emerges from the example in (84), where the same control test is applied to adjunct gerund clauses. The unexpressed subject of the adjunct gerund clause must, indeed, be controlled by the subject of the main clause in this example.

(84) ***Mama*$_i$** *a* *invitat* *-o* *pe* *Roxana* *înăuntru* *văzând* ___$_{i/^*j}$
mama has invited =her.ACC ACC Roxana inside seeing PRO
că *mai* *era* *timp*
that more was time
'Mother invited Roxana inside, seeing that there was still time'

The examples in (85a-b), containing an accusative or a dative experiencer in the main clause, show that this experiencer may also control the implicit subject of gerund adjuncts and, hence, confirm the validity of this test in Romanian.

(85) a. ***Mă*$_i$** *chinuia* *regretul*$_j$ *văzându* *-l* ___$_{i/^*j}$ *aşa* *trist*
me.ACC tortured.3SG regret.the seeing =him.ACC PRO so sad
'The regret was torturing me seeing him so sad'

b. ***Lui* *Victor*$_i$** *îi*$_i$ *plăcu* *fata*$_j$ *şi* *mai* *mult,*
DAT Victor him.DAT pleased.3SG girl.the even more much
auzind *-o*$_j$ ___$_{i/^*j}$ *cum* *cântă*
hearing =her.ACC PRO how sings
'Victor liked the girl much more after hearing her singing'

Given the evidence presented in Sections 3.2.4.2.1 and 3.2.4.2.2, I conclude that both tests related to control confirm that accusative and dative experiencers behave like canonical subjects. Section 3.2.4.2.1 has shown that, although they must be realized as clitics, they may be controlled by the subject of obligatory subject control verbs, while Section 3.2.4.2.2 has shown that they can occur in the position of a controller of PRO of an embedded clause in configurations which allow only control by subjects. In other words, accusative and dative experiencers may be controlled subjects of non-finite verb forms and controllers of PRO in configurations where this is only possible with subjects.

3.2.5 Raising

Another property of core arguments in many languages is the ability to raise. Raising, basically, occurs when a verb (without a subject of its own) takes over the subject of its complement clause. This "movement" is triggered by the finite predicate, known as *raising predicate*. The term *raising* has its origins in generative grammar: the constituent in question is assumed to "raise" from its initial, lower position as the subject of the non-finite verb, to the position of subject or object of the finite verb.

Van der Auwera & Noël (2011) distinguish three types of raising structures: (i) *Object-to-subject raising*, also called *tough* construction, in which the object of the non-finite clause raises to the position of subject of the finite clause as in (86). This kind of raising is triggered by adjectives meaning 'difficult' or 'easy'.

(86) *Această carte este uşor de citit*
    this   book is   easy SUP read
    'This book is easy to read'

(ii) *Subject-to-subject raising*, in which the subject raises from its lower position to the subject position of the finite verb, as in (87a-b). This kind of raising is found in most languages with verbs such as *appear, look, seem,* etc (cf. Van der Auwera & Noël 2011). In Romanian, it typically occurs with verbs such as *părea* 'seem' (87a) or *se întâmpla* 'happen' (87b).

(87) a. *Clara pare a citi o carte*
      Clara   seems   INF read   a   book
      'Clara seems to read a book'
    b. *Ştefan s- a întâmplat să ajungă mai devreme*
      Ştefan   SE= has happened SUBJ arrive.3SG more early
      'Ştefan happened to arrive earlier'

(iii) *Subject-to-object raising*, in which the subject of the lower verb becomes the object of the finite clause, as in (88). It is found with verbs such as *consider, believe, find,* etc. In Romanian, it occurs with verbs such as *crede* 'believe' (88).

(88) | Nu | -l | credeam | să | fie | capabil | de | atacuri | mișelești |
|---|---|---|---|---|---|---|---|---|
| not | =him.ACC | believed.1SG | SUBJ | be.3SG | capable | of | attacks | coward |

'I didn't think he was capable of coward attacks' (ziaruldevrancea.ro)

Two out of the three types of raising, namely subject-to-subject raising (87a-b) and subject-to-object raising (88), can be used as subjecthood tests. Hence, I will focus on these two types of raising, and their selecting predicates.[50]

3.2.5.1 Subject-to-subject raising

In Romanian, subject-to-subject raising is found with aspectual verbs (*începe* 'begin') and modals (*părea* 'seem', *putea* 'can', *trebui* 'must') (cf. Cotfas 2016). Raising verbs differ from control verbs in that they do not select, semantically, the raised argument, as shown in (89a-b) (see Van Riemsdijk & Williams 1986: 130, Borsley 1996: 133, Culicover 1997: 102), whereas a control predicate does (90a-b).[51]

(89) a. | Clara | pare | a | citi | o | carte | (Raising) |
|---|---|---|---|---|---|---|
| Clara | seems | INF | read | a | book | |

'Clara seems to read a book'

b. | *Clara | pare |
|---|---|
| Clara | seems |

(90) a. | Clara | îndrăznește | a | visa | (Control) |
|---|---|---|---|---|
| Clara | dares | INF | dream | |

'Clara dares to dream'

b. | Clara | îndraznește |
|---|---|
| Clara | dares |

'Clara dares'

---

[50] Romanian seems to be situated somewhere in between English and Greek with regard to the number of attested raising predicates. English is known to have several raising triggers (Noël 2001), while Greek is shown to have only the aspectuals *arxizo* 'start' and *stamatao* 'stop', alongside the verb *fenete* 'seem' (Alexiadou & Anagnostopoulou 2002). As for Romanian, the most clear raising trigger is *părea* 'seem', but modals and aspectuals are mostly also included among the raising predicates (Alboiu 2007, Cotfas 2016). However, modals such as *putea* 'can' and certain aspectuals, such as *prinde* 'start' occurring with an infinitive behave like complex predicates and are close to auxiliaries. Indeed, they show several signs of cohesion which are not found with other raising verbs like *părea* 'seem', among which, obligatory coreference, raising of object clitics, attachment of the negation to the head verb, absence of the infinitival particle *a* (specific to *putea* 'can') (cf. Nedelcu 2016: 243).

[51] The distinction between raising and control is highly controversial in Romanian, and also in recent studies on Balkan languages, where these mechanisms are also applied to subjunctive complement constructions. Typical obligatory control verbs (e.g. *îndrăzni* 'dare' or *încerca* 'try') have also been accounted for in terms of raising (cf. Alboiu 2007, Nicolae 2013: 247-262), whereas typical raising verbs such as aspectuals and modals have been accounted for in terms of obligatory control (cf. Landau 2004: 835). Since this controversy is not relevant here, I will not look further into it.

In languages such as Icelandic and Spanish, subjects can raise out of infinitive clauses. In Romanian, this phenomenon is sporadic because of the rareness of the infinitive clause (cf. 3.2.4.1). Nevertheless, several scholars (Rivero & Geber 2003, 2004; Geber 2006, 2011; Alboiu 2007, and Rivero 2009, among others) argue that Romanian allows the subject to raise not only out of infinitive clauses (cf. 91a), but also out of subjunctive clauses (cf. 91b).[52] In configurations with a subjunctive, the nominative subject of the subjunctive may remain in situ, i.e. postposed to the lower verb, or raises into preverbal position before the raising predicate, as in (91c).

(91) a. *Clara   pare   a   citi   o   carte*
      Clara   seems   INF   read.   a   book
      'Clara seems to read a book'
   b. *Ştefan   s-   a   întâmplat   să   ajungă   mai   devreme*
      Ştefan   SE=   has   happened   SUBJ   arrive.3SG   more early
      'Ştefan happened to arrive earlier'
   c. *(Marius)   poate /   trebuie /   începe   să   deseneze   (Marius)*
      (Marius)   can /   must /   begins   SUBJ   draw.3SG   (Marius)
      foarte   bine   (Marius)
      very   well   (Marius)
      'Marius can / must / is beginning to draw very well'

The question now arises whether accusative and dative experiencers may also raise in Romanian, and hence behave like subjects. The following examples show that, in present-day Romanian, experiencers cannot raise in Romanian (92a-b) to subject position when they are realized as only clitics. However, they can raise when they are realized as NPs (93a) or strong pronouns (93b), case in which they are doubled by a clitic that remains attached to the lower verb.

(92) a. **Îi   pare   să   placă   şcoala*
      him.DAT   seems   SUBJ   please.3SG   school.the
   b. **Te   pare   a   deranja   ceva*
      you.ACC   seems   INF   disturb   something
(93) a. *Lui   Mihai   pare   să   -i   placă   şcoala*
      DAT   Mihai   seems   SUBJ   =him.DAT   please.3SG   school.the
      'Mihai seems to like school'
   b. *Pe   tine   pare   a   te   deranja   ceva*
      ACC   you.ACC   seems   INF   you.ACC disturb   something
      'You seem to be disturbed by something'   (hotnews.ro)

In contrast, in the example in (94), the only instantiation of raising out of an infinitive clause found in my corpus for old Romanian, an example which dates from the end of

---

[52] Geber (2011) even distinguishes a third raising construction, out of indicative *că*-clauses, which I will not consider here.

the 18th century, the clitic experiencer raises together with the strong pronoun. This may suggest that, in old Romanian, the clitic could raise, together with the NP or the strong pronoun. Moreover, Rivero (2009: 2) shows that old Romanian permits NP and strong pronoun experiencers to occur without clitic doubling (cf. Rivero's example 30, given here in 95), as opposed to present-day Romanian, which requires the presence of clitics in most experiencer constructions (with the exception of quantificational and overt / null generic dative experiencers, as indicated by Rivero 2009: 8).

(94)	*mie*	*-m*	*începe*	*a*	*plăcea*	*poeticul*	
	me.DAT	=me.DAT	starts	INF	please	poetry.the	
	'I start to like the poetry'						(1760-1820_Budai Deleanu)
(95)	*Că*		*aşa*	*place*	*lui*	*Dumnezeu*	
	because		so	pleases	DAT	God	
	'Because in this way it pleases God'						(1710, A. Ivireanul, *Didahii*)

The ungrammaticality of the examples in (92a-b) above, where the accusative and dative experiencers are realized by clitics only, may certainly be explained by the obligatoriness of the clitics in definite and specific experiencer constructions in 21st century Romanian (see Dobrovie-Sorin 1993, Cornilescu 2004, cited by Giurgea 2017: 284; see also Rivero 2009).

At a closer investigation, besides subject-like arguments, other elements can occur in front of the raising predicate, in Romanian, namely focused direct objects (96a) and dative recipients (96b).

(96) a.	*Pe*	*Marius*	*pare*	*a*	*-l*	*fi*	*invitat*	*mai*	*devreme*
	ACC	Marius	seems	INF	=him.ACC	be	invited	more	early
	'S/he seems to have invited Marius earlier'								
b.	*Băiatului*		*pare*	*să*	*-i*	*fi*	*dat*	*un*	*cadou*
	boy.the.DAT		seems	SUBJ	=him.DAT	be.3SG	given	a	gift
	mai		*frumos*						
	more		beautiful						
	'To the boy s/he seems to have given (him) a nicer gift'								

Similarly, it has been shown that, in Icelandic and in German, elements such as adverbials, or a dummy subject can occupy this position, besides subject-like obliques. In spite of that, Barðdal (2006: 52) argues that in Icelandic, this evidence does not discredit the reliability of the subject-to-subject raising test. Instead, the fact that subject-like obliques occur in the same position as nominative subjects is considered as evidence in favor of their subject status.

In order to differentiate between these elements, I investigate if all fronted elements occupy the subject position of the raising predicate in Romanian, or if some of these elements occupy a focus position. Whether an element occurs in a focus or in a subject position can be tested by observing the behavior of quantifiers, which cannot occur in a

focus position (Rizzi 1982; Rizzi 1986; Cardinaletti 2004; Geber 2011), but are suitable in subject position (Rizzi 2005: 211).

In the following examples, the accusative and the dative experiencers containing a quantifier (97a-b) may raise into subject position, contrary to direct objects (98a) or recipient datives (98b) containing such a quantifier. The presence of quantifiers in these structures shows that the accusative and dative experiencers occur in subject position, just like the nominative *nimeni* 'nobody' in (99), and not in a focus position.

(97) a. *Pe*   *nimeni*   *nu*   *pare*   *a*   *îngrijora*   *ceva*
      ACC   nobody   not   seems   INF   worry   something
      'Nobody seems to worry about anything'
    b. *Nimănui*   *nu*   *pare*   *să*   *-i*   *placă*   *cartea*
      nobody.DAT   not   seems   SUBJ   =him.DAT   please.3SG   book.the
      'Nobody seems to like the book'

(98) a. **Pe*   *nimeni*   *nu*   *pare*   *să*   *-l*   *fi*   *invitat*   *mai*
      ACC   nobody   not   seems   SUBJ   =him.ACC   be.3SG   invited   more
      devreme
      early
    b. **Nimănui*   *nu*   *pare*   *să*   *-i*   *fi*   *dat*   *un*   *cadou*
      nobody.DAT   not   seems   SUBJ   =him.DAT   be.3SG   given   a   gift

(99) *Nimeni*   *nu*   *pare*   *să*   *înțeleagă*   *ceva*
    nobody   not   seems   SUBJ   understand.3SG   something
    'Nobody seems to understand something'

These examples show that the presence of the quantifier represents a reliable method to distinguish topicalized structures from neutral ones in the case of subject-to-subject raising. Furthermore, they confirm that the accusative and dative experiencers are not focused elements, but occur in subject position.

3.2.5.2 Subject-to-object raising

Subject-to-object raising is triggered in Romanian by predicates such as *crede* 'believe' and *şti* 'know', or *pune* 'put'. In contexts with *crede* 'believe', Romanian allows the subject to raise out of infinitive (100a), supine (100b), and participle (100c) clauses, and even out of subjunctive clauses (100d).[53]

(100) a. *Pe*   *tine*   *te$_i$*   *credeam*   *a*   *fi*   *de*   *încredere*
       ACC   you.ACC   you.ACC   believed.1SG   INF   be   of   trust
       'I thought you were trustworthy'   (diacronia.ro)

---

[53] Note that this is not the case with *şti* 'know', and with *pune* 'put, let', the raising possibilities being more limited with these verbs.

b. *Maşina   asta   o   credeam   de   furat*
   car.the   this   it.ACC believed.1sg   SUP   stolen
   'I thought this car was stolen'   (diacronia.ro)

c. *Vă   credeam   plecaţi*
   you.ACC   believed.1SG   gone.MASC.PL
   'I thought you were gone'   (scribd.com)

d. *Nu   -l   credeam   să   fie   capabil   de   atacuri*
   not =him.ACC believed.1sg SUBJ be.3SG   capable   of   attacks
   mişeleşti
   coward
   'I didn't think he was capable of coward attacks'   (ziaruldevrancea.ro)

Observe that in canonical contexts with clitic doubling such as (100a-b), the clitic raises together with the NP to the object position in the main clause. The fact that Romanian allows raising-to-object brings it closer to French (cf. 101a), to Brazilian Portugese (cf. 101b adapted from Martins & Nunes 2005: 13, example 19a), and even to Icelandic (cf. 101c, from Barðdal 2006, example 11).

(101) a. French:
       Je   te   croyais   partie
       I   you.ACC believed.1SG   gone.FEM.SG
       'I thought you were gone'
       b. Brazilian Portuguese:
       Eu   convenci   -a   a   viajar
       I   convinced.1SG   =her.ACC   INF   travel
       'I convinced her to travel'
       c. Icelandic
       Ég   laet   hana   borða   dagblað
       I   let.SG her.ACC   eat.INF   newspaper
       'I'll have her eat a newspaper.'

As for accusative and dative experiencers, the situation is different in Romanian, than, for instance, in Icelandic. As shown in the following examples, in contexts with clitic doubling, only NPs may raise to the object position of the main clause in Romanian, whereas the clitic must remain attached to the verb (102a-b). As for contexts without an overt NP as in (103a-b), they become ungrammatical when the subject-like clitic raises to the main clause.

(102) a. *Pe   băiat (*îl)   credeam   să   *(-l)   intereseze*
       ACC boy   him.ACC   believed.1SG SUBJ =him.ACC   interest.3SG
       matematica
       maths.the
       'I believed the boy to be interested in maths'

b. *Băiatului* (**îi*) *credeam* *să* **(-i)* *placă*
boy.the.DAT him.DAT believed.1SG SUBJ =him.DAT please.3SG

muzica
music.the

'I believed the boy to like the music'

(103) a. **îl* *credeam* *să* **(-l)* *intereseze* *matematica*
him.ACC believed.1SG SUBJ =him.ACC interest.3SG maths.the

b. **îi* *credeam* *să* **(-i)* *placă* *muzic-a*
him.DAT believed.1SG SUBJ =him.DAT please.3SG music.the

These examples show, on the other hand, that in Romanian, like in other languages, the subject-like oblique NP maintains its case marking when it raises to the object position of the main clause (102a-b) above.

Interestingly, the raised NPs occupy in Romanian a clause-initial position. As shown in the case of subject-to subject raising, this position can be occupied by other elements too, among which, focused accusative objects and dative recipients of the embedded verb, as shown in (104)-(105) below.

The question now arises what can differentiate in Romanian between these raised subject-like accusative or dative NPs (cf. 102 above) and focused accusative objects and dative recipients (104)-(105) below. In Icelandic, as opposed to German, for instance, raising-to-object of oblique subjects may be distinguished from topicalized objects of the embedded verb relying on word order (cf. Barðdal 2006: 48). In Romanian – a language in which word order is relatively free and sensitive to information structure – this tool is not available. For this reason, I rely on intonation, which may help in distinguishing between the two structures. More precisely, if focused accusatives and datives bear nuclear stress (marked in small caps and bold), this can suffice in distinguishing them from contexts with accusative or dative subject-like arguments, which, being neutral, may not bear nuclear stress. Therefore, following Giurgea (2017), I compare in the subsequent examples the intonation in two types of focus contexts, contrastive and non-contrastive focus.

(104) Non-contrastive focus

a. ***Pe*** ***Mihai*** *am* *crezut* *că* *l-* *a* *invitat* *mai*
ACC Mihai have.1SG believed that him.ACC= has invited more

devreme
early

'I thought s/he has invited Mihai earlier'

b. ***Lui*** ***Mihai*** *am* *crezut* *că* *i-* *a* *dat* *cartea*
DAT Mihai have.1SG believed that him.DAT= has given book.the

'I thought s/he gave the book to Mihai'

(105) Contrastive focus

a. ***Pe*** ***MiHAI***   *am*   *crezut*   *că*   *l-*   *a*   *invitat*
   ACC   Mihai   have.1SG   believed   that   him.ACC=   has.3SG   invited

   mai   *devreme*   ***nu***   ***pe***   ***Marius***
   more   early   not   ACC   Marius

   'I thought s/he has invited Mihai, and not Marius, earlier'

b. ***Lui*** ***MiHAI***   *am*   *crezut*   *că*   *i-*   *a*   *dat*
   DAT   Mihai   have.1SG   believed   that   him.DAT=   has.3SG   given

   cartea   ***nu***   ***Mariei***
   book.the   not   Maria.DAT

   'I thought s/he gave the book to Mihai, not to Maria'

These examples provide evidence that, in Romanian, the two constructions, with and without contrastive focus, differ with respect to their intonation structure. Non-contrastive focused accusative objects and dative recipients do not bear nuclear stress (marked in small caps and bold), just like clause-initial nominative subjects in ordinary active declarative clauses, whereas contrastive accusative objects and dative recipients bear nuclear stress.

The lack of nuclear stress on the clause-initial focused accusative objects and dative recipients, in Romanian, makes it impossible to distinguish between subjects and focused objects of the embedded clause. Therefore, subject-to-object raising cannot be considered as a relevant subject test in Romanian.

To summarize, this section has shown that subject-to-subject, but not subject-to-object raising, represents a relevant subject test in Romanian. I have provided evidence that, on the only condition of being doubled by a clitic, NP and strong pronoun experiencers in the accusative or in the dative can raise both from infinitive clauses, even though this is very rare, and from subjunctive clauses. This confirms the subject-like behavior of these experiencers, and the reliability of subject-to-subject raising as a subject test in Romanian. As for the subject-to-object raising test, this is not relevant in Romanian, since there is no way to distinguish the raised oblique subjects from focused accusative objects or dative recipients of the embedded verb.

3.2.6   Conjunction reduction

Another diagnostic for subjecthood is conjunction reduction. Indeed, in coordination the subject of the second clause can be left unexpressed on identity with the subject of the first clause. However, in Romanian canonical structures, nominative subjects may always be left unexpressed due to the pro-drop character of Romanian, whether occurring in coordination or not, as shown in (106a-b).

(106) a. ***Clara***   *a*   *venit*   *acasă*   *și*   ___   *a*   *pregătit*   *cina*

       Clara   has   come   home   and   *pro*   has   prepared   dinner.the

       'Clara came home and prepared dinner'

    b. ***Clara***   *a*   *venit*   *acasă.*   *După*   *câteva*   *minute*   ___   *a*

       Clara   has   come   home.   After   few   minutes   *pro*   has

       pregătit   *cina*

       prepared   dinner.the

       'Clara came home. After a few minutes, she prepared dinner'

As for non-canonical structures with accusative or dative subject-like arguments, this test yields a negative result because the clitic cannot be deleted as in (107)-(108), another consequence of the obligatoriness of clitics in Romanian. Therefore, this test is not applicable in identifying a subject in this language.

(107) ***Roxana*_i**   *a*   *picat*   *la*   *toate*   *examenele*   *dar*   *încă*   **(o_i)**   *mai*

       Roxana   has   failed   at   all   exams.the   but   still   her.ACC   more

       interesează   *istoria*

       interests   history.the

       'Roxana failed at all the exams but is still interested in history'

(108) ***Mara*_i**   *a*   *sosit*   *târziu*   *și*   **(i_i)=**   *a*   *plăcut*   *că*

       Mara   has   arrived   late   and   her.DAT   has   pleased   that

       mâncarea   *era*   *gata*

       food.the   was   ready

       'Mara has arrived late and was pleased that food was ready'

3.2.7 Deletion of the subject in imperatives

Deletion in imperatives is regarded as only compatible with nominative subjects and, hence, not applying to non-nominative subjects (cf. Barðdal 2006: 60-63, for Icelandic, Reis 1982: 186, for German). Possible incompatibilities are semantically motivated and are due to the non-agentivity of the relevant verbs. Since oblique subjects occur, in these languages, with such non-agentive predicates, more specifically predicates expressing states, which can never be ordered, they do not allow imperatives (Rögnvaldsson 1996: 48, Barðdal 2006: 54). This also holds for Romanian, in that deletion of the subject in imperatives is compatible with nominative subjects of agentive predicates (109), and of some non-agentive predicates such as *iubi* 'love' (110).

(109)   *Scrie!*

       write.IMP.2SG

       'Write!'

(110)   *Iubește*   *-mă!*

       love.IMP.2SG   =me.ACC

       'Love me!'

When it comes to non-nominative subject-like arguments in Romanian, these are typically experiencers, since most of the predicates express states. Hence, they cannot be used in the imperative (111), which makes that the test is not applicable as a subject diagnostic.

(111) **Placă -ţi!*
   please.IMP =you.DAT

3.2.8 Deletion of the subject in the telegraphic style

As for the test of deletion in the telegraphic style, Reis (1982: 190) considers that this test can apply to nominative subjects in German, but does not work with non-nominative ones. Notwithstanding, this test has been shown to be relevant in Icelandic, where a nominative subject, as well as an oblique experiencer can be left unexpressed in the telegraphic style, without resulting in ungrammaticality (Barðdal 2006: 55).

 To my best knowledge, this test has not been discussed in the literature on Romanian, most likely because of the pro-drop character of Romanian. Indeed, pronominal subjects referring to the speaker are always unexpressed. Nevertheless, as a native speaker I consider that, in structures with psychological verbs and non-nominative subject-like arguments, it is the subject-like argument that is left unexpressed in the telegraphic style, just like it is the case in canonical structures, and not the postverbal nominative NP. Hence, the most natural telegraphic version of a sentence like (112a) would be *plăcut fata* (lit. pleased girl.the.NOM), leaving unexpressed both the clitic and the auxiliary, as in (112b), rather than *mi-a plăcut* (lit. me.DAT has pleased), where leaving the postverbal nominative NP unexpressed leads to ambiguity. Observe that non-canonical structures pattern with canonical structures in this respect: the experiencer in (112a) patterns with the *pro* subject in (113a) in that they may be deleted, whereas both the direct object in (113b) and the target of the feeling (or the stimulus) in (112b) must be realized, in order to avoid ambiguity.

(112) a. *Mi-*   *a* *plăcut*   *fata*
    me.DAT= has pleased   girl.the
   b. *Plăcut*   **(fat-a)*
    pleased   girl-the
    'I liked the girl'
(113) a. *Am*   *trecut*   *examenul*
    have.1SG passed   exam.the
   b. *Trecut*   **(examenul)*
    passed   exam.the
    'I have passed the exam'

Based on the comparison of these examples, I advocate, in spite of the possible counter argument concerning the pro-drop characteristic of Romanian, that the deletion of subjects in telegraphic style is a reliable test in distinguishing between subjects and objects in Romanian.

3.2.9   Bare quantifiers in clause-initial position

The behavior of bare quantifiers is very helpful in assessing whether an element in clause–initial position is a focused element or not. It is generally acknowledged that quantifiers cannot occur in a focus position (Rizzi 1982, 1986; Cardinaletti 2004; Geber 2011), but are very suitable in subject position (Rizzi 2005: 211).

Romanian dative and accusative experiencers can be expressed by a bare quantifier in clause-initial position. Throughout the thesis, I use the substitution with bare quantifiers with respect to some subject tests, in order to rule out the possibility of an analysis of the experiencer as a fronted element. Some scholars, however, have used the clause-initial position of bare quantifiers as a subject test of its own (cf. Cuervo 2010 for Spanish, and Dumitrescu & Masullo 1996 cited in Rivero 2009 for Romanian).

In the following examples, the accusative and the dative experiencers realized by a bare quantifier (114a and c) may occupy a clause-initial position and have a neutral reading, contrary to direct objects (114b) or goal datives (114d).

(114) a. *Pe* *nimeni* | *nu* | *pare* | *a* | *îngrijora* | *ceva*
ACC nobody | not | seems | INF | worry | something
'Nobody seems to worry about anything'

b. **Pe* *nimeni* | *nu* | *l-* | *a* | *invitat* | *mai* *devreme*
ACC nobody | not | =him.ACC | has | invited | more early

c. *Nimănui* | *nu* | *-i* | *place* | *cartea*
nobody.DAT | not | =him.DAT | pleases | book.the
'Nobody likes the book'

d. **Nimănui* | *nu* | *-i* | *a* | *dat* | *un* | *cadou*
nobody.DAT | not | =him.DAT | has | given | a | gift

e. *Nimeni* | *nu* | *înţelege* | *nimic*
nobody | not | understands | nothing
'Nobody understands anything'

The grammaticality and the neutral reading of the structures in (114a and c), which contain bare quantifiers, constitute evidence in favor of a subject analysis of accusative and dative experiencers in Romanian. In these structures, the bare quantifiers occur, just like the nominative *nimeni* 'nobody' in (114e), in subject position and not in a focus position. For these reasons, I consider the test of bare quantifiers in clause-initial postion as a relevant and conclusive criterion, in Romanian, in identifying between subjects and direct objects or dative goals.

3.2.10 Secondary predication

Another subject criterion is the possibility of an experiencer in the dative or in the accusative to take a depictive, secondary predicate. This property has been presented as a subject test in Spanish by Fernandez-Soriano (1999: 123-124), since the secondary predicate is mostly subject-oriented. In Romanian, accusative and dative experiencers occurring in psychological constructions show this property as in (115a and c), although some Romanian native speakers may consider these examples less natural.

(115) a.? Pe *Roxana*　　　　 *o*　　　　　 *interesa*　　　　 *tipul*　　　　 **beată**
　　　　　Roxana.ACC　　　　 her.ACC　　　 interested.3SG　 guy.the　　　 drunk.FEM.SG
　　　　　'Roxana was interested about the guy when she was drunk'
　　　b. *Pe　 *băieţi*　　　　　 *i-*　　　　　 *a*　　 *certat*　　 *obosiţi*
　　　　　ACC　 bois.ACC.MASC.PL　 them.ACC=　 has　 scolded　 tired.MASC.PL
　　　c. ? *Roxanei*　 *îi*　　　　 *plăcea*　　 *tipul*　　 **beată**
　　　　　Roxana.DAT　 her.DAT　　 pleased.3SG　 guy.the　 drunk.FEM.SG
　　　　　'Roxana liked the guy when she was drunk'
　　　d. *Fetei*　　　　　　 *i-*　　　　　 *a*　　 *dat*　　　 *o*　　 *carte*　　 **enervată**
　　　　　girl.DAT.FEM.SG　 her.DAT=　 has　 given　 a　　 book　　 irritated.FEM.SG
　　　e. *Fata*　　　　　　　 *a*　　 *sunat*　　 *la*　 *poliţie*　　 **speriată**
　　　　　Girl.the.FEM.SG　 has　 called　 at　 police　　 scared. FEM.SG
　　　　　'The girl called the police while scared'

The accusative and the dative experiencers in (115a) and (115c) above show the ability to take secondary predicates, since *beată* 'drunk' bears each time on the experiencer. The fact that these experiencers can take secondary predicates shows that they pattern with the subject in the canonical structure in (115e). They occupy, therefore, a higher position in the structure as opposed to the direct object in (115b) and to the goal dative in (115d), which do not take secondary predicates. Consequently, I consider the ability of accusative and dative experiencers to take secondary predicates as relevant evidence in favor of their subject status.

3.3 Conclusions

This chapter has provided a description of the properties of nominative subjects and of subject-like arguments in Romanian, based on the properties suggested by Keenan (1976), namely coding properties such as nominative case and verb agreement, and behavioral properties, such as word order, reflexive binding, control, raising,

conjunction reduction, deletion in imperatives, deletion in telegraphic style, bare quantifiers in clause-initial position, and secondary predication.

It is well known that the canonical subject in Romanian is encoded in the nominative and triggers verb agreement. However, these coding properties are not applicable to subject-like arguments, which are either dative- or accusative-marked and cannot trigger verb agreement. Nevertheless, several behavioral tests show that accusative and dative experiencers in non-canonical subject constructions pattern with canonical, nominative subjects. Among these, the most reliable ones are word order, control, raising-to-subject, deletion in telegraphic style, bare quantifiers in clause-initial position, and secondary predication. Three tests are applicable to Romanian, but they are non-conclusive due to certain language-specific features such as pro-drop, the obligatoriness of dative clitics in structures with specific reading, less strict binding principles, or on semantic grounds. The non-conclusive tests are: conjunction reduction, deletion in imperatives, and binding. As for the raising-to-object, this test is unable to distinguish between subjects and objects in Romanian. Table 3.10 gives an overview of the subject tests used in my analysis on accusative and dative experiencers in Romanian. The following notation has been used: ✓ = relevant, ✗ = not relevant, NC = applicable but non-conclusive (due to different constraints).

Table 3.10   Subject tests for oblique subjects in Romanian

Subject test	Relevance in Romanian
1. First position in declarative clauses	✓
2. Subject-to-subject raising	✓
3. Deletion in telegraphic style	✓
4. Control - Only subjects may be controlled PRO	✓
5. Control - Only subjects may control PRO	✓
6. Bare quantifiers in clause-initial position	✓
7. Secondary predication	✓
8. Reflexivization (binding)	(✓)
9. Conjunction reduction	NC
10. Deletion in imperatives	NC
11. Subject-to-object raising	✗

Table 3.10 indicates that seven out of eleven examined subject tests are reliable and conclusive in distinguishing between subjects and objects in Romanian. As for binding, (noted with (✓)), it is less convincing, snce binding rules are less strict in Romanian. Among the other three tests, two are non-conclusive due to some language-specific

constraints, although they are applicable (conjunction reduction, deletion in imperatives), while one test yields a negative result (raising-to-object).

In addition, in these oblique experiencer constructions, the stimulus, which is encoded in the nominative and mostly triggers verb agreement, and is hence, traditionally analyzed as the subject, is deficient with respect to these specific tests. Furthermore, the nominative subject in canonical structures, as well as the accusative and the dative experiencers in non-canonical structures coincide, in Romanian, with the leftmost argument of the argument structure of the verb, in a neutral, declarative sentence, what argues in favor of their subjecthood.[54]

---

[54] Cf. Eythórsson & Barðdal (2005) and their definition of subject.

Chapter 4   The MIHI EST construction as a complex predicate construction

This chapter situates the Romanian MIHI EST construction among other experiencer constructions, more particularly complex-predicate constructions. Complex-predicate constructions combine an experiencer and a complex predicate, which contains a light verb or a copula, and a predication (cf. Goldberg 1995, Michaelis & Lambrecht 1996: 242, Bickel 2004, Verhoeven 2007: 31–32, and Le Mair et al. 2017, who label it *compositional predicate*).[55]

In what follows, I use the term *complex-predicate constructions* when I refer to constructions containing a complex predicate. The *complex-predicate constructions* (116) differ from constructions containing psych verbs (117) in that the psychological meaning emerges from the combination of a light verb and a state with an experiencer, whereas psych verbs carry alone the psychological meaning of the construction. In *complex-predicate constructions*, the state is mostly realized by a noun, but it can also be realized by an adjective, or even by an adverb.

(116)   *The woman **has an aversion** for mushrooms*   (complex predicate construction)
(117)   *The woman **hates** mushrooms*   (psych-verb construction)

This chapter is structured as follows. After a brief introduction of two different types of complex-predicate constructions in Romanian, I concentrate my attention on the MIHI EST construction, and on its Latin ancestor, the MIHI EST pattern. The last section reiterates the research questions, and places the present study in the typological context of SAE languages.

---

[55] Also called *non-congruent constructions* (Verhoeven 2007: 88), or *metaphorical expressions* (Reh 1998: 11), these constructions are opposed to *congruent expressions*. In non-congruent constructions the verb refers to a different semantic domain than the whole expression, while congruent constructions contain psychological verbs of the types *temere*, *preoccupare* and *piacere* (Reh 1998: 11), cf. also Section 3.2.1.

4.1   Two types of complex-predicate constructions in Romanian

As shown by Van Peteghem & Ilioaia (In prep.), Romanian has several *complex-predicate constructions.* These constructions involve the copulative *fi* 'be', or light verbs such as *apuca* 'seize', *lua* 'take', *veni* 'come', *trece* 'pass', etc., which are not *psych* verbs by themselves, but may select a noun expressing a psychological state and an experiencer. For time and space considerations, I focus in the present work on constructions containing a state realized by a noun, excluding those containing a state realized by an adjective or an adverb. Depending on the determination of the state noun, two types may be distinguished: (i) *the determined state noun type* (118) and (ii) *the bare state noun type* (119).

(118)   *Mă   ia   somnul*
        me.ACC   takes   sleep.the
        'I start feeling sleepy'

(119)   *Mi-   e   dor*
        me.DAT=   is   longing.
        'I miss'

4.1.1   The determined state noun type

The determined state noun type is represented in Romanian by two constructions depending on the case of the experiencer, which may be the accusative or the dative. The accusative experiencer occurs with verbs of the CAPIO-type (from the Lat. *capio, capere* 'take, seize'), meaning 'take, seize' (cf. Rom. *apuca* 'seize', *lua* 'take', *trăsni* 'fulminate', *lovi* 'hit'), combination which enforces an inchoative interpretation to the state event (120).[56] The dative experiencer is used with cessative verbs (from the Lat. *cesso, cessare* 'cease, stop'), this combination contributing a cessative interpretation of the state event (cf. Rom. *trece* 'pass', *pieri* 'die', *se duce* 'leave from', *se tăia* 'to be cut off', *se potoli* 'calm down') (121).[57] The state noun is usually a definite NP, sometimes an indefinite NP.

(120)   *Mă   apucă   foamea*
        me.ACC   seizes   hunger.the
        'I start feeling hungry'

---

[56] For this use of the label inchoative, see Vivès (1984).

[57] I propose this label for the cessative aspectual meaning, by analogy with the inchoative aspectual meaning conveyed by certain verbs.

(121) *Mi-*   *a*   *trecut*   *somnul*
me.DAT=   has   passed   sleep.the
'I don't feel sleepy anymore'

I will label the construction with *CAPIO*-type verbs and accusative experiencers the *CAPIO inchoative construction*,[58] and the structure containing *cessative*-verbs occurring with dative experiencers the *cessative construction.* In both the *CAPIO inchoative* and the *cessative* constructions, the state noun is encoded in the (unmarked) nominative when realized by an NP (120) and (121) above, but it can also be realized by a clause (cf. 122 and 123).

(122) *Câteodată*   *mă*   *apucă*   *să*   *iau*   *câte*   *un*   *blogger*
sometimes   me.ACC   seizes   SUBJ   take.1SG   at_a_time   one   blogger
la   *puricat*
SUP   cleaned
'Sometimes I feel like criticising some blogger'   (aguritza.ro)

(123) *Mi-*   *a*   *trecut*   *să*   *experimentez*
me.DAT=   has   passed   SUBJ   experiment.1SG
'I don't feel like experimenting anymore'   (bloguluneitipeoarecare.wordpress.com)

Furthermore, the noun may take a complement denoting the stimulus of the state, mostly realized by a PP (124)-(125), or by a finite (126) or a non-finite clause (127).

(124) *Mă*   *apucă*   *dorul*   *de*   *tine*
me.ACC   seizes   longing.the   of   you.ACC
'I start missing you'

(125) *Mi-*   *a*   *trecut*   *dorul*   *de*   *ea*
me.DAT=   has   passed   longing.the   of   her
'I don't miss her anymore'

(126) *Mi-*   *a*   *trecut*   *pofta*   *să*   *experimentez*
me.DAT=   has   passed   craving.the   SUBJ   experiment.1SG
'I don't feel like experimenting anymore'

(127) *L-*   *a*   *luat*   *dorul*   *de*   *a*   *merge*   *acasă*
him.ACC=   has   taken   longing.the   of   INF   go   home
'He started longing for going home'

Benedetti (2013b) examines similar structures containing the verb *prendere* 'take, seize' in Italian, and shows that these constructions trace back to a classical Latin structure, in which the verb *capio* 'seize, take' occurs with an experiencer encoded in the accusative (128a), structure labeled the *"transitive" type*, or with a nominative experiencer (128b),

---

[58] Note that (Benedetti 2013a, 2013b) refers in her papers to Italian structures containing Latin *CAPIO*-type verbs, such as *prendere* 'take, seize' as the *prendere*-constructions and to corresponding Latin structures as the *capere*-constructions (Benedetti 2013b: 132). Nevertheless, she abandons these labels in favor of labels such as *transitive*, *plane*, or *inversion* constructions (cf. Benedetti 2013b: 124).

which she calls the *"plain" type*.[59] As pointed out by Benedetti (2013b: 122–124), a corresponding structure exists in ancient Greek as well, with the verb *lambánein* 'seize, take', which shows both the "transitive" type (129a), and the "plain" type (129b).[60] She assumes that the persistence of this construction in Latin has been reinforced by the existence of the corresponding construction in Greek. How far in time the CAPIO *inchoative* construction can be traced back is difficult to determine due to lack of sufficient data.

(128) a. *Metus*   *capit*   *Caesarem*   (Classical Latin)
          fear.NOM   seizes   Caesar.ACC
          'Caesar becomes afraid'   (example 3b from Benedetti 2013b: 123)
       b. *Caesar*   *metum*   *capit*
          Caesar.NOM   fear.ACC   seizes
          'Caesar becomes afraid'

(129) a. *Phóbos*   *lambánei*   *Théōna*   (Ancient Greek)
          fear.NOM   seizes   Théōn.ACC
          'Théōn becomes afraid'
       b. *Théōn*   *lambánei*   *phóbon*
          Théōn.NOM   seizes   fear.ACC
          'Théōn takes fear'

In the Romance daughter languages, both combinations have been preserved through the lexical replacement of *capio* 'seize take' with other corresponding verbs, meaning 'seize, take' (in Italian *prendere* 'take, seize', in Spanish *invader* 'invade', in French *saisir* 'grasp', *envahir* 'invade', or *prendre* 'take'). As for Romanian, verbs such as *prinde* 'catch', *lua* 'take', *apuca* 'seize', *cuprinde* 'grasp', *invada* 'invade' correspond to the Latin *capio* 'seize take'. These verbs occur, in the Romance daughter languages, either with an accusative experiencer (130a-c), or with a nominative one (131a-c).

(130) a. *M-*   *a*   *cuprins*   *frica*   (Romanian)
          me.ACC=   has   seized   fear.the
          'I became afraid'
       b. *La*   *peur*   *m'*   *a*   *saisi*   (French)
          the   fear   me.ACC=   has   seized
          'I became afraid'
       c. *Lo*   *prese*   *paura*   (Italian)
          him.ACC   seizes   fear
          'The fear seized him'

---

[59] Note that the pattern *metum capio* is the aspectually inchoative variant of *metum habeo* 'I have fear', with which it shares some syntactic properties (Benedetti 2013b: 124).

[60] As Benedetti (2013b: 126) points out, this verb corresponds semantically to Lat. *capere* 'seize take', although etymologically, they are not related.

(131)   a. *Băiatul*   *a*   *prins*   *frică*   (Romanian)
        boy.the   has   caught   fear
        'The boy has taken fear'

    b. *Le*   *garçon*   *a*   *pris*   *peur*   (French)
        the   boy   has   taken   fear
        'The boy has taken fear'

    c. *Il*   *ragazzo*   *prese*   *paura*   (Italian)
        the   boy   took   fear
        'The boy has taken fear'

In addition, Benedetti (2013b) signals that the *CAPIO inchoative* pattern develops in Modern Italian into a new structure, which she calls *the "inversion" type*, in which a *CAPIO*-type verb occurs with an experiencer in the dative (132) (Benedetti 2013b: 126).

(132)   *Gli*   *prese*   *paura*   (Italian)
        him.DAT   took   fear
        'He became afraid'

As for the origin of the cessative construction, things are not clear. Besides in Romanian (133), the construction is also attested in Italian (134a), Spanish (134b), and Portuguese (134c).[61]

(133)   *Mi-*   *a*   *trecut*   *somnul*   (Romanian)
        me.DAT   has   passed   sleep.the
        'I don't feel sleepy anymore'

(134)   a. *Mi*   *è*   *passata*   *la*   *fame*   (Italian)
        me.DAT   is   passed   the   hunger

    b. *Se*   *me*   *ha*   *pasado*   *el*   *hambre*   (Spanish)
        REFL   me.DAT   has   passed   the   hunger

    c. *Passou*   *-me*   *a*   *fome*   (Portuguese)
        passed.3SG   =me.DAT   the   hunger
        'I don't feel hungry anymore'

Nevertheless, a simple query in the Romanian corpus reveals that this construction occurs for the first time at the end of the 19th century. The available data suggest that the cessative construction is an innovation of the daughter languages rather than an inheritance from Latin, but this should be further investigated.

---

[61] I relied on the World Wide Web in order to verify the existence of such structures in these languages.

4.1.2   The bare state noun type – the MIHI EST construction

The bare state noun type is represented by the MIHI EST construction, containing the verb *fi* 'be', and a state noun (135a). This construction has also an inchoative version, in which a verb such as *veni* 'come' or *se face* 'become' combines with a state noun. In the present study, I label the inchoative construction the *VENIO inchoative* construction (135b).[62] In both constructions, the experiencer occurs in the dative, and the state noun is generally a bare noun or, very rarely, an indefinite NP.

(135) a. *Mi-*   *e*   *dor*   *de*   *casă*
       me.DAT   is   longing   of   home
       'I miss home'
    b. *Îmi*   *vine*   *somn /*   *mi*   *se*   *face*   *somn*
       me.DAT   comes   sleep /   me.DAT   SE   makes   sleep
       'I begin feeling sleepy'

Just like in the case of the determined state noun construction, the noun may take a complement denoting the stimulus of the state, realized either by a PP (136a), or by a finite or a non-finite clause, as in (136b-c).

(136) a. *Li-*   *era*   *milă* *de*   *acei*   *copii*
       them.DAT= was   pity   of   those   children
       'They had pity for those kids'
    b. *Mi-*   *a*   *fost* *frică* *să*   *pun*   *întrebarea*
       me.DAT=   has   been   fear   SUBJ   put.1SG   question.the
       'I was afraid to ask the question'
    c. *Mi-*   *era*   *groază* *văzând*   *atâtea*   *insecte*
       me.DAT=   was   terror   seeing   so_many   insects
       'I was terrified seeing so many insects'

It is worth mentioning the existence in Romanian of another structure of the MIHI EST type, without a state noun.[63] This structure contains a dative experiencer, the verb *fi* 'be' and a PP (137), a finite (138a), or a non-finite clause (138b).

(137) *Nu*   *mi-*   *e*   *de*   *părinți,*   *ci*   *mi-*   *e*   *de*   *copii*
     not   me.DAT=   is   of   parents   but   me.DAT=   is   of   children
     'I'm not concerned about the parents, but about the children'

---

[62] I label this construction the *VENIO inchoative* construction, given that this construction, containing an equivalent of the Latin verb *venio* 'come', occurs in several other languages with the same verb or an equivalent, and always conveys to the construction an inchoative meaning (for Italian, Fedriani 2011: 321; for Early Vedic, Danesi & Barðdal 2018: 15; for Nepali, Barðdal & Eythórsson 2018: 258).

[63] Bauer (2000: 180) mentions the existence in Early Latin of a so-called *verbal MIHI EST construction* (i), which could be the ancestor of the raising structure with the modal *fi* 'be', presented in this section.

(138) a. *Mi-*   *e*   *să*   *nu*   *se*   *supere*

     me.DAT=   is   SUBJ   not   him.ACC   upset.3SG

     'I'm afraid that he will mind' (= 'I don't want him to be upset')

  b. *Nu*   *mi-*   *e*   *a*   *glumi*   *despre*   *ceva*   *aşa*   *serios*

     not   me.DAT= is   INF   joke   about   something so   serious

     'I don't feel like joking about something so serious'

The structures in the examples above have in common the fact that the verb *fi* 'be', which is a neutral verb without a distinct meaning of its own, actualizes a modal meaning in combination with the constituent expressing the stimulus.[64] Indeed, as observed by Barbu (2015: 330–331), it is the predication that occurs with the verb *fi* 'be' that discriminates between the different meanings this verb can take, just like is the case with the corresponding verbs in other languages (for English, Borsley 1991: 143; for Russian, Augustinova 2006: 13).[65]

Following Palmer's. (2001) classification of event modalities, Barbu (2015: 320–321) postulates that the Romanian modal *fi* 'be' expresses either *deontic*, or *dynamic* modality. Deontic modality is externally conditioned and expresses a (physical or moral) constraint, necessity, possibility, permission, or predestination ('be meant for'), etc, as illustrated in Barbu's example 2, given here as (139). The dynamic modality, instead, refers to internal conditionings, and expresses the individual's willingness or ability to do something (cf. Barbu's example 3 given in 140).

(139) *dacă* **mi-**   *e*   **să**   *moriu (...)*   (Deontic modality)

     if   me.DAT   is   SUBJ   die.1SG

     'If it's meant for me to die ...'   (1760–1820, Budai-Deleanu)

(140) *Îmi*   *este*   *de*   *cântat*   (Dynamic modality)

     me.DAT   is   SUP   sung

     'I feel like singing.'

Focusing on the modal uses of *fi* 'be', Barbu (2015) distinguishes between two main types of structures: (i) structures with a raising *fi*, containing the modal *fi* with a deontic meaning (139), and (ii) structures with a control *fi*, which contain the modal *fi* with a dynamic meaning (140).

Concerning the MIHI EST construction, however, this distinction does not stand, since the MIHI EST structures with a deontic modal *fi* 'be' cannot be analyzed as raising structures, as Barbu (2015) argues for the structures containing a deontic modal *fi* 'be', in

---

[64] It is well known that in Romanian *fi* 'be' is categorized as a predicative, copular, auxiliary, or modal verb (Dragomirescu 2013: 197). As a modal verb, *fi* 'be' actualizes several meanings: obligation, possibility, or permission, but also volition, capability or non-volition (fear, concern).

[65] In her approach, Barbu (2017) proposes an analysis in terms of raising, not only for the modal *fi* 'be', but also for the predicative, copular, (semi)auxiliary *fi* 'be' (Barbu 2017: 560–561).

general. Therefore, I propose, with respect to the MIHI EST construction, to maintain Barbu's (2015) distinction between (i) structures with a deontic modal *fi* 'be', and (ii) structures with a dynamic modal *fi* 'be', but I argue that both types instantiate control structures.

In what follows, I will focus on these MIHI EST structures and I will first identify and illustrate those containing a deontic modal *fi*, then the ones containing a dynamic modal *fi* 'be'.

4.1.2.1 MIHI EST patterns with a deontic modal *fi* 'be'

Based on the available data, I show in this section that, when the deontic modal *fi* 'be' is part of the *verbal MIHI EST* construction (Bauer 2000: 180), the combination instantiates only control structures, a raising analysis being excluded. This goes against the general approach to the verb *fi* 'be', considered as a raising verb by several scholars such as Avram (1999), Cornilescu (2009), Ionescu (2013), and Barbu (2015). Taking a closer look at the values of the modal *fi* 'be', Barbu (2015) shows that an analysis of this verb exclusively in terms of raising is contradicted by the data. She argues that, in general, the structures with a deontic modal *fi* 'be' are raising structures. While this may be true for the impersonal structures, or for the ones in which the verb *fi* 'be' has a nominative subject (141), this is not the case for structures in which the verb *fi* 'be' occurs with a dative, as in the verbal MIHI EST structure in (142).

(141) *...sămânța   ce   este   a   se   semăna*
      seed.the   that   is   INF   se.PASS   sow
      '... the seed that must be sown.'   (19th century, ex. 11c from Barbu 2015: 322)

(142) *Caută   ce   -mi   fu   a   păți*
      Look   what   =me.DAT   was.3SG   INF   happen
      'Look what had to happen to me.'   (17th century, ex. 11b from Barbu 2015: 322)

Indeed, in the example in (142), the verb *fi* 'be', although used as a deontic modal, cannot be a raising verb because the dative is clearly part of the higher verb *fi* 'be', and not assigned by the lower verb *păți* 'happen', which is a nominative-selecting verb in Romanian. In contrast, in (141), the raised nominative subject *ce* 'what', a relative pronoun having *sămânța* 'seed' as an antecedent, clearly belongs to the lower verb, *semăna* 'sow', which assigns the nominative case. It is well known that, in raising contexts, the raised subject should maintain the same case, as assigned by the lower verb, to which it actually belongs.

Barbu (2015) herself points out that structures in which the verb *fi* 'be' occurs with a dative are mostly found in the old language, and are specific to religious texts. As for the present-day language, these structures tend to specialize in the experession of dynamic modality, occurring with the verb *fi* 'be' as a control verb. Nevertheless, dative structures with the deontic modal *fi* 'be', expressing an obligation, a physical or moral constraint, or

predestination ('be meant for'), which, according to Barbu's alaysis, contain the raising verb *fi* 'be', cannot be analyzed in terms of raising: neither the examples from older Romanian, nor the ones from the present-day language. Observe the following examples:

Physical or moral constraint / obligation

fi + INF

(143) *că*   *era*   *popă (...)*   *şi*   *-i*   *era*   *a*   *mearge*   *şi*   *el*
that   was   priest   and   =him.DAT   was   to   go   and   he
cu   *-mpăratul*
with   =emperor.the
'Because he was a priest (...), and he also had to go with the emperor'   (1682, Dosoftei, V. S)

fi + SUBJ

(144) *Pre*   *Dumnezău,*   *cui*   *are*   *hi*   *să*   *-l*   *slujească,*
ACC   God   whom.DAT   have   be.COND.3SG   SUBJ   =him.ACC   serve.3SG
ei   *-l*   *pierd*
they-   =him.ACC   lose
'They who have to serve God lose him.'   (1643 Varlaam, C.; example 12b, Barbu 2015: 322)[66]

Predestination ('be meant for')[67]

fi + INF

(145) *? Nu*   *i-*   *a*   *fost*   *tânărului*   *[a*   *-i*   *plăcea*
not   him.DAT=   has   been   young_man.the.DAT   INF   =him.DAT   please
de   *fata*   *aleasă*   *de*   *părinţii*   *lui]*
of   girl.the   chosen   by   parents.the   his. POSS
'It wasn't meant for the young man to like the girl chosen by his parents'

---

[66] This example is presented by Barbu (2015) to illustrate the raising construction conveying a necessity/ obligation meaning. Nevertheless, in my data, this example must have been collected from another version of the same text, since it occurs without the accusative clitic *-l* 'him.ACC', what reveals another possible meaning of the example, causing the relative pronoun *cui* 'whom' to refer to the object *pre Dumnezău*, 'ACC God.ACC' and not to the *pro* subject of the sentence, coreferential with *ei* 'they'. This new interpretation cannot be analyzed as a MIHI EST construction, not even as a dative + *fi*.3SG as analyzed in Barbu (2015: 322), since the subject of *fi* 'be' is not the dative, but a *pro*, co-referential with *ei* 'they'. Observe also the anacoluthic structure, in which two coreferential elements have different cases (*pre Dumnezău* – accusative, and *cui* – dative).

[67] This structure alternates in Romanian with another structure, seemingly more frequent, in which a past participle occupies the predicate position. This past participle expresses a way in which the specific event is predestined. The most common ones are *dat* 'given', *scris* 'written', and more rarely, *destinat* 'destined'.

(146) *? Nu   i-   a   fost   tânărului   [a   cânta*

      not   him.DAT=   has   been young_man.the.DAT   INF   sing

      împreună   cu   ea]

      together   with   her

      'It wasn't meant for the young man to sing with her'

(147) **Nu   a   fost   tânărul   [a   cânta împreună   cu   ea]*

      not   has   been young_man.the   INF   sing together   with her

fi + SUBJ

(148) *dacă **mi-**   e   **să**   moriu (...)*

      if   me.DAT   is   SUBJ   die.1SG

      'If it's meant for me to die ...'   (1760–1820, Budai-Deleanu)

(149) **dacă **eu**   e   **să**   mor (...)*

      if   I   is   SUBJ   die.1SG

(150) *dacă e   **să**   mor (...)*

      if   is   SUBJ   die.1SG

      'If it's meant for me to die ...'

The examples (143) and (144) show, indeed, that the deontic modal *fi* 'be' cannot be a raising verb, since the raised subject does not maintain its case. As for (145), this example seems to instantiate a raising structure, in which the dative subject of the lower verb, *plăcea* 'please' raises to the subject position of the verb *fi* 'be', while maintaining its dative case. Example (146), however, contradicts this analysis and reveals that the dative goes actually with the verb *fi* 'be', and does not belong to the lower verb. This is demonstrated by the fact that, when the lower verb is replaced with a nominative-taking verb, like *cânta* 'sing', the dative subject of the higher verb does not change into a nominative, since such a structure would be ungrammatical (147). These contexts in (143)-(147) can be analyzed as clear instances of control.

As for (148), just like the previous examples, it cannot be interpreted as a raising structure due to the impossibility of the raised subject to occur in the nominative (149). It is worth mentioning, however, that this example has an impersonal variant (150), which is perfectly acceptable and grammatical in Romanian. Such impersonal structures expressing deontic modality are analyzed by Barbu (2015) as raising structures. A more detailed analysis of these structures would be necessary, but it would go beyond the purpose of the present study.

As Barbu (2015) observes, the patterns with the dative experiencer are mostly found in the old language and are specific to religious texts. In present-day Romanian, such examples are more difficult to interpret, since there is no clear distinction between the moral constraint, the obligation, and the 'be meant for' interpretation of the modal *fi* 'be' (151).

(151) **Nu   ne   este   nouă   a   judeca   aceasta**
     not   us.DAT   is   us.DAT   INF   judge   this
     'We must not judge this/ It is not meant for us to judge this'   (impantokratoros.gr)

To my best knowledge, most of these structures are not found in languages other than Romanian and Russian (152). In Russian, this structure occurs only with the predestination meaning (Moore and Perlmutter 2000). As for the variant in which a past participle such as *dat* 'given', *scris* 'written', or *destinat* 'destined' occupies the predicate position, such structures also exist in French (153) and in Bulgarian (154).[68]

(152) *Borisu   ne   istratit'   tak   mnogo   deneg   na   sebja*   (Russian)
     Boris.DAT   not   spend.INF   so   much   money   on   self
     It's not (in the cards) for Boris to spend so much money on himself.
                                       (ex. 7, Moore & Perlmutter 2000: 377)

(153) *Il   ne   lui   a   pas   été   **donné**   de   voir ...*   (French)
     it   not   him.DAT   has   NEG   been   given   INF   see
     'He didn't have the opportunity to see ...'

(154) *Ne   mu   e   bilo   **pisano**   da   se   ojeni   za   Klara*   (Bulgarian)
     not   him.DAT is   been   written   SUBJ   REFL   marry   for   Klara
     'It wasn't written for him to marry Klara'

Based on the available data on Romanian, I showed in this section that the pattern DAT + *fi* + inf / subj, in which the verb *fi* 'be' is a deontic modal expressing an obligation, a physical or moral constraint, or a predestination ('be meant for'), instantiates a control rather than a raising structure. This analysis is built on the evidence that the dative in the higher clause goes with the verb *fi* 'be', and does not belong to the lower verb.

4.1.2.2   MIHI EST patterns with a dynamic modal *fi* 'be'

The patterns introduced in this section are instances of the MIHI EST structure in which the verb *fi* 'be' takes on a modal interpretation, and occurs with a finite or a non-finite clause. Depending on the clause, the structure has a different meaning: it can express willingness or the lack of it, or it may convey the meaning 'to care' or 'to be afraid'. It can even express an irreal situation or a dream.

    Barbu (2015: 324–325) argues for a control analysis of the modal *fi* 'be' in the patterns presented in this section. In her view, the control *fi* has two argument positions: one is occupied by the dative experiencer, which is the controller, while the other is realized by a small clause in contexts with obligatory control (155), or by a subordinate clause in structures with optional control (156).

---

[68] Cf. personal communication with Vassil Mostrov, native speaker of Bulgarian.

(155) *Mie*   *mi-*   *e*   *a*   *dansa*   (Obligatory control)
       me.DAT   me.DAT=   is   INF   dance
       'I feel like dancing'

(156) *I$_i$-*   *e*   *să*   *nu*   *poticnească*   ___$_{i/j}$ (Optional control)
       him.DAT=   is   SUBJ   not   stumble.3SG=PL   PRO
       'S/he is afraid that s/he stumbles'

She explains the control analysis by the fact that, in these structures, two different semantic roles are assigned by two different verbs: the *experiencer*, assigned by *fi* 'be', and the *agent*, assigned by the verb in the small clause. This clearly contradicts the raising analysis proposed in the literature for the same structures by for instance Cornilescu (2009), since a raising verb cannot assign semantic roles.

These structures have in common the fact that they all combine a dative experiencer, the verb *fi* 'be' and a clausal complement, which attributes a *modal attitude* to the structure as a whole (concern, feeling like, feeling as if), (cf. also Cornilescu 2009: 212).[69] The first two attitudes coincide with the ones proposed by Barbu (2015), namely willingness and concern (157)–(160). To her list, I add a third pattern which expresses a (mostly) neutral attitude toward an irreal situation, imagined or dreamed by the experiencer (161a–b).

Willingness, or the lack of it

fi + INF

(157) *mie*   *mi-*   *e*   *a*   *dansa*
       me.DAT   me.DAT=   is   INF   dance
       'I feel like dancing'

fi + de + SUP

(158) *Dacă*   *nu*   *ţi-$_i$*   *a*   *fost*   *de*   *cumpărat*   ___$_{i/*j,...}$
       if   not   you.DAT=   has   been   SUP   bought   PRO
       'If you didn't want to buy,...'   (1892 Creangă: *Amintiri din copilărie*)

Concern or care

fi + neg + subj

(159) a. *I$_i$-*   *e*   *să*   *nu*   *poticnească*   ___$_{i/j}$
       him.DAT=   is   SUBJ   not   stumble.3SG=PL   PRO
       'S/he is afraid that s/he stumbles'

---

[69] For this reason, Cornilescu (2009: 212) considers that, in the structures expressing a modal attitude, the verb *fi* 'be' is treated as a *propositional attitude* verb. She uses the term *propositional* for structures in which the dative experiencer is envisaging the *proposition* of a specific event as a *possibility* and is attaching an emotion to it.

b. *Mi-*   *e*   *să*   *nu*   *se*   *supere*
  me.DAT=   is   SUBJ   not   REFL   upset.SUBJ.3SG
  'I'm afraid that he will mind'

fi + *că*-clause

(160) a. *I̵-*   *e*   *că*   ___ᵢ/ⱼ   *se*   *arde*
  him.DAT=   is   that   PRO   REFL   burns
  'S/he is afraid of burning her/himself'
 b. *mi -*   *e*   *că*   *te*   *-i*   *supăra*
  me.DAT=   is   that   you.ACC   FUT   upset.INF
  'I'm afraid that you will get upset.'   (ex. 17b, Barbu 2015: 325)

Feeling as if (irrealis)

fi + *ca și când*-CLAUSE/ *parcă*-CLAUSE

(161) a. *Îiᵢ*   *era*   *ca_și_când*   ___ᵢ/ⱼ *s-*   *ar_fi*   *jucat*   *cu*   *focul*
  her.DAT   was   as_if   PRO   REFL   be.COND.3SG   played   with   fire.the
  și   ___ᵢ/ⱼ   *s-*   *ar_fi*   *ars*
  and   PRO   REFL   be.COND.3SG   burned
  'She felt as if s/he would have played with fire and s/he got burned'
 b. *Îi*   *era*   *parcă*   *fiica*   *ei*   *suferă*
  her.DAT   was   as_if   daughter.the   her   suffers
  de-   *o*   *grea*   *boală*
  of   a   heavy   illness
  'She felt as if her daughter would suffer from a serious illness'

As opposed to the examples in (157)–(158), where *fi* 'be' conveys willingness or the lack of it, and displays obligatory control, in the structures in (159)– (160) the dative experiencer optionally controls PRO. The same holds for the examples in (161a-b), which express an irreal situation introduced either by *parcă*, or by *ca și când*, both meaning 'as if'. This situation is always experienced like in a dream by the experiencer, but PRO is coreferential either with the experiencer himself, or with another person. Indeed, on a raising analysis, these structures with optional control in (159)-(161) would remain inexplicable.

As for the pattern expressing a skill, or a personal ability, identified by Barbu (2015), it contains, indeed, a control *fi* 'be'. These structures differ from the previous ones in that the verb *fi* 'be' occurs with an adjective or with an adverb. Observe, however, that, in these structures, the finite or non-finite clause does not occupy a predicate position itself, but is a complement of the construction (162a-b). Therefore, the meaning of the whole construction is conveyed this time by the adjective or the adverb, and not by the clausal complement. This makes the construction more comparable with the MIHI EST construction with a noun, and the similarities and differences between them constitute an extraordinary topic for further research.

(162) a. *Mi-*   *e*   **uşor**   *să*   *cânt*
     me.DAT=   is   eas(il)y   SUBJ   sing.1SG
     'I can sing easily.'   (ex. 18a, Barbu 2015: 326)

    b. *Lui*   *Ion*   *îi*   *era*   **greu**   *să*   *plece*
     DAT   John   him.DAT   was   hard(ly)   SUBJ   leave.3SG
     'John could hardly/wouldn't leave.'   (ex. 18b, Barbu 2015: 326)

It has to be noted that the two values of the modal *fi* 'be' distinguished above are ambiguous. Examples like (163a-b) can instantiate either deontic (163a) or dynamic modality (163b), as observed by Barbu (2015). She considers this fact as an additional argument against the analysis of the verb *fi* 'be' exclusively as a raising verb in these structures, as it was claimed in the literature. In the examples below, I preserved her notation in order to differentiate between the two identical examples. Recall that, against the clear evidence shown by the data, she considers deontic modal *fi* 'be' as a raising verb, hence her specific notation, whereas the dynamic modal *fi* 'be' is considered as a control verb.

(163) a. ***Nimănui*ᵢ**   *nu*   *-i*   *este*   [SC *e*ᵢ.AGENT   ***a***   *-l*   **săruta** *pe*   *Domnul]*
     nobody.DAT   not   =him.DAT   is   PRO   INF   =him.ACC   kiss   ACC   God
     'Nobody is meant to kiss God.'   (ex. 19a, Barbu 2015: 327)

    b. ***Nimănui*ᵢ,EXPERIENCER**   *nu*   *-i*   *este*   [SC *e*ᵢ.AGENT   ***a***   *-l*   **săruta**
     nobody.DAT   not   =him.DAT   is   PRO   INF   =him.ACC   kiss
     pe   *Domnul]*
     ACC   God
     'Nobody feels like kissing God.'   (ex. 19b, Barbu 2015: 327)

4.1.2.3   The inchoative variant of MIHI EST

Let us now turn to the inchoative variant of the MIHI EST construction, the *VENIO inchoative* construction. As already mentioned at the beginning of this section, this inchoative construction contains the movement verb *veni* 'come' (164a-b), or the verb *se face* 'become' (165a-b). In combination with a state and an experiencer, these verbs assign an inchoative meaning to the construction as a whole. Both structures may occur with a stimulus realized either by a noun or by a finite or a non-finite clause.

(164) a. ***Îmi***   *vine*   *dor*   *de*   *tine*
     me.DAT   comes   longing   of   you.ACC
     'I begin to miss you'

    b. ***Îmi***   *vine*   *dor*   *să*   *plec*   *la*   *munte*
     me.DAT   comes   longing   SUBJ   depart.1SG   at   mountain
     'I begin to miss going to the mountains'

(165) a. ***Mi***   *se*   *face*   *dor*   *de*   *tine*
     me.DAT   SE   makes   longing   of   you.ACC
     'I begin to miss you'

b. *Mi*	*se*	*face*	*dor*		*să*	*plec*	*la*	*munte*
me.DAT	SE	makes	longing		SUBJ	depart.1SG	at	mountain

'I begin to miss going to the mountains'

It is worth mentioning that *veni* 'come', but not *se face* 'become', occurs also with a modal value (166) and (167), when it is not followed by a state noun. Just like in the case of the verbal MIHI EST construction, *veni* 'come' co-occurs in this structure with an experiencer in the dative and a finite (166a) or a non-finite clause (166b).

(166)   a. *Îmi*   *vine*   *să*   *plec*   *la*   *munte*
        me.DAT   comes   SUBJ   depart.1SG   at   mountain
        'I feel like going to the mountains'

        b. *Îmi*   *vine*   *a*   *plânge*
        me.DAT   comes   INF   cry
        'I must cry'

(167)   *Mi   *se*   *face*   *să*   *plec*   *la*   *munte*
        me.DAT   SE   makes   SUBJ   depart.1SG   at   mountain

The modal *veni* 'come' is similar to the modal *fi* 'be' in that it has both the deontic and the dynamic values, expressing either uncontrollable necessity (cf. 'I must' in 166b), or willingness (cf. 'I feel like' in 166a). Note that both instantiate control structures, and, although in both structures the control is obligatory, a raising analysis is excluded.

4.1.2.4   Origin of the bare state noun type constructions

It is well known in the literature that the MIHI EST construction traces back to the Latin MIHI EST pattern, although more recent studies point toward much deeper roots, going back to Proto-Indo-European (cf. Section 4.2, *infra*). Interestingly, Benedetti (2013b: 134) establishes a connection between the Latin MIHI EST and the Latin structure using the verb *capere* 'take, seize', in terms of non-inchoative vs. inchoative construction. However, as shown in Section 4.1.1, the CAPIO *inchoative* construction belongs to a different type in Romanian, namely the determined state noun type. As opposed to the structure with the verb *capere*, the MIHI EST construction did not survive in other Romance languages, except in Romanian.

As for the VENIO *inchoative* construction, there is, to my best knowledge, no clear indication in the literature about its origin. Nevertheless, the fact that it also exists in other Romance languages, as well as in Icelandic and in South Asian languages such as Nepali,[70] points toward an Indo-European origin.

---

[70] Cf. Barðdal & Eythórsson (2018: 2).

In Romance, the existence of the *VENIO inchoative* construction is confirmed in the literature for Italian (168) (cf. Bentley 2006: 118-119, and Fedriani 2011: 321). For Spanish[71] and Portuguese, I have had to rely on the World Wide Web in order to document examples of the *VENIO inchoative* construction. For Spanish I have found on the web 73 occurrences (169), while Portuguese examples are rather scarce: for European Portuguese I found six occurrences, all with a determined state noun (170a), for Brasilian Portuguese eight occurrences with a determined state noun and three occurrences with a bare state noun (170b). It has to be noted that in French, the use of the construction is more restricted, in that it always co-occurs with one and the same state noun *envie* 'mood, will, desire' (171).

(168) *mi è venuta fame* (Italian)
me.DAT is come hunger
'I begin feeling hungry'

(169) *el hambre me ha venido de_golpe* (Spanish)
the hunger me.DAT=ACC has come suddenly
'Suddenly, I begin feeling hungry'

(170) a. *veio -me a fome* (European Portuguese)
came.3SG =me.DAT the hunger
'I begin feeling hungry'

    b. *me veio (a) fome* (Brazilian Portuguese)
me.DAT came.3SG the hunger
'I begin feeling hungry'

(171) *Il me vient/ prend l' envie de ...* (French)
it me.DAT=ACC come/ take the= mood/desire to ...
'I begin feeling like ....'

Moreover, the French construction shows many similarities with the *CAPIO inchoative* construction, since the state noun is necessarily determined. In addition, the verb *venir* 'come' freely alternates with the verb *prendre* 'take', and the clitic form is ambiguous between the accusative and the dative, due to syncretism. In the light of the discussion above, I suggest that the French venir *'come'* in this structure has only its dynamic modal value, imposing obligatory control and expressing willingness (171), as opposed to Romanian, where *veni* 'come' expresses both deontic as well as dynamic modality.

---

[71] The present example is from Jose Antonio Molinero Reina, *The War of the Saurians:* Libro Uno*: La Primera Batalla* (2014).

4.2   The Latin MIHI EST construction

This section focuses on the path the MIHI EST construction has followed from the Latin MIHI EST pattern to the Romanian experiencer MIHI EST construction. I will first present the traditional approach, which considers the Latin predicative possession construction as the ancestor of the Romanian MIHI EST construction. This view implies that the predicative possession construction has undergone a development from concrete possession to abstract possession, which triggered the experiencer reading of the dative, instead of the possessive reading, leading to the new experiencer construction (Benveniste 1966; Bauer 1996; Bolkestein 1983, 2001; Bauer 2000: 174, 193). Subsequently, I will introduce a more recent view on the origin and the development of the MIHI EST construction proposed by Barðdal et al. (2012), Barðdal & Smitherman (2013), Danesi, Johnson & Barðdal (2017), and Danesi & Barðdal (2018). This new approach demonstrates that the experiencer meaning of this construction was available already in Early Latin, and confirms that the origin of the dative experiencer construction goes far beyond Latin.

The traditional approach on the MIHI EST pattern claims that this pattern traces back to Proto-Indo-European, where it was the canonical construction for predicative possession (cf. Benveniste 1966, Bauer 1996). However, in most European languages, this structure has been replaced by a HABEO construction, described by Benveniste (1966: 197) as an "*être-à inversé*" ('an inversed *be of*'). This analysis of 'having' as a reversed 'being of' has been taken over by many linguists, especially in Generative Grammar, which views the relation between MIHI EST and HABEO as a case of "predicate inversion" (cf. among others, Kayne 1993 and Den Dikken 1998). Both 'have' and 'be' are analyzed as copulas in that they are unable to assign semantic roles. These are assigned by the copula's complement (i.e. the direct object of 'have', and the predicate noun of 'be'). However, the two constructions differ in their encoding of the subject: in the MIHI EST construction, the possessee is in the nominative controlling verb agreement, and the possessor is in the dative, whereas with HABEO the possessor is in the nominative (172).

	Subject	Predicate	
(172)	*Liber est*	*puero*	'The book belongs to the child'
	Puer habet	*librum*	'The child has a book'

In most SAE languages, the MIHI EST construction has given way to the HABEO construction. This is due to the fact that, being transitive, the HABEO construction corresponds better to the European type, which tends to transitivity (cf. Bauer 1996, Haspelmath 2001a, Seržant 2013). Moreover, the HABEO construction is more canonical because the nominative NP tends to be definite and human. This is not the case in the MIHI EST

construction, in which the nominative, which denotes the possessee, is indefinite and mostly inanimate, whereas the dative argument, which denotes the possessor, is definite and human. According to Benveniste (1966), the indefiniteness of the nominative possessee in the MIHI EST construction is a crucial property, which differentiates it from another possessive construction, i.e. the belonging construction, in which the possessor is in the genitive (173c). In the belonging construction, the nominative possessee is generally definite and is, therefore, more appropriate for a subject than the indefinite possessee of the MIHI EST construction. Benveniste argues that, with respect to the referential value of the possessee, the MIHI EST construction corresponds in fact to the HABEO construction with a nominative possessor, rather than to the belonging construction with a genitive possessor.

(173)	a. *Puero*	*est*	*liber*		POSSESSION
	child.DAT	is	book.NOM	'The child has a book'	
=	b. *Puer*	*habet*	*librum*		
	child.NOM	has	book.ACC	'The child has a book'	
≠	c. *Liber*	*est*	*pueri*		BELONGING
	book.NOM	is	child.GEN	'The book belongs to the child'	

However, according to Bolkestein (1983, 2001) the difference between the constructions in (173a) and (173c) is related neither to the definiteness of the possessee, which is not marked in Latin, nor to its topicality, since the order of the constituents can vary in both structures (cf. also Bauer 1996). Bolkestein (1983, 2001) argues that the essential difference between the two case markings of the possessor is that the genitive possessor functions as a predicate, whereas the dative possessor is an argument. The predicative character of the genitive possessor is due to the adnominal character of the genitive, which typically marks modifiers. In contrast, the dative is generally assigned to verbal complements or adjuncts, and never to adnominal complements. It is therefore not suitable for predicate marking, but rather for argument marking. This different functioning of the possessor gives rise to semantic differences between the two constructions: the genitive structure expresses a permanent possession characterizing the possessee, whereas the dative structure expresses a temporary or contingent possession (Bolkestein 1983: 13).

Due to its similarity with the HABEO pattern, the MIHI EST pattern quickly receded for the HABEO structure in classical Latin. Bolkestein (1983, 2001) and Bauer (1996: 244-245, 2000: 174, 193) show that MIHI EST is used as a possessive construction with concrete nouns only in pre-classical Latin, especially in Plautus. These scholars note also that, in classical Latin it mostly occurs with inalienable relationships, which leads Bolkestein (2001) to the conclusion that the dative is an experiencer rather than a possessor.

While acknowledging the formal and functional similarities between the possessive and the experiencer constructions, Danesi & Barðdal (2018) challenge the hypothesis presented above.

Based on the Vedic data, Danesi & Barðdal (2018: 23) consider the dative construction as a multifunctional expression, whose meaning covers two core semantic fields namely experience and happenstance, and three less core semantic areas, i.e. evidentiality, modality, and possession. Their findings suggest that in the Indo-European daughter languages the dative construction is not restricted to expressing possession but is a subconstruction of a family of constructions, in which the subject-like argument is in the dative, namely the oblique subject construction (Danesi & Barðdal 2018: 23). This fact is supported by several examples from Plautus and Terence, as the one in (174), in which MIHI EST occurs already in Old Latin with an abstract noun, i.e. *pudor* 'shame' (cf. also Bennett 1914: 164, cited by Bauer 2000: 181), and not much later, in classical Latin, as stated by Bolkestein (1983, 2001) and Bauer 1996: 244-245).

(174) *Credam,*   *pudor*   *si*   *quoiquam*   *lenoni*   *siet*
      believe.FUT.1SG   shame   if   any   pimp.DAT   be.SUBJ.3SG
      'I will believe it, if any pimp is ashamed'   (Plautus, *Curculio* 58)

Furthermore, Barðdal et al. (2012), Barðdal & Smitherman (2013), Danesi, Johnson & Barðdal (2017), and Danesi & Barðdal (2018), show that the oblique subject construction found in the Indo-European daughter languages (i.e. Latin, Greek, Indic, Germanic, Slavic, etc) can be reconstructed for Proto-Indo-European. Their studies show that the dative in the oblique subject construction, to which the Latin MIHI EST construction belongs, has always had an experiencer role, in addition to other roles such as possessor, protagonist of a happenstance or a modality event. They argue that it must have shown subject properties already in Proto-Indo-European.

Indeed, without mentioning the dative subject hypothesis, Bolkestein (1983) herself points out several examples in which the dative possessor takes on subject properties, such as (175), an example from Plautus, where the dative pronoun *ei* is the antecedent of the reflexive *sui*. As is well known, the possibility of controlling reflexive pronouns is one of the basic criteria of subject and is central to the dative subject hypothesis (see Haspelmath 2001a; Barðdal 2002; Barðdal & Eythórsson 2003, among many others).

(175) *Erat*   *ei*   *hospes*   *par*   *sui*
      was   him.DAT   friend   similar   his.REFL.GEN
      'He had a friend similar to him'   (Plaute, Rud. 49)

Along the same lines, Pooth et al. (2018) argue that the use of non-canonical subjects in the Indo-European daughter languages is a remnant and an analogical extension of the semantic marking of the experiencer and of the undergoer in Early Proto-Indo-European. Despite the change in the alignment system – from a semantic to a transitive system –,

traces of the Early Proto-Indo-European experiencer construction survived and have evolved in the daughter languages, according to the particular situation of each language.

Baldi & Nuti (2010: 260–61, cited by Fedriani 2011: 312) show that this is also the case for the MIHI EST construction. In their study, they observe that in Old Latin the MIHI EST construction with an experiencer interpretation was more frequent than the experiencer HABEO construction, with a ratio of 35 (MIHI EST) to seven (HABEO) occurrences. More precisely, they show that in the texts of an early author like Plautus, MIHI EST was the dominant construction for expressing states, whereas HABEO was used only occasionally with the same abstract nouns. This is in line with the hypothesis presented above, and confirms that the Latin MIHI EST construction is one of the instances of the Proto-Indo-European dative subject construction. In light of these details, I hypothesize that the Pre-Latin experiencer MIHI EST construction receded in classical Latin, in favor of the HABEO construction, but became again frequent in later stages, giving rise to the productive experiencer MIHI EST construction in Romanian, in which the dative experiencer is the topical argument and the state noun is the predicate. This new construction has become the most natural way to express psychological or physiological states in Romanian.

4.3   The Romanian MIHI EST construction

The existence and the seeming expansion of the MIHI EST experiencer construction, among other experiencer constructions in Romanian, has led Haspelmath (2001a), based on the study by Bossong (1998), to classify Romanian among SAE-peripheral language groups such as East-Slavic and Baltic. However, the MIHI EST structure has never been studied in detail, neither for Contemporary nor for Old Romanian. As a result, several questions related to the origins and evolution of this structure in Romanian, are not dealt with in the literature, a gap that I aim to fill through the present research.

Hence, in the remainder of this dissertation, the following research questions will be addressed:

(i)   How does the set of state nouns occurring in the MIHI EST construction evolve throughout the centuries?

(ii)   How does the competition between the MIHI EST and the HABEO construction evolve in Romanian during the course of the history?

(iii)   To which extent is the dative experiencer in the MIHI EST construction a genuine syntactic subject?

(iv)   Is the MIHI EST construction expanding or regressing in Romanian?

Finding an answer to these questions is crucial for the study of Romanian, but also for the study of the Romance languages in general, these outcomes being of a considerable importance for typology, as well as for comparative linguistics. Therefore, in what follows I will narrow down my attention on the MIHI EST construction in Romanian, observing its behavior and its evolution from the first attested texts until today.

Chapter 5   Corpus and methodology

5.1   Choice of the corpus

The choice of a corpus can be very challenging, since several factors have to be considered. In the case of Romanian, the privilege of choosing did not really exist, since, at the time I started my study, in 2016, there was, to my best knowledge, no 'official' corpus freely available for scholars neither for present-day Romanian,[72] nor for modern or old Romanian.[73] In my study, for present-day Romanian I have used the web corpus Romanian Web 2016 (*roTenTen16*), which contains over two billion words and is provided by Sketch Engine, an electronic platform that collects web corpora for several languages. By its accessibility and its user-friendliness, the Sketch Engine platform has proven to be a valuable tool for the present study, providing me with a comprehensive corpus for present-day Romanian. As for pre-21st century Romanian, I compiled my own corpus. Fortunately, the Sketch Engine platform made available the necessary tools and storage place, facilitating in this way the creation of such a corpus.

---

[72] A notable project is the one titled CoRoLa (Cristea et al. 2019), a project of creating a representative corpus of Contemporary and Present-Day Romanian, containing written and oral materials from 1945 until now. This project started in 2012 and the corpus was launched in December 2017. CoRoLa – *Corpus de referință pentru limba română contemporană* 'CoRoLa – Reference corpus for contemporary Romanian', is the final product of a project initiated by the Romanian Academy through the Research Institute for Artificial Intelligence "Mihai Drăgănescu" (București), in collaboration with the Institute of Theoretical Informatics (Iași). URL: http://89.38.230.23/corola_sound_search/

[73] The CETRV - *Corpus electronic al textelor românești vechi* (1521 – 1640) 'The electronic corpus of old Romanian texts' – has been created within a project conducted by Al. Gafton and his team from the University Al. I. Cuza (Iași). The aim of the project is to compile an electronic database of Old Romanian texts. For studies which target the oldest period of Romanian only, this is a valuable tool. However, this corpus is too limited for my research because it does not go beyond the 17th century. Another problem is that it does not allow searching for specific patterns, but only for specific lemmas.

5.2 Why a web corpus?

One may fairly ask why I prefer to use a web corpus for my research. As mentioned in the previous section, when this study started there were no official corpora available to researchers working on Romanian, with the only exception of the Europarl 7, a parallel corpus created from the European Parliament Proceedings, containing texts from the period 2007–2011. However, this resource contains mostly legal documents and is therefore not suitable and it is too limited with respect to the topic of the present study, which concerns verbs and predicates expressing psychological and physiological states.

The choice for a web corpus proved to be a very good resource for a study on experiencer constructions. Today it is on the web, on blogs and forums, and on social media, that one expresses most naturally one's feelings. Additionally, the selected corpus contains a large number of examples from news platforms, and even literary texts, which creates diversity in genre and register. Moreover, I considered it vital to focus on texts as recent as possible in order to obtain a proper description of current tendencies.

Needless to say, a web corpus contains a large amount of noise which, inevitably, is exported together with the relevant examples and needs to be cleaned manually from the dataset. Just to mention a few sources of noise, these are recurrent texts on different websites, dictionary websites, translation websites, etc.

In spite of the noise, the web corpus provides an abundance of relevant instances of the construction under scrutiny. The examples in my dataset for present-day Romanian represent three quarters of the total relevant instances, the other quarter being examples from pre-21st century Romanian.

Furthermore, earlier studies on corpora have shown that the World Wide Web can be considered as a representative language corpus, even though it is both unbalanced and uncontrolled for. Keller & Lapata (2003) suggest that the enormous size of the Web compensates for these limitations.

5.3 Creation of a corpus for the pre-21st century Romanian

The beginning of writing in Romanian is characterized by a mix of influences. (Stan 2013: 21) points out that during the period between the 16th and the 20th centuries, spoken Romanian comes in contact with Slavon, Latin, Hungarian, German, Neo-Greek, Italian, and Turkish. Some of the syntactic patterns used in these languages have entered the language through translations. These translated texts are for the most part religious texts (The Bible 1688), legends (*Legenda duminicii* 1601), historical writings (the

chronograph translated by monk Mihail Moxa at Cozia Monastery in 1620), juridical writings (*Pravila lui Coresi* 1560–1562), and moralizing novels (Alexander Romance 1620).

My corpus contains both translated and original texts.[74] It goes without saying that original texts are the most interesting ones, although the number of original Romanian texts representative for the period between the 16th and 20th centuries, and especially for the 16th century, is very limited. Among these, I can mention short writings such as Princely or Episcopal Chancery documents, bills of sale, letters, notes, prefaces, epilogues, or longer writings, such as historical texts (Gr. Ureche ~1725, M. Costin 1700–1750, Constantin Cantacuzino 1700–1750), religious texts (Antim Ivireanul 1709; 1722–1725), and literary texts (Cantemir ~1705).

The compilation of the corpus for old Romanian was a challenging but very enriching experience, since not all the existing texts have been already digitized and very few are available for researchers.[75] Therefore, building this corpus may not have been possible without the kind cooperation of other Romanian scholars, who provided me with a part of the documents in digital format.[76] Digital libraries with open access represented another important source of old documents in digital format.[77] Hence, all available old texts and a selection of modern texts have been collected either in pdf or in txt format. The pdf files have been converted into editable pdf files, then into txt files, with the help of optical character recognition (OCR) in Acrobat. At this point I made use of the tools and the storage place offered by Sketch Engine,[78] which allowed me to upload the txt files and to tag and syntactically annotate (parse) my own Romanian corpus for the period between the 16th and 20th centuries. This corpus contains in total almost eight million words.

---

[74] As for the versions of the texts for old Romanian, I selected, where this was possible, the ones that have been used in the Oxford version of *The syntax of old Romanian* (2016).

[75] Several projects of digitizing Old Romanian texts have been pursued: Eugen Munteanu, *Monumenta Linguae Dacoromanorum*, 2006–2007, which focuses on the *Biblia de la Bucureşti* 1688; Cristea et al. (2007), *Thesaurus Dictionary of the Romanian Language*, 2003–2007, etc.

[76] I am profoundly grateful to prof. Dana Niculescu from the University of Amsterdam and to prof. Camelia Stan, from the University of Bucharest, who provided me with a part of the texts for my corpus.

[77] To mention just a few names: *Medievalia* (https://medievalia.com.ro/), Biblioteca Digitală a Bucureştilor (*Dacoromanica*) (http://www.digibuc.ro/), Biblioteca Digitală BCU Cluj (http://dspace.bcucluj.ro/), Biblioteca Digitală BCU Iaşi (http://dspace.bcu-iasi.ro/), and the CETRV (http://www.textvechi.ro/acasa).

[78] The storage size for own corpora was, at that time (2016-2017), limited to 10 million words.

5.4   Periodisation of Romanian

In comparison with the Western Romance languages, which show a certain continuity of texts during the transition from Latin to a new daughter-language, this transition period is very difficult to reconstruct for Romanian due to the lack of preserved written texts. Moreover, written sources continue to be scarce even after the attestation of the first Romanian written text. Hence, instead of well-defined chronological boundaries, the researchers propose relative chronologies, as mentioned by Ionescu-Ruxăndoiu 2010: 195; cited in Pană Dindelegan 2013b: 4).

The periodization proposed by Gheție (1997: 52–53) remains one of the most comprehensive relative chronologies, adopted also in *The Grammar of Romanian* (2013) and in *The Syntax of Old Romanian* (2016), both coordinated by Pană Dindelegan (2013c, 2016). Gheție (1997) distinguishes six different stages of Romanian:

(i)   Early Old Romanian, starting with the first original Romanian written text, in 1521, until the moment when the first collection of legal, canonical and civil laws in Romanian was printed at the Govora Monastery, in 1640;[79]

(ii)   Late Old Romanian, from 1640 until 1780, when the first important grammar of Romanian, *Elementa linguae daco-romanae sive valahicae*, was printed;

(iii)   Pre-Modern Romanian is a very short period, between 1780 and 1830; this period is representative because it brings the beginning of a conscious period of 'Romanization', or of 'Re-Romanization' of Romanian (Pană Dindelegan 2013b: 5);

(iv)   Modern Romanian covers the period between 1830 and 1899, and is characterized by stylistic diversification and an increasing production of genuine literary works;

(v)   The entire 20th century is classified as Contemporary Romanian;

(vi)   Present-Day Romanian includes all Romanian productions after 1989, until today.

For distinguishing the different stages of Romanian in the diachronic study, I first wanted to adopt this periodization. However, my dataset does not contain a balanced number of examples representing all the above-mentioned periods. In the whole corpus, I found very few examples of the earliest period for the simple reason that there are few texts available. I found very few examples too for what Gheție calls Pre-Modern Romanian, this time because it is a very short period.

Therefore, for the purpose of the present study, I distinguish only between four periods. Early and late old Romanian are merged into one period, which I call old Romanian, and pre-modern and modern Romanian are merged into what I call modern Romanian. Hence, the four periods of Romanian I work with here are the following:

---

[79] The first original Romanian written text is a letter, *Scrisoarea boierului Neacșu din Câmpulung* 'The letter of Neacșu from Câmpulung', dating from 1521.

(i)   Old Romanian, corresponding roughly to the 16th–18th centuries (ORom);
(ii)   Modern Romanian, covering the 19th century (MRom);
(iii)   Contemporary Romanian, corresponding to the 20th century (CRom);
(iv)   Present-day Romanian: the 21st century (PDRom).

5.5   Compilation of the dataset

The dataset I used for my study is preserved in an excel document that contains a certain number of examples extracted from the above-mentioned corpora, and manually annotated according to a selected number of variables. This dataset has been used for the computational analysis of the data and for the visualization of the outcomes. There are two stages in the creation of the dataset for the present corpus study.

5.5.1   Phase I

The first stage in my research has as objective to establish the list of nouns occurring in the MIHI EST construction. Therefore, I searched for examples with a dative pronoun, the verb *fi* 'be' and a noun, both in present-day Romanian and in the pre-21st century Romanian. In the roTenTen16 corpus, which counts approximately two billion words, as well as in the self-made corpus for the pre-21st century Romanian, a specific query was run, targeting combinations such as [DAT *fi* N] (cf. *mi-e lene* 'I feel lazy'), where N stands for any noun. The corpus for present-day language returned 154 492 sentences containing the targeted structure.

From these examples, I extracted a random sample of 100 000 instances, based on which I have generated a frequency list. As for the pre-21st century Romanian, the query returned 2 277 instances, which have all been preserved and based on which a frequency list was generated. For instance, for the targeted combination [DAT *fi* N] (cf. *mi-e lene* 'I feel lazy'), the frequency list normally provides structures such as *li-e lene* 'they feel lazy', *v-a fost lene* 'you were lazy', *îţi va fi lene* 'you will feel lazy'. In order to reduce the size of the frequency list and the amount of work needed to process it, the frequency list has been generated based on lemmas. To get a glimpse of how the frequency list looks like, Figure 5.1 displays the top rows of the frequency list for the 21st century. A great amount of unwanted hits was filtered out manually, from both sets of structures. Eventually, from the extracted samples, a number of 95 nouns were identified for present-day Romanian and 29 nouns for pre-21st century Romanian, the most frequent among them expressing

physiological states, such as hunger, thirst, or psychological states, such as fear, shame, etc.

		Lemma	↓ Frequency	Per million tokens		
1	☐	eu fi frică	9,523	3.03		⋯
2	☐	eu fi dor	9,119	2.90		⋯
3	☐	eu fi teamă	8,898	2.83		⋯
4	☐	el fi frică	5,147	1.64		⋯
5	☐	eu fi rusine	2,715	0.86		⋯
6	☐	el fi teamă	2,527	0.80		⋯
7	☐	eu fi milă	2,186	0.70		⋯
8	☐	el fi rusine	2,145	0.68		⋯
9	☐	eu fi rușine	1,761	0.56		⋯
10	☐	noi fi frică	1,543	0.49		⋯
11	☐	tu fi frică	1,541	0.49		⋯
12	☐	el fi loc	1,539	0.49		⋯
13	☐	eu fi silă	1,321	0.42		⋯
14	☐	eu fi foame	1,288	0.41		⋯
15	☐	el fi rușine	1,200	0.38		⋯
16	☐	eu fi si	1,187	0.38		⋯
17	☐	tu fi teamă	1,146	0.36		⋯
18	☐	eu fi lene	1,088	0.35		⋯

Figure 5.1   Snapshot of the frequency list for the 21ˢᵗ century (Sketch Engine)

5.5.2   Phase II

The second phase of the corpus study aims to search for all structures that can host the gathered nouns, besides the MIHI EST construction, both in present-day Romanian, and in the pre-21ˢᵗ century Romanian. Structures such as the HABEO construction, and other verb constructions with a nominative, an accusative or a dative experiencer were found. In order to get the maximum number of each of the selected structures for each noun, and to minimize the amount of noise, very specific queries were created for each of the nouns from the list collected during the first phase of the study. An example of a simplified query is given in (176), below:

(176) [DAT v N]|[NOM v N]|[ACC v N]

where 'v' can be any verb, including *fi* 'be' or *avea* 'have', and N is each time replaced with one of the nouns from the list gathered in the first study.

Several peculiarities of the present-day language were considered, such as irregularities in the usage of diacritics (cf. *frică* vs. *frica*, *scârbă* vs. *scarbă/ scârba/ scarba/ scîrbă/ scîrba/ scirbă/ scirba,* etc), or spelling modifications for pragmatic reasons (cf. *frig* vs. *friiiiig/ frică/ fricăăăă/ fffrică*, etc). As for the particularities of old Romanian, special attention was given to archaic forms (cf. *hi* in place of *fi* 'be', *pohtă* instead of *poftă* 'craving', *seate* for *sete* 'thirst', etc.) or to inconsistencies in the already mentioned use of diacritics (cf. *scârbă* vs. *scarbă/ scârba/ scarba/ scîrbă/ scîrba/ scirbă/ scirba,* etc). Note that when working with old texts, other orthographic peculiarities may occur which cannot always be predicted (cf. for instance, words broken by brackets). These issues can explain the greater amount of noise that was extracted from the corpus for the pre-21st century Romanian.

After considering all the predictable situations, I ran the query for each noun. From the total hits per noun, a sample of 200 sentences was taken, my goal being to collect approximately 100 relevant examples per noun, in order to generate the final dataset. It has to be noted that this was not possible for all the nouns and for all periods, due to size-related limitations.

By means of these queries, I was able to retrieve 16 550 examples for all periods of Romanian, including the noise. These examples were extracted and saved in an Excel document. After manually filtering the noise, my sample counts 8 458 examples to be analyzed, among which 4 828 examples (57 %) with a dative experiencer, 3 412 (43 %) with other kinds of experiencers, and 2 784 examples of the MIHI EST type only. It has to be noted also that the relevant examples from present-day Romanian are more numerous than the ones from the pre-21st century Romanian. More precisely, 74 % (6 248) of the examples represent present-day Romanian, and 26 % (2 209) the pre-21st century Romanian.

5.6   Annotation of the data

The next step was to annotate the examples with respect to a series of variables, i.e. factors that carry specific labels, which can be counted or measured, and which can take on potentially different values. For instance, the variable Case (of the experiencer) can take the following values: NOM, ACC, DAT, or GEN. In what follows, the most important variables are briefly described.

5.6.1   Translated vs. Original

The distinction between original and translated sources is relevant especially for the examples from the pre-21st century Romanian. As for present-day Romanian, since I work with a web corpus, it is more difficult to distinguish translated from original texts, unless this is clearly specified. Therefore, observations and judgements related to this parameter are made only for the pre-21st century Romanian, based on the information related to the dating and the origin of most of the old texts, provided in Timotin (2016: 1–7).[80]

5.6.2   Period

The entire timespan, from the first attested text until the present-day language, is covered by a representative number of texts. For what concerns the dating of old texts, I rely on the sources used in the *Oxford Syntax of old Romanian*, which, for instance, takes into account filigranology, i.e. the study of watermarks, considered to be safer than the analysis of linguistic particularities in dating the old texts (Timotin 2016: 5). For the sake of precision, the cited examples from the pre-21st century Romanian are accompanied, where possible, by the year and the name of the source. As for the examples from present-day Romanian, the source web page is given.

5.6.3   The case of the experiencer

Besides experiencers in the dative, the dataset contains examples of experiencers in other cases, such as nominative and accusative. An accurate annotation of this parameter is of a great importance for the analysis of the data.

5.6.4   Verb

The majority of the examples, especially the ones instantiating the MIHI EST pattern, showed no great difficulties in annotating for this variable. Nevertheless, in other types of experiencers structures there is a wealth of possibilities which can fill the verb slot: either they are psychological verbs pertaining to any of the three classes proposed by Belletti & Rizzi (1988), or non-psych verbs, in which case the psychological meaning emerges from their co-occurrence with a state noun (Verhoeven 2007: 88).

---

[80] I am profoundly indebted to Prof. Camelia Stan from the University of Bucharest, for her very useful remarks related to the dating and the origin of the texts in the corpus for pre-21st century Romanian.

5.6.5 Complementation and stimulus

In experiencer constructions with a state noun, the stimulus, when expressed, occupies a complement position (cf. *mi-e milă de tine* 'I feel pity for you'). Nevertheless, it has to be noted that the MIHI EST construction is not always followed by a stimulus, but it can be followed by other elements too (*mi-e atât de milă, încât nu pot să mă uit* 'I feel so pity, that I can't even look at it/at them'). Therefore, these variables have been annotated separately in my dataset.

With respect to complementation, the following patterns have been observed: [preposition + N], [preposition + non-finite clause], [*că*-clause], [*să*-clause], [other complementizer + clause], etc. As for the stimulus, it has to be noted that when it refers to a human or to an animate entity, it is mostly realized by a PP, whereas when it refers to a non-animate entity it is mostly realized by a clause. When no stimulus is expressed, I use the notation NA (not applicable), as for all variables in the dataset.

5.6.6 Word order

The word order variable is crucial for the present study, since word order is one of the subject diagnostics used in the study of the MIHI EST structure. In the Excel-document, it occupies three columns, since it is annotated for the experiencer clitic, for the nominal experiencer, realized by a NP or a strong pronoun, as well as for the state noun.

Bearing in mind that clitics usually precede the verb, I expect no variation in the position of the clitics. As for the nominal experiencer and for the state noun, they may occur in preverbal or postverbal position both in present-day language and in pre-21st century Romanian.

5.6.7 Determination and modification

A particularity of the MIHI EST pattern is that the postverbal noun referring to a state is mostly a bare noun. Nevertheless, nouns with a determiner are also possible. The majority of determiners occur when the state noun has a modifier; however, this is not always the case. Among the most frequent determiners occurring in the MIHI EST structure are the indefinite articles (cf. *o* 'a (FEM)', *un* 'a (MASC)') and indefinite pronouns such as *nicio* 'any' (cf. *nu mi-e nicio jenă de chestia asta* 'I don't feel any shame about this thing'). There are situations when a determiner occurs without a modifier (cf. *că mi- e o sete ...* 'because I'm so thirsty ...'), case in which the punctuation or the intonation plays the role of a modifier. When present, the modifier is realized by an adjective (cf. *îi era o rușine nespusă* '(s)he felt an untold shame'), by an adverb (cf. *mi- era foarte sete* 'I was very thirsty'), by an NP (cf. *mi-era o foame de lup* 'I was hungry as a wolf'), or by a clause (cf. *îmi e o foame de behăi* 'I'm so

hungry that I bleat'). In Chapter 7 *infra*, it will be investigated which modifiers are preferred in the MIHI EST structure, what is their relation with the determiner, and whether they occur equally in all historical periods of Romanian.

5.6.8   Semantic class and polarity

The semantic class is a very important variable in determining the semantic domain from which the nouns in the MIHI EST structure come. Besides its significance for the study of the productivity of the construction under scrutiny, this variable may show whether this pattern has the ability to coerce nouns from different semantic fields into the psychological or physiological meaning expressed by this pattern. Indeed, besides the physiological or psychological state nouns, nouns from other semantic fields have been found to occur in the MIHI EST construction recently, such as events (*plecare* 'departure'), time (*iarnă* 'winter'), meteorological phenomena (*furtună* 'storm'), etc. In order to identify possible nuances and relations in the analysis of these structures, the polarity of the state has been also noted in terms of negative vs. positive.

5.6.9   First attestation & Etymology

The annotation for these two variables concerns the state noun. The first attestation variable is meant to provide more insight into the construction's ability to attract new nouns or even verbs. The question arises whether the construction attracts neologisms, or whether it is more appealing for words that already exist in the language. As for the etymology of the noun, it provides information on nuances and tendencies in the selection of new elements (domestic vocabulary or neologisms) to fill the state noun slot.

5.6.10 Type source-text & Source

My aim was to gather a representative number of texts for pre-21st century Romanian as well as for the present-day language. Although the desired variation in genres was not possible for all periods of Romanian, I considered it important to include this information under the variable labeled *Type source-text*, and to annotate it for the whole dataset. This information is given based on another variable, called *Source*, which mentions the work to which a specific example belongs, the year of publication, where possible, and the author.

Chapter 6   Nouns entering the MIHI EST construction

This chapter explores which nouns can enter the MIHI EST construction. Its aim is to gather a well-defined list with the nouns that can co-occur with a dative experiencer and the verb *fi* 'be', both in present-day Romanian (6.2), and in pre-21ˢᵗ century Romanian (6.3). While doing this, I aim to observe to which extent this construction combines with new nouns and whether they belong to the same semantic field of physiological or psychological states (6.4). The last section presents the evolution of the competition between *fi* 'be' and *avea* 'have' in constructions with state nouns (6.5), followed by conclusions (6.6).

6.1   Dataset

Using the advanced query option on Sketch Engine, all examples containing the verb *fi* 'be', preceded by a dative clitic and followed by a noun were gathered from both the corpus for present-day Romanian, roTenTen16 provided by Sketch Engine, and the corpus for the pre-21ˢᵗ century Romanian, created by myself. By means of an advanced query, I searched all structures of the type [DAT *fi* N] (cf. *mi-e lene* 'I am lazy'), where N stands for any noun that can combine with an experiencer in the dative and the verb *fi* 'be'.

The query returned 154 492 examples for present-day Romanian, which I have restricted to a random sample of 100 000 examples, the maximum of data that can be exported from Sketch Engine. The same query in the corpus for the pre-21ˢᵗ century Romanian returned 2 278 examples, which were all preserved.

For each of the two samples, I have automatically generated a list of different sequences (of the type *mi-e foame* 'I am hungry' or *li-era foame* 'they were hungry') using the Frequency tool on the platform. The full list was then exported to Excel and manually annotated. After removing the noise, the remaining examples were centralized and a final list with the nouns occurring in this construction in all periods of Romanian was made. Due to imperfections in tagging or inconsistency in the use of diacritics in the older texts, the query returned a larger amount of noise for pre-21ˢᵗ century Romanian than for present-day Romanian. In spite of that, the data for pre-21ˢᵗ century Romanian allowed

me to collect 29 different state nouns occurring in the MIHI EST construction, whereas the corpus for the present-day language yields 95 different nouns used in this construction.

6.2 The set of nouns in present-day Romanian

Before investigating the inventory of nouns occurring in the MIHI EST construction, it is worth discussing a few ambiguous cases found among the results returned by the query used to extract nouns in the MIHI EST construction. These cases involve mostly adverbs (177) and adjectives (178), which under certain circumstances can be used as nouns as well (179a-d), and a few other elements that are used either as interjections, adverbs, or nouns.

(177) *Mi-* *e* *bine*
   me.DAT is good
   'I feel good'

(178) *Mi-* *e* *cald*
   me.DAT is warm
   'I feel warm'

(179) a. *Mi-* *e* *rău*
    me.DAT is bad
    'I feel bad'

  b. *Bătrânul*   *merge* *rău*
    old.man.the   walks badly
    'The old man walks with difficulty'

  c. *Omul* *rău* *nu* *cunoaşte* *liniştea*   *sufletească*
    man.the bad not knows tranquility.the of.soul
    'A bad person does not know/have peace of mind'

  d. *Răul* *făcut* *se* *plăteşte* *întotdeauna*
    bad.the done SE.PASS pays always
    'One must always pay for the bad he did'

In spite of their ambiguity, such elements are found in other Indo-European languages as well, in which they are analyzed either as adjectives or as adverbs. These examples confirm that Romanian MIHI EST occurs also with adjectives and adverbs. Due to space and time limitations, these structures will not be addressed in the present study.

As for the other elements, some of them are of a special interest, such as *vai* 'woe', *amar* 'bitterness, pain', *haram* 'haram, forbidden', *halal* 'bravo'. Since they are categorized in

Romanian dictionaries as interjections,[81] but also as nouns or adverbs, it is surely not by accident that these words have been returned by the query, their categorization being rather ambiguous. *Vai* 'woe', *amar* 'bitterness', and *haram* 'haram, forbidden' signal adversity, or even danger involving a participant, whereas *halal* 'bravo' was first specific to a situation in which a participant receives words of appreciation and it has evolved into an ironical use in the present-day language.[82] The examples below illustrate how these words are used in present-day Romanian (180)-(183), where they seem to be rather frequent.

(180) *dacă   nu –   vai   nouă!*
      if   not   woe   us.DAT
      'If not, it's not good for us!'   (compania.ro)

(181) *Şi   mi-   e   frig   şi   mi-   e   amar   şi (...)*
      and   me.DAT=   is   cold   and   me.DAT=   is   bitterness   and
      'And I feel cold and I am bittered, and ...'   (clubdaciiliberi.ro)

(182) *dacă   seamănă   cu   tine,   haram   de   ei!*
      if   resemble.3PL=SG   with   you   haram   of   them
      'If they are alike you, God be with them!'   (europeea.ro)

(183) *BRAVOS,   domnule   primar!   HALAL   SĂ   -ŢI   FIE!*
      bravo,   mister.the   Mayor   halal   SUBJ   =you.DAT   be.3SG
      'Congratulations, Mr. Mayor, you should be ashamed!'   (protecuci.ro)

The problematic categorization of these elements is not specific to Romanian. English and German have also the element "woe" (cf. Engl. *woe*, Germ. *weh*), which is classified as an interjection, and as an adverbial element, and has developed nominal and adjectival uses (cf. OED 1989; Kluge 2002). Barðdal et al. (2013: 350) reconstruct the path of "woe" in five sub-branches of Indo-European, namely Indo-Iranian, Italic, Baltic, Slavic and Germanic. One of the findings of this insightful study is that there is nothing in the history of "woe", neither the morphological form, nor the syntax, which can guide toward its analysis as a noun. Notwithstanding, Barðdal et al. (2013) show that this change happens in English, as well as in German: the nominal use of "woe" has been signaled in English in the 5th century and in German in the 8th century (Barðdal et al. 2013: 337). Therefore, according to Barðdal et al. (2013), this element shows no syntax at all outside the MIHI EST construction in the early and archaic Indo-European languages. Instead, it is clearly an interjection, used as an adverbial element in the MIHI EST construction in these Indo-European languages.

---

[81] Cf., among others, DEX online (https://dexonline.ro/), *Micul dictionar academic* – MDA (https://www.webdex.ro/online/micul_dictionar_academic), *Dictionarul etimologic roman* – DER (https://www.webdex.ro/online/dictionarul_etimologic_roman).

[82] The Romanian *halal*, a dedicated exclamative word, is used as an interjection, and has a deprecatory meaning in the present-day language {Citation}. In the past, it expressed positive evaluation. It comes from a Turkish noun, with Arabic origins *halâl* 'lawful' and means 'gratitude, grace, luck'.

Nevertheless, since these items are ambiguous between a nominal and an adverbial use, I have not included them in the present study, but hope to study them in a later research.

Turning to the set of 95 nouns extracted for present-day Romanian, a thorough examination of the selected nouns reveals that the most frequent among them are abstract nouns that refer to physiological or psychological states. Table 6.1 provides a complete list of these nouns with their token counts (i.e. the amount of instantiations of a specific type, in this case, the occurrences of a specific noun) sorted, in descending order, from the most frequent to the least frequent noun.

Table 6.1   Set of nouns occurring in the MIHI EST construction in present-day Romanian

Noun	English translation	Token count
lene	laziness	190
dor	longing1	188
ruşine	shame1	186
scârbă	disgust1	179
teamă	fear2	175
frică	fear1	174
silă	disgust2	168
lehamite	boredom	164
foame	hunger1	154
sete	thirst	153
jenă	embarrassment	150
ciudă	rancor1	137
frig	cold	123
greață	nausea	97
milă	pity1	89
groază	terror	73
târşă	disgust/horror1	46
jale	grief1	44
somn	sleep	31
necaz	rancor2	30
fomică	hunger2	28
ruşinică	shame2	28
răcoare	coolness	24
poftă	craving	16
grijă	worry	15
fomiță	hunger3	9
oroare	horror	8
oftică	rancor3	7
târşeală	disgust/horror2	5
alean	longing2	5
fricuță	fear4	4
toamnă	autumn	4
spaimă	fear3	3
chef	mood/ disposition	3
linişte	tranquility	3
repulsie	repulsion	2
speranță	hope1	2
nelinişte	anxiety	2
dezgust	disgust3	2

Noun	English translation	Token count
tăcere	silence2	2
pace	peace	2
doruleț	longing3	2
așteptare	waiting	2
durere	pain	2
furie	anger	2
melancolie	melancholy	2
cântec	song	1
zid între inimi	wall between hearts	1
atât de primavera	so_springtime	1
foșnet	rustle	1
bucurie	joy	1
freamăt	thrill	1
fericire	happiness	1
miluță	pity2	1
tumult	turmoil	1
muzică	music	1
rană de cântec în gând	wound of unsung song	1
nădejde	hope2	1
iarnă	winter	1
cuvânt nerostit	untold word	1
soare	sun	1
descânt	incantation	1
larmă	jangle	1
nerăbdare	impatience	1
lună	moon	1
nesomn	wakefulness	1
vis	dream	1
ninsoare	snow	1
mâhnire	sorrow	1
noapte_de_alb	night_of_white	1
deznădejde	despair	1
nor de ploi prea albastre plângând	cloud of crying blue rains	1
căldură	warmth	1
nostalgie	nostalgy	1
invidie	envy	1
ochi de priviri	eyes_of_gaze	1
șoaptă	whisper	1
dezamăgire	disappointment	1
jelanie	grief2	1
fum	smoke	1
joc de copii	play_of_kids	1
blândețe	gentleness	1
amurg	twilight	1
panică	panic	1
curcubeu	rainbow	1
pasăre rară	rare_bird	1
tristețe	sadness	1
pică	rancor4	1
vânt	wind	1
plecare	departure	1

Noun	English translation	Token count
zborul din aripi	flight_of_wings	1
ploi	rains	1
zor	rush2	1
furtună	storm	1
gând de ploaie	thought_of_rain	1
Total		2 784

As is evident from Table 6.1, only a few of the nouns occurring in the MIHI EST pattern, namely 26 %, show a high token count in present-day Romanian. For practical reasons, I consider, in the present study, nouns with occurrences between 15 and 190 as having a medium-to-high token count: *lene* 'laziness' (190), *dor* 'longing' (188), *ruşine* 'shame' (186), *scârbă* 'disgust' (179), *teamă* 'fear' (175), *frică* 'fear' (174), *poftă* 'appetite' (16) and *grijă* 'worry' (15). A smaller number of nouns (22 %), with occurrences between two and nine, are considered as having a medium-to-low token count: *fomiţă* 'little hunger' (9), *oftică* 'resentment/rancor' (7), *târşeală* 'disgust/shame/fear' (5), *dezgust* 'disgust' (2). Eventually, a third much larger group, 52 %, contains nouns occurring only once in this construction, called *hapax legomena* (i.e. unique occurrences in a given dataset): *bucurie* 'happiness', *tristeţe* 'sadness', *invidie* 'envy'. Figure 6.1 visualizes the token counts of the set of nouns for present-day Romanian, without the hapax legomena, which I sometimes refer to also with the label *hapax*.
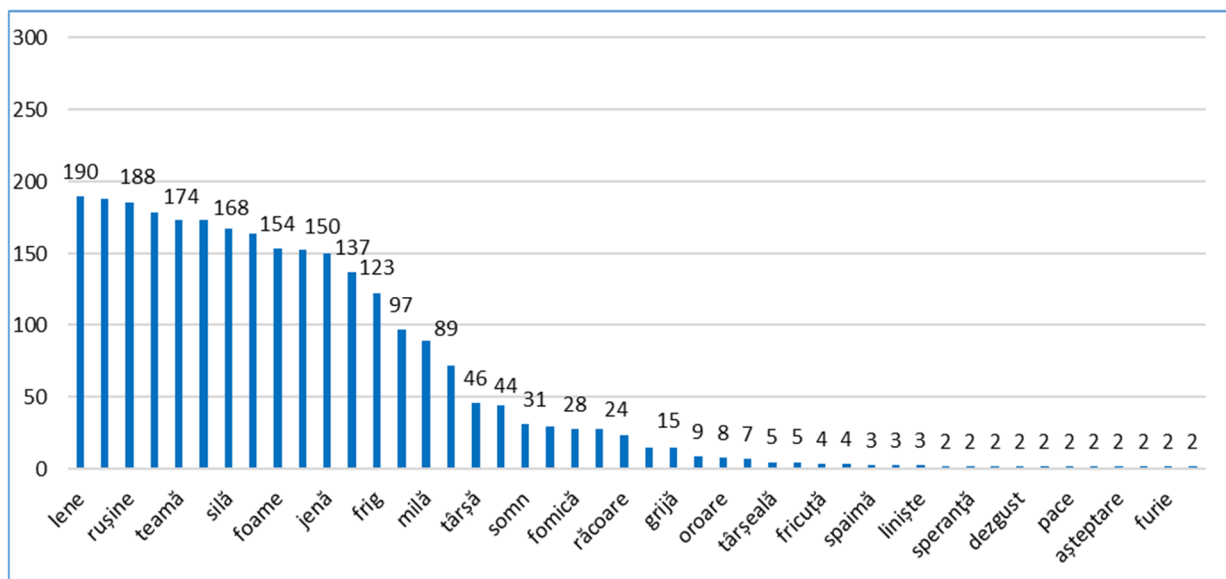
Figure 6.1   State nouns in the MIHI EST construction in the 21st century (without hapaxes)

6.3   The set of nouns in pre-21st century Romanian

As for the 29 state nouns occurring in the MIHI EST pattern between the 16th and the 20th centuries, all of them are abstract nouns expressing physiological or psychological states. Table 6.2 provides an exhaustive list of these nouns with their token count.

Table 6.2   Set of nouns occurring in the MIHI EST construction in pre-21st century Romanian

Noun	English translation	Token count
frică	fear1	280
ruşine	shame1	205
milă	pity1	174
teamă	fear2	102
foame	hunger1	79
dor	longing1	63
sete	thirst	49
somn	sleep	41
frig	cold	37
silă	disgust2	32
scârbă	disgust1	22
greaţă	nausea	20
ciudă	rancor1	17
grijă	worry	9
lene	laziness	9
spaimă	fear3	8
groază	terror	8
necaz	rancor2	7
răcoare	coolness	5
jale	grief1	4
poftă	appetite1	3
grabă	rush1	2
jind	craving2, longing3	2
lehamite	boredom	2
bucurie	joy	1
nevoie	need	1
zor	rush2	1
deznădejde	despair	1
târşeală	2disgust/shame/fear	1
Total		1 185

Contrary to present-day Romanian, the majority of the nouns (45 %) in pre-21st century Romanian have a medium-to-high token count, between 17 and 280 tokens: *frică* 'fear' (280), *ruşine* 'shame' (205), *greaţă* 'nausea' (20) and *ciudă* 'envy' (17). A smaller group (38 %) has a medium-to-low token frequency, between two and nine: *lene* 'laziness' (9), *groază* 'horror' (8), *jind* 'craving, longing' (2) and *lehamite* 'boredom' (2). As for the *hapax legomena*, they form the smallest group (17 %), containing five elements: *bucurie* 'happiness' (1), *nevoie* 'need' (1), *zor* 'rush' (1), *deznădejde* 'despair' (1) and *târşeală* 'disgust/shame/fear' (1).

Figure 6.2 graphically illustrates the token count of the set of nouns for the pre-21st century Romanian, without the hapax legomena.
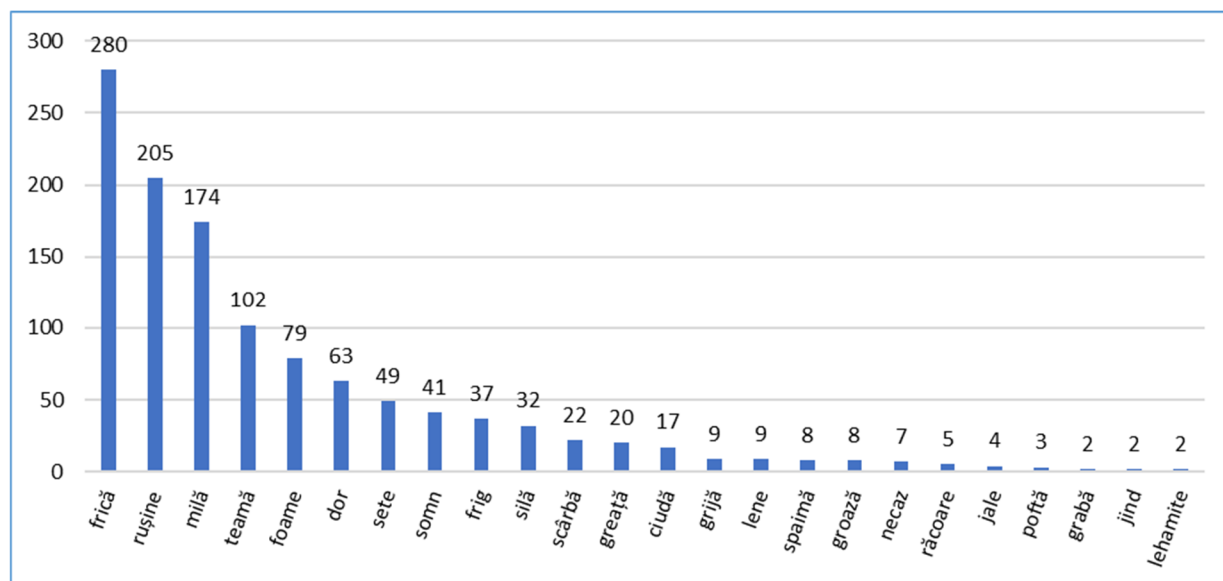
Figure 6.2   State nouns in the MIHI EST construction in the 16th–20th centuries (without hapaxes)

6.4   Etymology and innovation

A closer look at the etymology of the nouns occurring in the MIHI EST construction, both in the present-day language and in pre-21st century Romanian, shows that the majority are old nouns, either of Latin or Slavic origin. A few of these disappear from this construction in the 19th century (*nevoie* 'need') or the 21st century (*jind* 'craving, longing') – even though these nouns are still used in present-day Romanian in other constructions. New nouns are introduced with every period: *teamă* 'fear' in 19th century, *târşeală* 'disgust/horror', *jenă* 'embarrassment' in the 20th century, and *târşă* 'disgust, horror' in the 21st century, among others. Table 6.3 illustrates these dynamics, affecting the noun slot of the MIHI EST construction across all periods of Romanian. It also indicates the approximate moment in time when a certain noun starts to occur in, or it disappears from this construction.

Table 6.3   Dynamics of the set of nouns across the centuries (16th–21st centuries)

Old Romanian			Modern Romanian	Contemporary Romanian	Present-day Romanian
16th–18th	17th	18th	19th	20th	21st
LATIN: *sete* 'thirst' SLAVIC: *milă* 'pity', *scârbă* 'disgust' GREEK: *frică* 'fear' DERIVED: (lat.) *greață* 'nausea'					
SLAVIC: *nevoie* 'need'					
	LATIN: *rușine* 'shame', *dor* 'longing', *foame* 'hunger', *frig* 'cold' BULGARIAN: *grijă* 'care, worry'				
		SLAVIC: *ciudă* 'envy', *groază* 'terror' DERIVED: (sl.) *poftă* 'craving', (alb.) *bucurie* 'joy', (lat.) *teamă* 'fear'			
			LATIN: *somn* 'sleepiness' FRENCH: *jenă* 'embarassment' SLAVIC: *silă* 'disgust', *lene* 'laziness', *jale* 'grief', *necaz* 'rancor' BULGARIAN: (?) *lehamite* 'boredom' UNKN. ETYMOL.: *spaimă* 'fear' DERIVED: (lat.) *teamă* 'fear', (lat.) *răcoare* 'coldness'		
			DERIVED: (bg.) *grabă* 'rush, hurry'		
				TURKISH.: *zor* 'rush, hurry' UNKN. ETYMOL.: *târșeală* 'disgust' DERIVED: (sl.) *deznădejde* 'discouragement'	
				DERIVED: (sl.) *jind* 'desire'	
					LATIN: *toamnă* 'autumn', SLAVIC: *nădejde* 'hope' HUNGARIAN: *alean* 'longing' NEO-GREEK: *furtună* 'storm' UNKN. ETYMOL.: *târșă* 'disgust' DERIVED: (lat.) *blândețe* 'softness'

Based on the present data, there is evidence that the MIHI EST construction attracted new nouns in every historical period of Romanian (cf. among others, *ciudă* 'rancor', *nevoie* 'need' in old Romanian, *grabă* 'hastiness/ rush' and *poftă* 'craving' in modern Romanian; *jale* 'grief' and *greață* 'nausea' in contemporary Romanian). Moreover, starting with contemporary Romanian, the construction freely allows the occurrence of synonymous nouns in it. For instance, the derived Slavic *jind* 'craving, longing', synonym

of *poftă* 'craving', occurs in the MIHI EST pattern starting with the 20ᵗʰ century. A century later, it disappears from this construction, although it continues to be used in other types of constructions until today. The Turkish *zor* 'hastiness/ rush', synonym of the derived *grabă* 'hastiness/ rush', occurs in the dative experiencer construction starting with the 20ᵗʰ century, and it keeps being used until today, as opposed to *grabă*, which entered the construction much earlier, and disappears from it in the 21ˢᵗ century.

Not surprisingly, most of these nouns seem to occur also in the HABEO construction, where they combine with a nominative experiencer, in order to express the same meaning as the MIHI EST construction. This competition between the MIHI EST and HABEO constructions is not a novelty, since it is known to have already existed in Classical Latin. An in-depth study of this competition, although very tempting, goes far beyond the aims of this dissertation. Hence, in the next section, I describe a small study I carried out, on the relation between the two constructions in Romanian across the centuries.

Furthermore, during the 21ˢᵗ century, a larger number of state nouns, among which also synonyms of the already existing ones, are allowed in the MIHI EST construction. For instance, *poftă* 'craving' has another synonym, the Turkish *chef* 'appetite/ will', which has been allowed in the MIHI EST pattern in present-day Romanian. Likewise, occurrences of alternative nouns such as *spaimă* 'fear', synonym of *frică 'fear'* and of *teamă 'fear'* are found in the present-day language. Moreover, I have also pointed out that the MIHI EST construction shows, during the past decades, a tendency to admit nouns from different semantic fields, such as event (*plecare* 'departure'), time (*atât de primăvară* 'so springtime'), communication (*cuvânt nerostit* 'untold word'), etc. (cf. Ilioaia 2020). These new nouns are used metaphorically in the MIHI EST construction, and are being coerced into the construction's meaning, that of a physiological or a psychological state (184a-b).

(184) a. *Mi-*   *e*   *atât*   *de*   *primăvară*   *încât*   *șoptesc*
me.DAT=   is   so   of   springtime   that   whisper.1SG
întunericului   *să (...)*
dark.the.DAT   SUBJ
'I feel so springtime that I whisper the dark to (...)'   (facebook.com/permalink.php)

   b. *mi-*   *e*   *cuvânt*   *nerostit ...*   *mi-*   *e (...)*
me.DAT=   is   word   untold   me.DAT=   is
'I feel like an untold word, I feel (...)'   (alexsmallthings-desprenimic.blogspot.com)

Recall that, as Lauwers & Willems 2011: 1220) highlight, in Construction Grammar coercion constitutes a major argument in favor of the existence of constructions as independent form/meaning pairings. The ability of a construction to change the meaning of a lexical item that occurs in it has been used as an empirical test to argue for constructions carrying a particular meaning on their own, irrespective of the lexical items that instantiate them. The evidence presented in this section confirms, once again, the construction status of the MIHI EST structure.

6.5 Evolution of the competition between *fi* 'be' and *avea* 'have' with state nouns

The tendency to innovation observed with the MIHI EST construction in Romanian raises several questions. One of them is whether the other constructions in which these state nouns occur are as frequent as the MIHI EST construction. Among all possible competing constructions, with an accusative, or a dative experiencer, partially described in Chapter 4, or constructions with a nominative experiencer, I focus on the transitive HABEO construction, in which a state noun occurs with the verb *avea* 'have' and a nominative experiencer (185). This construction carries the same meaning as the MIHI EST construction and has replaced the MIHI EST construction in other Romance languages like French, Italian and Spanish.

(185) *După   acéia,   Gligorie-vodă,   având   grijă   și   mare*
      after   that   Gligorie-voivode   having   worry   and   big
      frică, [...]
      fear
      'Then, since Gligorie-voivode was worried and very afraid, [...]'
                                      (1659 Constantin Cantacuzino, *Letopisețul*)

In my dataset, the total number of examples containing a MIHI EST onstruction (3 969 occurences) is higher than the number of examples with the HABEO construction (3 038 occurences). At a closer look, however, the internal dynamics of these and of other competing constructions is more complex (cf. Figure 6.3).
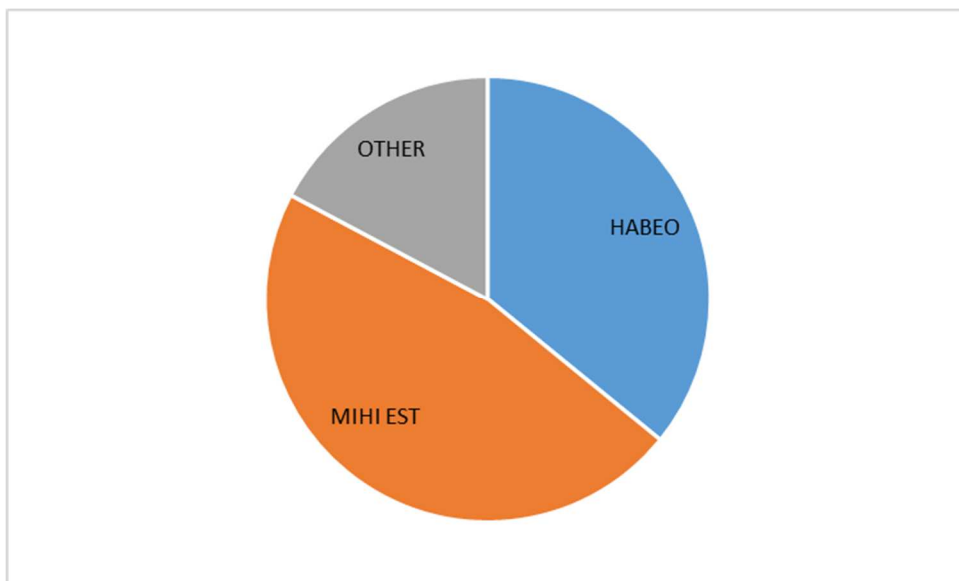

Figure 6.3   Repartition of experiencer constructions in my dataset

As expected, not all the nouns on my list can occur in the HABEO construction. Indeed, less frequent nouns and hapax legomena, especially the ones that belong to other semantic fields (186a), cannot occur in the HABEO construction, or if they do, they may convey a totally different meaning, which is not relevant here (186b).

(186) a. *Mi-*   *e*   *toamnă*
       mi.DAT=   is   autumn
       'I feel melancholic'
    b. *Anul*   *trecut*   *am*   *avut*   *o*   *toamnă*   *de*   *vis*
       year.the   past   have.1SG   had   an   autumn   of   dream
       'Last year, I had the autumn of my dreams'

Moreover, the less frequent nouns expressing a state, which have entered recently into the MIHI EST construction, are expected to have a high frequency in the HABEO construction. My hypothesis is that the state nouns on my list first occurred in the HABEO construction and, over time, have shifted to the MIHI EST construction, which has become more appropriate for expressing physiological and psychological states. Hence, I claim that the structure [*habeo* + state noun] occurred more frequently in old Romanian than in the present-day language. Indeed, the HABEO construction typically occurs in present-day Romanian with concrete nouns and becomes more suitable for expressing predicative possession in Romanian.

My hypothesis that in old Romanian the [*habeo* + state noun] construction was more frequent than the MIHI EST construction is based on an earlier case study of two of the most frequent nouns entering the MIHI EST construction in Romanian: *frică* and *teamă*, both meaning 'fear' (Van Peteghem & Ilioaia 2017). This study shows that *frică* was present in both constructions in the 16th century, but was more frequent in the HABEO construction, whereas *teamă* occurred in the HABEO construction at the end of the 16th century, when it was attested in the language for the first time, and shifted to the MIHI EST construction later on, in the 19th century.

In what follows, I first give an overview of the competition between the two constructions, based on the entire dataset. Subsequently, I describe a few diachronic case studies in order to investigate per noun if, indeed, the shift takes place from the HABEO to the MIHI EST construction.

Figure 6.4 shows the competition between the two constructions for the most frequent nouns in the complete dataset, for all periods taken together. Note that, for readability purposes, the graph does not plot the nouns with a relative frequency under 0,2 %. The frequencies shown in this graph have been obtained by dividing the absolute frequency of each noun to the total number of occurrences in the dataset, i.e. 7 007 occurrences. From this graph, it can be observed that the nouns occurring less frequently in the MIHI EST construction still have a higher frequency in the HABEO construction, whereas the more frequent a noun is in the MIHI EST construction, the less it occurs in the HABEO construction.
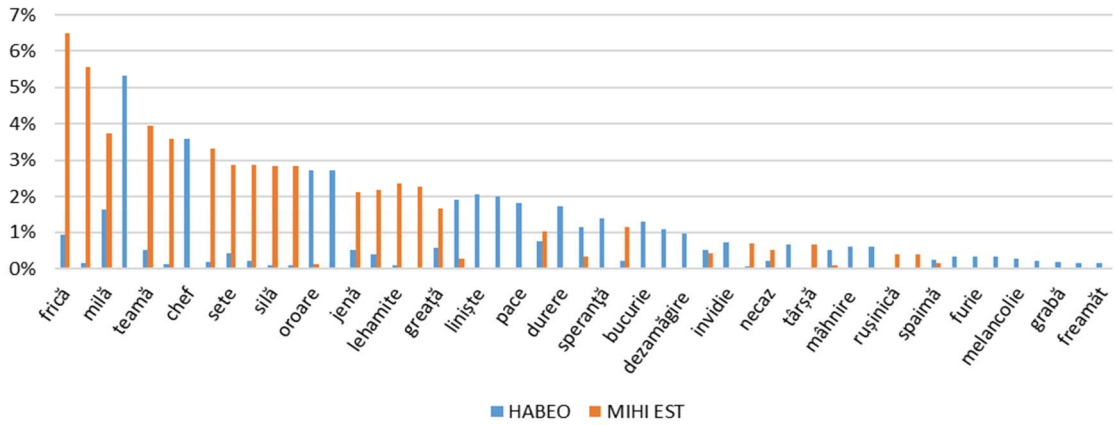
Figure 6.4   Competition HABEO vs. MIHI EST – frequent nouns (16th–21st)

As for the evolution of this rivalry, Table 6.4 gives an overview of the competition between the two constructions throughout the centuries. Since the sets of examples for each historical period are not of equal size, the absolute figures are followed by relative figures in the table. I have calculated the relative figures by dividing the absolute figures to the total number of examples in the dataset (7 007). These data are visualized in the graph in Figure 6.5, which clearly demarcates that it is only in the first historical period that the HABEO construction was dominant. Starting with the 19th century, the MIHI EST construction prevails, which corroborates my hypothesis.

Table 6.4   Competition HABEO vs MIHI EST - diachronic perspective

	HABEO	%	MIHI EST	%	Total	%
16th–18th	232	3,3%	159	2,3%	391	5,6%
19th	195	2,8%	335	4,8%	530	7,6%
20th	301	4,3%	691	9,9%	992	14,2%
21st	2 310	33,0%	2 784	39,7%	5 094	72,7%
Total	3 038	43,4%	3 969	56,6%	7 007	100,0%
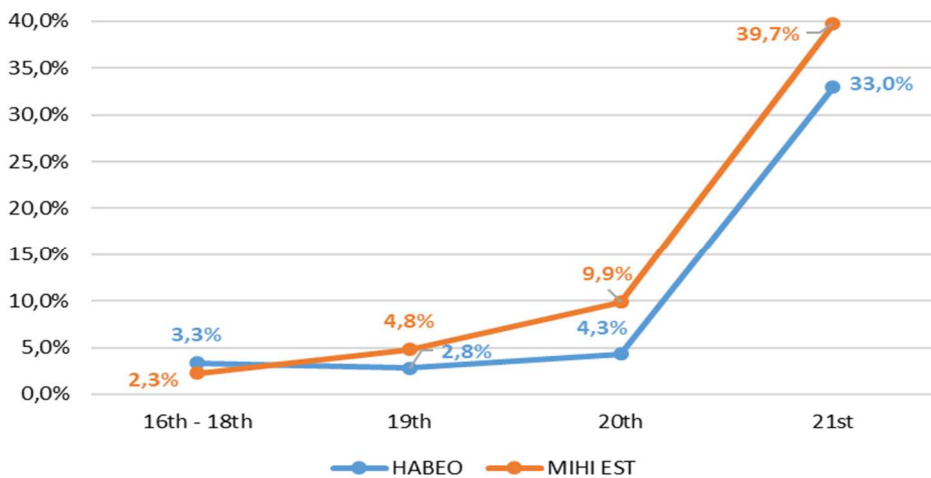
Figure 6.5   Competition HABEO vs. MIHI EST - diachronic perspective

The high frequency of the HABEO construction in the 21ˢᵗ century can be explained by the high number of new nouns entering the MIHI EST construction, nouns which already occur in the HABEO construction.

In what follows, I focus on a limited number of nouns and I investigate whether these nouns occurred first in the HABEO construction, and which of the two rival constructions prevails throughout the centuries. From the list of nouns given in Sections 6.2 and 6.3, in this chapter, I have selected the following nouns: *frică* 'fear', *teamă* 'fear', *dor* 'longing', *scârbă* 'disgust', *poftă* 'craving', *milă* 'pity', *grijă* 'worry, care', *spaimă* 'fear', *grabă* 'rush, hurry' and *nevoie* 'need'.

The first two selected nouns, *frică* and *teamă*, have already been investigated in Van Peteghem & Ilioaia (2017) who show that both nouns occurred first in the HABEO construction, before shifting to the MIHI EST construction. Yet, these nouns entered into the language at different moments and from different sources: *frică*, from the Gr. *φρίκη*, is attested at the end of the 15ᵗʰ century, but might be older, whereas *teamă*, derived from the verb *teme* 'to fear' (< Lat. *timere*), is first attested in the late 16ᵗʰ century. Table 6.5 and Table 6.6 give the absolute and the relative frequencies respectively for *frică* and *teamă* for all the historical periods. Their evolution throughout the centuries is graphically illustrated in Figure 6.6 and Figure 6.7.

Table 6.5   *Frică* 'fear' - HABEO vs. MIHI EST construction

	Frică 'fear'					
	HABEO	%	MIHI EST	%	Total	%
16ᵗʰ–18ᵗʰ	31	6,0%	20	3,8%	51	9,8%
19ᵗʰ	14	2,7%	87	16,7%	101	19,4%
20ᵗʰ	9	1,7%	173	33,3%	182	35,0%
21ˢᵗ	12	2,3%	174	33,5%	186	35,8%
Total	66	12,7%	454	87,3%	520	100,0%
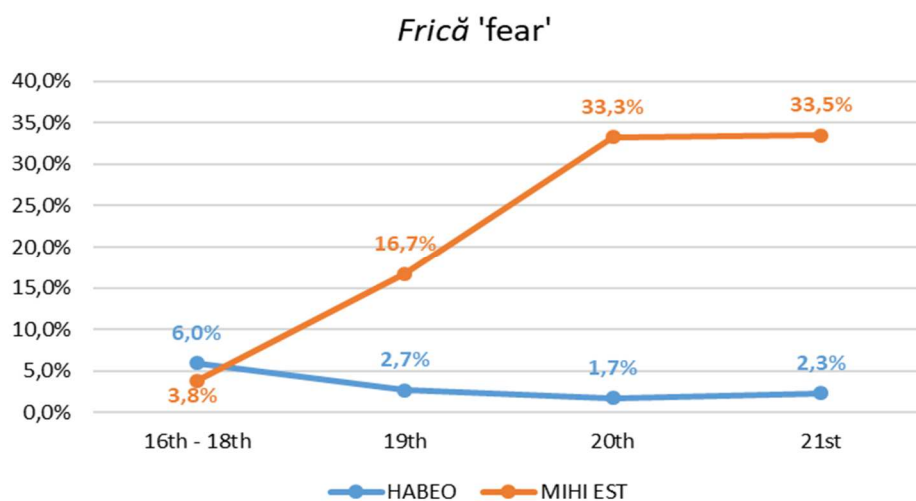
Figure 6.6   *Frică 'fear'* - HABEO vs. MIHI EST construction

Table 6.6   *Teamă 'fear'* - HABEO vs. MIHI EST construction

	Teamă 'fear'					
	HABEO	%	MIHI EST	%	Total	%
16th–18th	5	1,6%	0	0,0%	5	1,6%
19th	9	2,9%	5	1,6%	14	4,5%
20th	10	3,2%	97	31,1%	107	34,3%
21st	11	3,5%	175	56,1%	186	59,6%
Total	35	11,2%	277	88,8%	312	100,0%
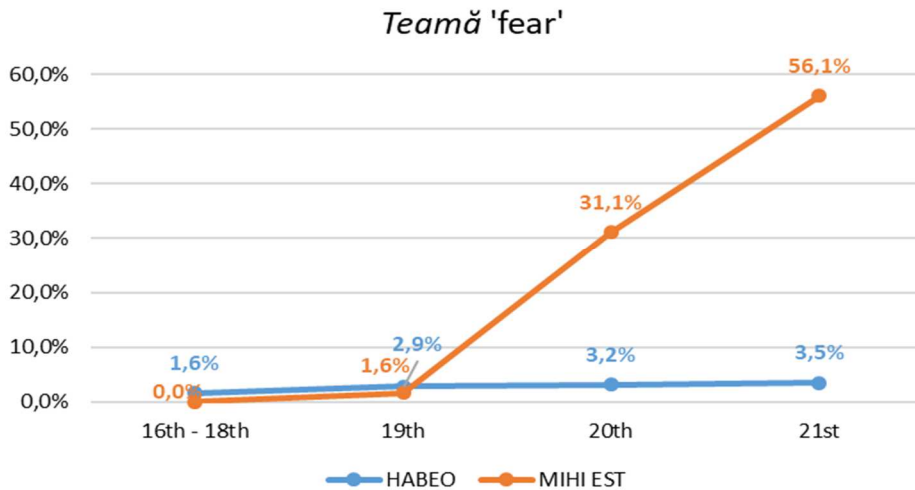
Figure 6.7   *Teamă 'fear'* - HABEO vs. MIHI EST construction

The two graphs on *frică* 'fear' and *teamă* 'fear' confirm that *frică* was more frequent in the HABEO construction than in the MIHI EST construction during the first historical period (16th–18th centuries). The situation has completely reversed in the 19th century, when the MIHI EST construction becomes by far the dominant construction with *frică*. As for *teamă*, this noun does not appear in the MIHI EST construction in the 16th or 17th centuries and is rare also in the 18th century. It then occurs exclusively in the HABEO construction. It was only in the 19th century that *teamă* became more common, and occurred in both constructions, with the HABEO construction as the most dominant one. The shift to the MIHI EST construction as the dominant construction became visible only in the 20th century.

A similar evolution is observed for the noun *dor* 'longing' (< Lat. *dolus*), which was used, during the first historical period, exclusively with the HABEO construction, exactly like *teamă*, but which switched to the MIHI EST in the 19th century. As evident from Table 6.7, with *dor*, and as it was the case with *frică* and *teamă*, the MIHI EST construction prevails ever since. Figure 6.8 visualizes the dynamics of the competition between the two constructions with *dor* 'longing' throughout the centuries.

Table 6.7   *Dor* 'longing' - HABEO vs. MIHI EST construction

	Dor 'longing'					
	HABEO	%	MIHI EST	%	Total	%
16th-18th	3	1,2%	0	0,0%	3	1,2%
19th	2	0,8%	20	7,7%	22	8,5%
20th	2	0,8%	43	16,5%	45	17,3%
21st	2	0,8%	188	72,3%	190	73,1%
Total	9	3,5%	251	96,5%	260	100,0%
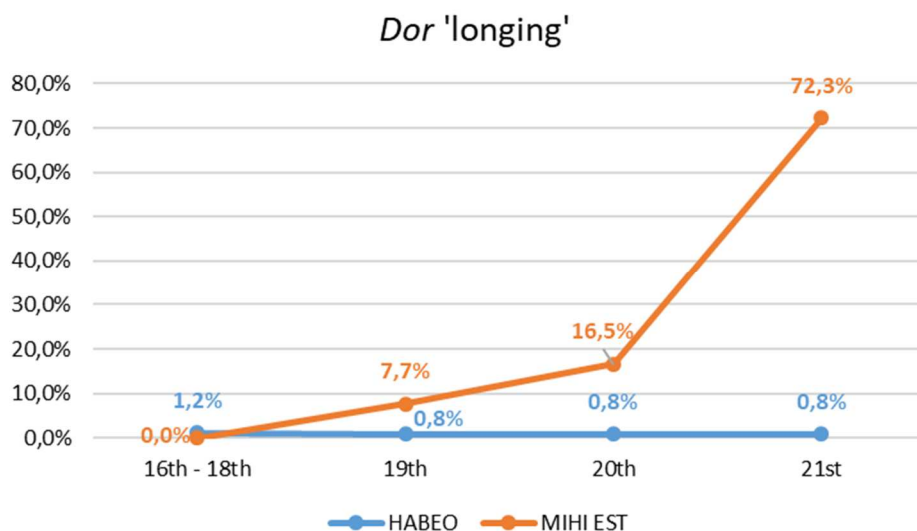
Figure 6.8   *Dor* 'longing' - HABEO vs. MIHI EST construction

As for the noun *scârbă* 'disgust', it occurred in both constructions during the 16th to 18th centuries, yet most frequently in the HABEO construction, as the absolute figures show in Table 6.8, below. As opposed to *dor* 'longing', which continues to occur in the HABEO construction until the 21st century, *scârbă* has completely disappeared from the HABEO construction already in the 19th century. Note also that its frequency in the MIHI EST construction increased significantly, from five occurrences in the 19th century, to 179 occurrences in the 21st century, as can be seen in Table 6.8 and gauged from the graph in Figure 6.9

Table 6.8   *Scârbă* 'disgust' - HABEO vs. MIHI EST construction

	Scârbă 'disgust'					
	HABEO	%	MIHI EST	%	Total	%
16th-18th	14	6,5%	2	0,9%	16	7,4%
19th	0	0,0%	5	2,3%	5	2,3%
20th	0	0,0%	15	7,0%	15	7,0%
21st	0	0,0%	179	83,3%	179	83,3%
Total	14	6,5%	201	93,5%	215	100,0%
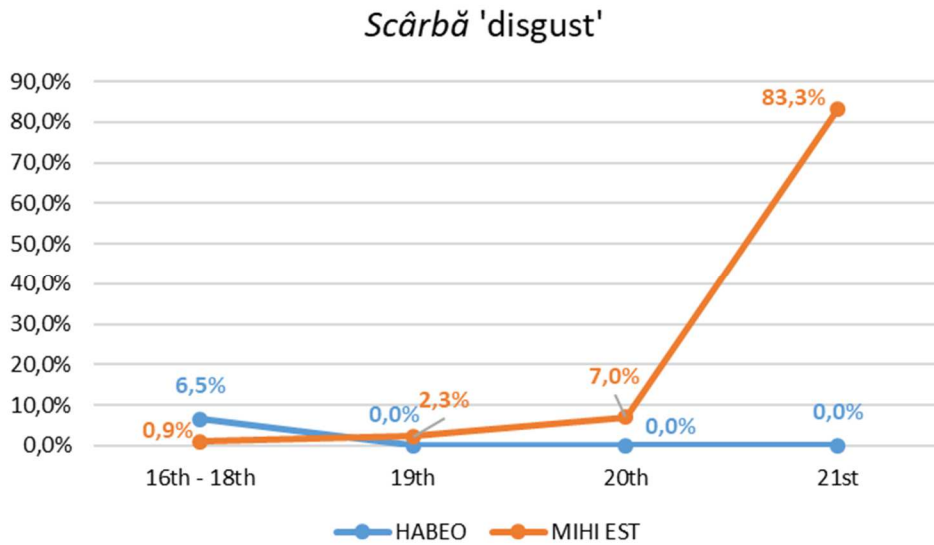
Figure 6.9   *Scârbă* 'disgust' - HABEO vs. MIHI EST construction

The evolution of *grijă* 'worry, care' is completely different. It first occurred in both constructions, the HABEO construction being the dominant one, then, in the 19th century, was found only in the HABEO construction. As the facts in Table 6.9 show, it disappeared from both constructions in the 20th century, and appeared again in the 21st century, in the MIHI EST construction only. I could not find an explanation for this peculiar evolution, but I would certainly not exclude the possibility of this being a result of insufficient relevant examples for certain historical periods. In Figure 6.10, this evolution is visually represented.

Table 6.9   *Grijă* 'worry, care' - HABEO vs. MIHI EST construction

			Grijă 'worry'			
	HABEO	%	MIHI EST	%	Total	%
16th-18th	43	41,7%	9	8,7%	52	50,5%
19th	36	35,0%	0	0,0%	36	35,0%
20th	0	0,0%	0	0,0%	0	0,0%
21st	0	0,0%	15	14,6%	15	14,6%
Total	79	76,7%	24	23,3%	103	100,0%
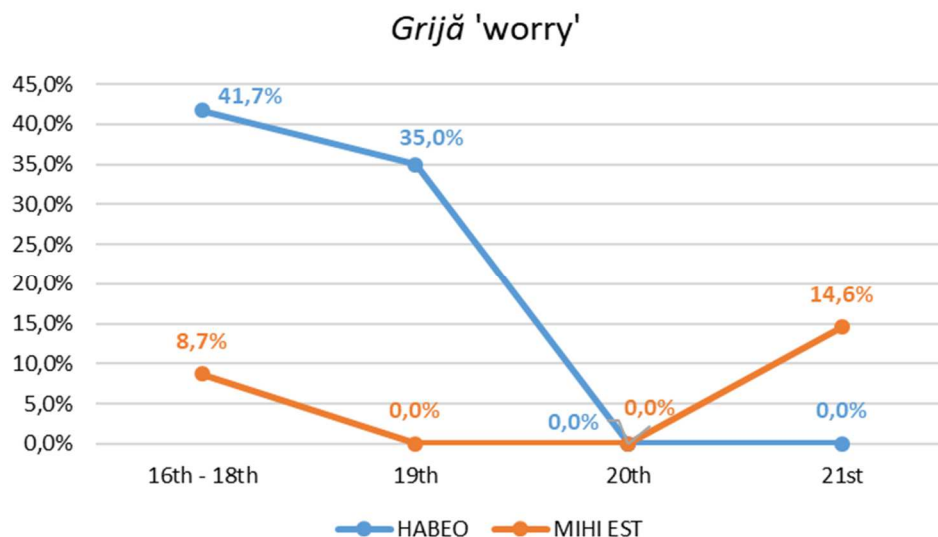
Figure 6.10 *Grijă* 'worry, care' - HABEO vs. MIHI EST construction

In the following, I focus on what *milă* 'pity' and *poftă* 'craving' have in common and why their frequency in the HABEO construction increased alongside with their frequency in the MIHI EST construction throughout the centuries.

The noun *milă* (Sl. *milŭ*) 'pity' is an old one, attested very early, which occurs rather frequently in the language in all kinds of structures. This noun contradicts my hypothesis that the state nouns occur first in the HABEO construction, before being recruited by the MIHI EST construction. Indeed, *milă's* occurring in the MIHI EST construction is first attested at the end of the 15th century, in the first translated text, whereas occurrences of *milă* in the HABEO construction are not found in my dataset before the beginning of the 17th century. At a closer look at the distribution of its occurrences, it turns out that the occurrences of *milă* in translated texts outnumber the occurrences of this noun in original texts between the 15th-16th centuries (11 in translated texts vs. six in original texts). However, this does not change the fact that this noun, indeed, is first attested in the MIHI EST construction, and only a century later in the HABEO construction.

As for *poftă*, it is a noun derived early from the verb *pofti* (< Sl. *pohotěti*), and it is still much used in the present-day language, in different types of structures. The graphs in Figure 6.11 and in Figure 6.12 show that there is a crucial difference between the two nouns: with *milă*, the dominant construction is the MIHI EST construction, whereas *poftă* occurs mostly in the HABEO construction.[83] Table 6.10 and Table 6.11 give an overview of their absolute frequencies for each period, as well as the relative facts. Note that, in parallel, the two nouns also occur in the rival construction, to a different degree.

---

[83] Besides the MIHI EST and the HABEO constructions, *poftă* 'craving' occurs also with other verbs with dative experiencers, including structures like *mi se face poftă* and *îmi vine poftă*, both meaning 'my appetite comes', in order to express the same physiological state of craving.

Table 6.10   *Milă* 'pity' - HABEO vs. MIHI EST construction

Milă 'pity'						
	HABEO	%	MIHI EST	%	Total	%
16th-18th	5	1,3%	65	17,2%	16	7,4%
19th	20	5,3%	43	11,4%	5	2,3%
20th	18	4,8%	66	17,5%	15	7,0%
21st	72	19,0%	89	23,5%	179	83,3%
Total	115	30,4%	263	69,6%	215	100,0%
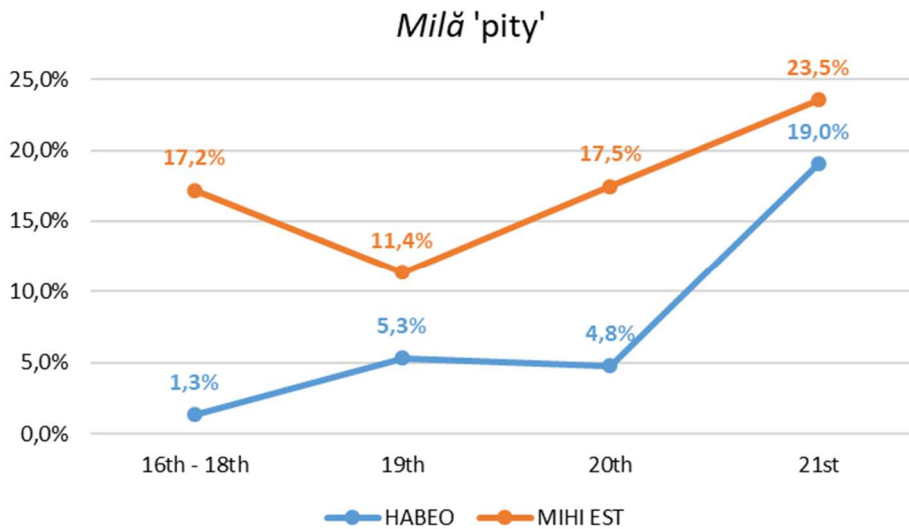
Figure 6.11 *Milă* 'pity'- HABEO vs. MIHI EST construction

Table 6.11   *Poftă* 'craving' - HABEO vs. MIHI EST construction

Poftă 'craving'						
	HABEO	%	MIHI EST	%	Total	%
16th-18th	6	3,9%	1	0,7%	7	4,6%
19th	13	8,5%	1	0,7%	14	9,2%
20th	24	15,7%	1	0,7%	25	16,3%
21st	91	59,5%	16	10,5%	107	69,9%
Total	134	87,6%	19	12,4%	153	100,0%
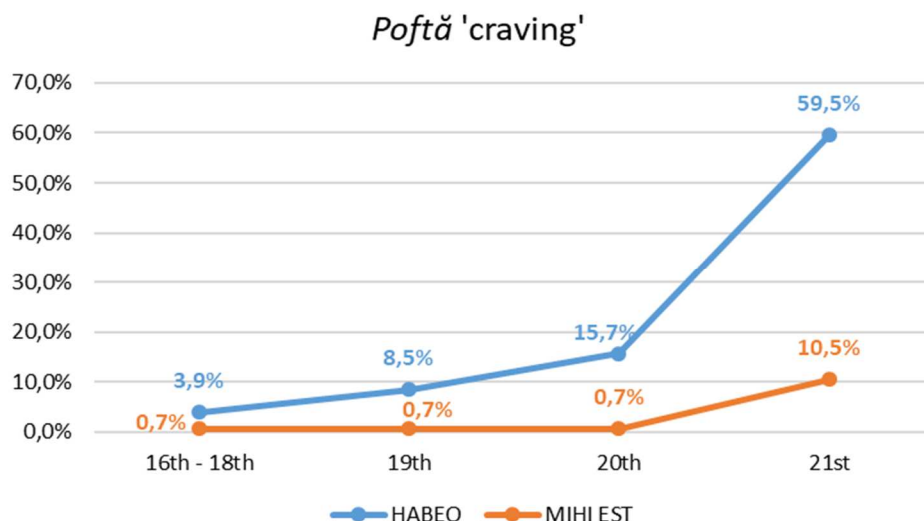
Figure 6.12 *Poftă* 'craving' - HABEO vs. MIHI EST construction

The surprising evolution of *poftă* raises the question whether certain nouns may show a different tendency, decreasing in frequency in the MIHI EST construction, while the HABEO construction becomes dominant. Among the nouns which occurred in the MIHI EST construction during the first two historical periods (i.e. from the 16th century, until the end of the 19th century), three nouns show this evolution: *spaimă* 'fear', *grabă* 'hurry, hastiness' and *nevoie* 'need', switching to the HABEO construction. Note that they are early attested nouns, two of them (*spaimă* and *nevoie*) at the end of the 15th century, and *grabă* in the 17th century.

Spaimă, as opposed to the other two nouns, can still be found in the MIHI EST construction in the 21st century, but its frequency has decreased to half, as it can be seen in Table 6.12. The other two nouns, *grabă* and *nevoie*, disappear completely from the MIHI EST construction in the 21st century and are now only found with the HABEO construction.

Table 6.12 *Spaimă* 'fear'- HABEO vs. MIHI EST construction

	HABEO	%	MIHI EST	%	Total	%
			Spaimă 'fear'			
16th-18th	0	0,0%	0	0,0%	0	0,0%
19th	0	0,0%	6	20,7%	6	20,7%
20th	3	10,3%	2	6,9%	5	17,2%
21st	15	51,7%	3	10,3%	18	62,1%
Total	18	62,1%	11	37,9%	29	100,0%
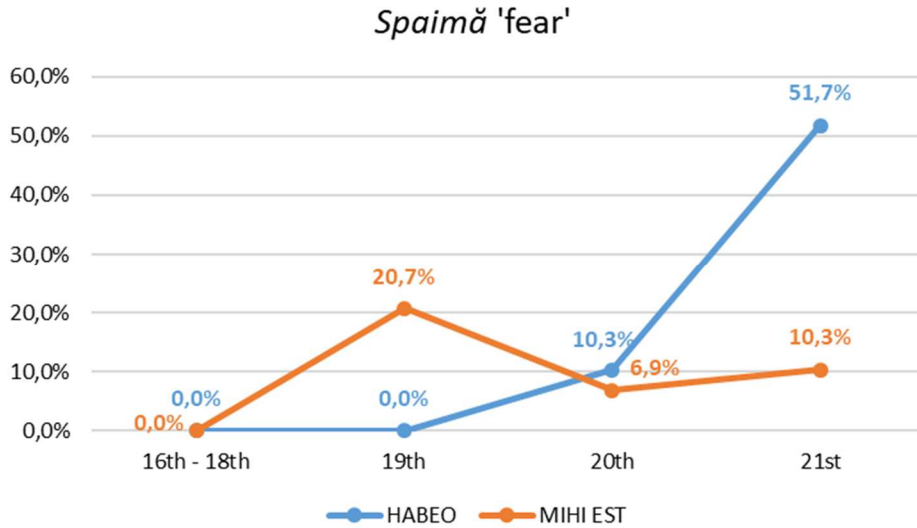
Figure 6.13 *Spaimă* 'fear'- HABEO vs. MIHI EST construction

Nevertheless, it has to be mentioned that these nouns have never really been deep-rooted in the MIHI EST construction, given that their highest frequency in this construction is six occurences for *spaimă*, (cf. Table 6.12) and only one occurrence for each of the other two nouns (cf. Table 6.13 for *grabă* and Table 6.14 for *nevoie*). Moreover, *nevoie* is the only one to occur frequently in the HABEO construction, as shown in Table 6.14. The other two nouns occur scarcely in the HABEO construction: *spaimă* increases from three occurrences in the 20th century to 15 occurrences in the 21st century, whereas *grabă* increases from three occurrences in the 19th century to eight occurrences in the 21st century. In the graphs in Figure 6.13–Figure 6.15, these facts have been plotted, facilitating a visual comparison between the three nouns.

Table 6.13 *Grabă* 'rush, hurry' – HABEO vs. MIHI EST construction

Grabă 'rush, hurry'						
	HABEO	%	MIHI EST	%	Total	%
16th–18th	0	0,0%	0	0,0%	0	0,0%
19th	3	21,4%	1	7,1%	4	28,6%
20th	1	7,1%	1	7,1%	2	14,3%
21st	8	57,1%	0	0,0%	8	57,1%
Total	12	85,7%	2	14,3%	14	100,0%
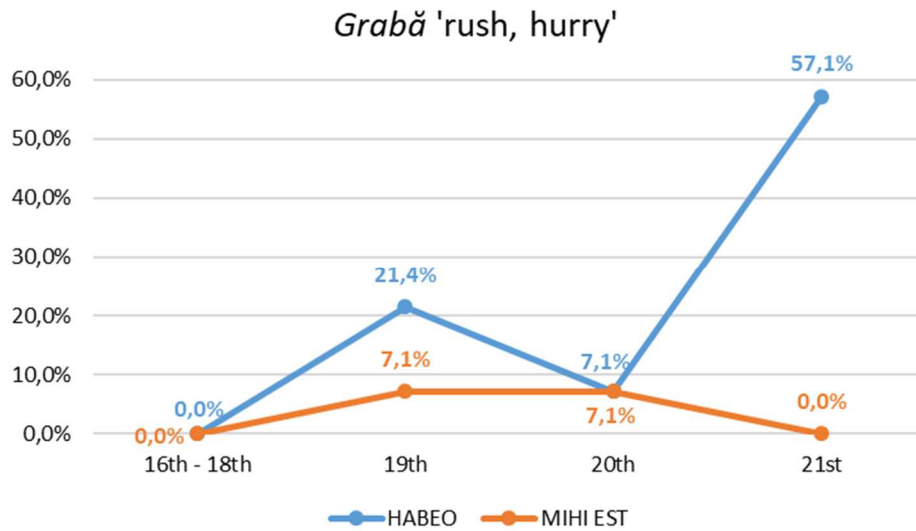
Figure 6.14 *Grabă* 'rush, hurry' - HABEO vs. MIHI EST construction

Table 6.14 *Nevoie* 'need' - HABEO vs. MIHI EST construction

Nevoie 'need'						
	HABEO	%	MIHI EST	%	Total	%
16th-18th	12	3,2%	1	0,3%	13	3,5%
19th	47	12,6%	0	0,0%	47	12,6%
20th	125	33,4%	0	0,0%	125	33,4%
21st	189	50,5%	0	0,0%	189	50,5%
Total	373	99,7%	1	0,3%	374	100,0%
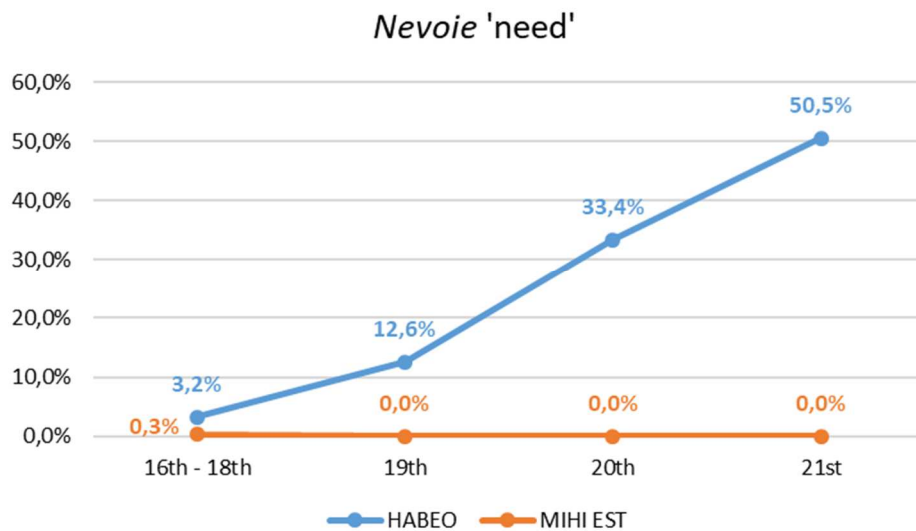
Figure 6.15 *Nevoie* 'need' – HABEO vs. MIHI EST construction

In the case of *spaimă*, a possible explanation for its behavior can be the competition with the other two nouns meaning 'fear', *frică* and *teamă*, which are also among the most frequently-used nouns in the MIHI EST construction. The evolution of *grabă* is more

difficult to explain, unless the occurrence in the 21st century of the noun *zor* 'rush, hurry', which conveys the same meaning as *grabă* when used in these constructions, may be seen as an appropriate justification.

As for *nevoie*, its evolution is not surprising, since this noun has been attested only once in the MIHI EST construction, in a translated text from the 16th century, and it never occurs again in the MIHI EST construction, as indicated in Table 6.14, above. Indeed, in the following centuries, *nevoie* occurs exclusively with the HABEO construction in my dataset. Unless this unique occurrence is a result of an insufficiently documented period, it may point toward an imitation of a structure from the source text, and hence an accidental occurrence of this noun in the MIHI EST construction in Romanian.

In order to have a better overview of the data, I represent the evolution of the above commented nouns in the two constructions in a single graph (Figure 6.16). This two-axis column chart displays the competition between the HABEO and the MIHI EST construction for each of these nouns per historical period.
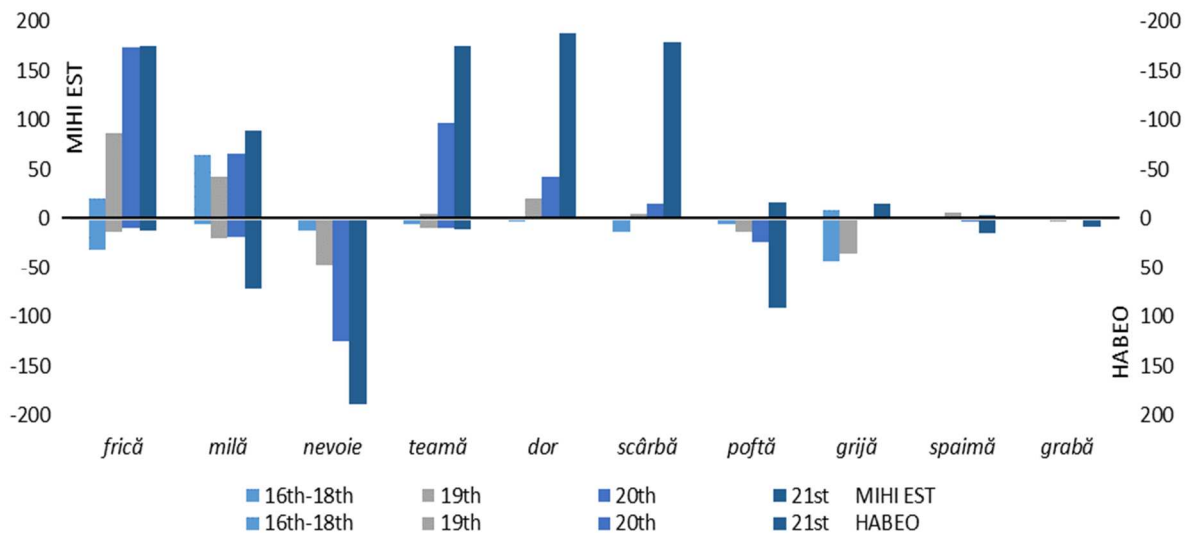
Figure 6.16 Case studies – competition between HABEO and MIHI EST

The graph clearly shows that, in the 16th-18th centuries, most of the nouns of my case study occur predominantly in the HABEO construction (the inversed bars), and become more frequent in the MIHI EST construction (the bars on the upper side of the graph) later on. Yet, the evolution of nouns such as, *poftă*, *grijă*, or *spaimă* indicates an opposite tendency: these nouns increase in frequency in the HABEO construction, and decrease in the MIHI EST construction. This raises the question of how many of the most frequent nouns in the 21st century have the HABEO construction as dominant construction. In order to obtain such a ranking, I calculate the relative frequency in present-day Romanian of all the nouns occurring in the two constructions taken together, and sort them by their total (cf. Table 6.15). In Figure 6.17, I have plotted the ten most frequent nouns in the 21st century.

Table 6.15   The ten most frequent nouns in the 21st century – HABEO vs. MIHI EST

Nouns	HABEO	%	MIHI EST	%	Total	%
chef 'mood'	196	2,8%	3	0,0%	199	2,8%
lene 'laziness'	6	0,1%	190	2,7%	196	2,8%
oroare 'horror'	184	2,6%	8	0,1%	192	2,7%
rușine 'shame'	4	0,1%	186	2,7%	190	2,7%
dor 'longing'	2	0,0%	188	2,7%	190	2,7%
nevoie 'need'	189	2,7%	0	0,0%	189	2,7%
frică 'fear'	12	0,2%	174	2,5%	186	2,7%
teamă 'fear'	11	0,2%	175	2,5%	186	2,7%
jenă 'embarrasment'	36	0,5%	150	2,1%	186	2,7%
sete 'thirst'	27	0,4%	153	2,2%	180	2,6%
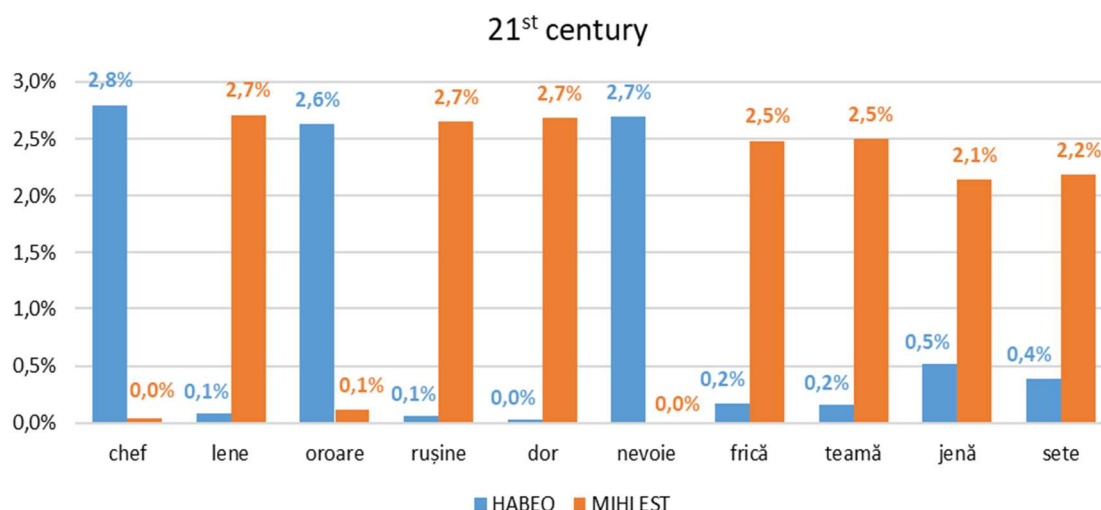
Figure 6.17 The ten most frequent nouns in the 21st century – HABEO vs. MIHI EST

Out of the ten nouns from my graph, seven have the MIHI EST as the dominant construction to express states, whereas three of them prefer the HABEO construction. One of these three nouns is *nevoie* 'need', which, as already mentioned, has been found in my dataset occurring only once in the MIHI EST construction, in a translated text and, hence, may be considered as an accidental occurrence, since it did not develop further in this construction. The other two nouns are *chef* 'mood' and *oroare* 'horror', which have both newly entered the MIHI EST construction in the 21st century. Since these two nouns just started occurring in the MIHI EST construction in the 21st century, *chef* counting three occurrences and *oroare* eight occurrences, their evolution is difficult to predict. Their high frequency in the HABEO construction corroborates my hypothesis that the state nouns first occur in the HABEO construction before shifting to the MIHI EST construction, during the first three historical periods of Romanian.

Based on the evolution of the state nouns described in this case study, I suggest that once a noun has been recruited by the MIHI EST construction, one of the following paths will be taken:

(i) The newly entered noun decreases (immediately or within a century) in frequency in the HABEO construction, or disappears completely from it, while it increases in frequency in the MIHI EST construction. This is the path followed by: *frică* 'fear', *teamă* 'fear', *dor* 'longing', *scârbă* 'disgust', and *grijă* 'worry, care';

(ii) The newly entered noun keeps occurring in both constructions throughout the centuries, with increasing frequencies in each of them. This is the case of: *milă*, 'pity', *poftă* 'craving', and *spaimă* 'fear';

(iii) The newly entered noun occurs for a period in both constructions, but it continues to increase in frequency in the HABEO construction while decreasing its presence in the MIHI EST construction. This evolution has been observed for *grabă* 'rush, hurry', and *nevoie* 'need'.

From these case studies, I conclude that the state nouns in my list, indeed, tend to first occur in the HABEO construction and then in the MIHI EST construction, with the only exception of the noun *milă* 'pity', which is first attested in my dataset in the MIHI EST construction. For some nouns, the change occurs very abruptly, whereas for other nouns it takes centuries. The change is for some nouns definitive, whereas other nouns continue to occur in both constructions.

This study also highlights that, although most of the nouns occurred first in the HABEO construction, they did not necessarily have deep roots in this construction when they were recruited by the MIHI EST. A possible explanation for this is that these nouns had recently entered in the language too. In contrast, some of the nouns entering the MIHI EST construction in the 21st century show a considerably high frequency in the HABEO construction, since they were borrowed in the language one century earlier (*oroare*, in the 20th century), or even two centuries earlier (*chef*, in the 19th century). Given that, it is difficult to predict whether the new nouns entering the MIHI EST construction in the 21st century will be able to abandon the HABEO construction, in which they seem rather deep-rooted. Note also that the nouns entering the MIHI EST construction in the 21st century and belonging to other semantic fields do not occur first in the HABEO construction, as opposed to the nouns coming from the semantic field of states and emotions. This is a sign that the MIHI EST construction is now able to recruit its own nouns, independent of their link with the HABEO construction, and coerces them into its psychological and physiological meaning.

In any case, a clear tendency can be observed throughout the centuries, namely the attraction force exercised by the MIHI EST construction on state nouns in Romanian, and especially on the state nouns occurring in the HABEO construction, causing it to become less preferred in expressing psychological or physiological states.

6.6   Conclusion

In this chapter, I have presented the methodology I applied in order to gather the list of nouns occurring in the MIHI EST construction. From the collected data, a number of 29 different nouns have been extracted for the pre-21ˢᵗ century Romanian, and 95 nouns have been found to occur in the MIHI EST construction in the present-day language.

Before observing more closely the set of selected nouns, I presented a few ambiguous cases of adjectives, adverbs and interjections, which can sometimes be analyzed as nouns too in Romanian. Because they are occasionally analyzed as nouns, these elements have been returned by my query in the corpus. Although they represent extremely interesting cases, I have filtered them out, due to the ambiguity in their analysis, considering them as valuable material for further research.

Across the centuries, the set of nouns turns out to be very dynamic, since with every historical period, new nouns enter this construction, while others are discontinued in it. In the case of the new nouns entering the MIHI EST construction, few of them are new borrowings or newly formed words through morphological derivation, whereas others have already existed in the language for some time. As for the nouns that no longer occur in the construction, most of them still occur in other constructions in the present-day language, whereas a small number have fallen into disuse.

Furthermore, I have observed the evolution of the competition between the MIHI EST and the HABEO constructions, with respect to the extracted nouns. This study shows a clear tendency for the MIHI EST construction to attract mostly nouns which already occur in the HABEO construction. I propose three possible evolutions of the recruited nouns with respect to their use in the two construction: (i) the noun increases in frequency in the MIHI EST construction while it disappears from the HABEO construction; (ii) the noun increases in frequency in both constructions, or (iii) the noun increases in frequency in the HABEO construction while decreasing in frequency in the MIHI EST construction. These conclusions are obtained based on the data I have been able to gather. A thorough study, based on a larger dataset, is required in order to validate these claims.

Chapter 7   Identifying the subject of the MIHI EST construction

In Chapter 3 of this dissertation, I have shown that accusative and dative experiencers occurring in psych-verb constructions behave like nominative subjects rather than as objects in canonical structures. Such behavior has been observed in many languages, where certain dative arguments, and even accusative and genitive marked arguments, take on several subject properties and, hence, behave like syntactic subjects.[84]

The hypothesis that Romanian has non-canonical subject marking is not new. Romanian has been argued to have quirky dative subjects by several scholars. However, this issue has mainly been discussed with respect to psychological verbs such as *plăcea* 'please' (cf. Rivero & Geber 2003; Rivero & Geber 2004; Geber 2006; Geber 2011; Alboiu 2007; Rivero 2009), and more rarely and less extensively with respect to the MIHI EST construction (Cornilescu 2009). The main reasoning for analyzing the dative argument as a non-canonical subject comes from its behavior in raising constructions (Rivero & Geber 2003, 2004; Geber 2006, 2011; Alboiu 2007, Rivero 2009, Cornilescu 2009), or more rarely word order (cf. Giurgea 2017: 307-308). In this chapter, I focus my attention first on the dative experiencer and then on the postverbal noun.[85]

7.1   The status of the dative experiencer

This section will investigate to what extent the dative experiencer in the MIHI EST construction displays subject properties. Obviously, it lacks the two main coding properties of the subject: case and verb agreement. In what follows, I examine the relevance and the outcome of the relevant subject tests described in the first part of this

---

[84] For Germanic languages (Icelandic, Faroese, Old Norse, Old and Modern German), see Zaenen, Maling & Thráinsson (1985), Greenberg & Franks (1991), Barðdal & Eythórsson (2003), Sigurðsson (2004), among many others; for Slavic (Russian, Slovenian, Bulgarian), see Greenberg & Franks (1991), Moore & Perlmutter (2000), Sigurðsson (2004), Rivero & Savchenko (2005), Fleisher (2006), Madariaga (2011); for Baltic languages (Lithuanian and Latvian), see Seržant (2013).

[85] This chapter is, partially, an altered version of a submitted article written together with my supervizor, Prof. Marleen Van Peteghem. For the purpose of consistency with the rest of the dissertation, I use in this chapter the first person singular, although the claims belong equally to both of us.

dissertation, and apply them to structures containing the MIHI EST construction. The following tests are investigated: word order (7.1.1), binding (7.1.2), control (7.1.3), raising (7.1.4), and deletion in telegraphic style (7.1.5), bare quantifiers in clause-initial position (7.1.6), and the ability to take secondary predicates (7.1.7).

7.1.1   Word order

As mentioned in Section 3.2.2, in Chapter 3, word order is considered one of the basic properties of subjects in Indo-European languages (cf. Keenan 1976). [86] However, investigating word order is not very simple, since it is not always clear which is the basic word order of a language. This is also the case in Romanian, which is argued to have a relatively free word order. The linear position of the subject shows indeed great freedom and is very sensitive to information structure.

It is generally assumed that Romanian displays two basic word order patterns, SV and VS, the former being the most frequent one, whereas the latter occurs with certain verbs and in rhematic sentences (Vasilescu 2013: 537), or when the sentence is considered not to have a topical subject (cf. Giurgea 2017). In generative studies, for instance, Romanian is argued to be a VS language. Certain scholars (cf. Dobrovie-Sorin 1987; Cornilescu 1997; Alboiu 2002) consider that, since there is no preverbal position dedicated to subjects, preverbal subjects must be seen as themes, or topics (unless focus-fronted). By contrast, other scholars (cf. Motapanyane 1994; Ştefănescu 1997; Hill 2002, and Dobrovie-Sorin 2011, cf. also Footnote 28) claim that preverbal subjects cannot be treated as topics or foci in certain cases, and that the preverbal position of subjects is an A-position.

Given the uncertainty created by this debate, Geber (2011: 90) considers word order inconclusive for identifying subjects in Romanian, especially non-canonical ones. One of her reasons is the fact that the dative experiencer is mostly expressed by a clitic pronoun, which, like all clitic pronouns in Romanian, is usually attached to the left of the verb, except for the imperatives, in which case the clitic is postposed to the verb.

Notwithstanding, with respect to certain languages the neutral word order of some core arguments in a specific construction coincides with their order in the argument structure of that construction. A case in point is German, where, as shown by Barðdal & Eythórsson (2003, 2018), the position of the subject in a neutral declarative sentence coincides with its position in the argument structure of the specific event. When the subject is defined as the leftmost argument of the argument structure (Eythórsson &

---

[86] As mentioned in Chapter 3, word order was considered initially a coding property, along with case and verb agreement (cf. Keenan 1976). Nevertheless, in recent studies, word order is argued to be a behavioral property (cf. Le Mair et al. 2017, and Barðdal et al. forthcoming)).

Barðdal 2005), the position of the arguments becomes crucial in differentiating between subject arguments and object arguments.

Moreover, it is often argued that, when the verb has two arguments, the subject is typically the first in the linear word order, because it is usually the most topical one. Furthermore, Giurgea (2017) supports a multifunctional position account in which topical and non-topical subjects occupy the same preverbal position (Giurgea 2017: 306). Since each of these contrasting arguments depends on the theory it originates from, only a thorough usage-based study can clarify which type of word order Romanian subjects show.

Such an in-depth study goes beyond the scope of this dissertation. Nonetheless, a smaller case study can already give us a glimpse of what we should expect from such an investigation. Therefore, in what follows, I describe a case study I carried out on the word order in Romanian, based on the dataset described in Chapter 6. In this case study, I compare the word order in the MIHI EST structure with the word order in the canonical Nom-Acc structure, which is what we find with a nominative experiencer. Note that some of the results concerning the structure with canonical subjects have already been described in Section 3.2.2, in Chapter 3, where I compared these results with the outcomes of the study presented by Van Peteghem (2016), on pain constructions with an accusative experiencer.

The present case study is based on a part of my dataset, more precisely on a total of 7 235 examples. Among these examples, 3 266 contain the canonical structure with a nominative subject and an accusative object, and 3 969 examples contain the MIHI EST construction, hence, an experiencer in the dative occurring with a state noun in the nominative. Following the example of other scholars,[87] I aim to compare in this study the word order of the experiencer in the dative, on the one hand, and of the state noun encoded in the nominative, on the other hand, with the word order of the subject and that of the object in canonical Nom-Acc structures. This comparison will cover the whole period between the 16th and the 21st centuries.

A first aspect to consider in this study is the existence of *pro*, i.e. an unrealized argument, in both canonical and non-canonical structures, as a result of the pro-drop character of Romanian. In my dataset, from the 3 266 examples with a nominative experiencer, 2 368 contain a *pro* nominative subject. As for the examples containing the MIHI EST construction, from the 3 969 examples, 3 546 do not contain doubled NPs or strong pronouns. Recall that, in this language, the dative experiencer can also occur as a full NP, or a strong pronoun, which must be doubled by a clitic pronoun, Romanian being a clitic doubling language. Therefore, the examples enclosing a clitic only are interpreted in this study as corresponding to the ones with *pro,* since the experiencer NP or the strong

---

[87] Cf. Le Mair et al. (2017) and Barðdal et al. forthcoming.

pronoun remain unrealized. Table 7.1 and Table 7.2 report on these figures for the nominative experiencer construction and for the MIHI EST construction, respectively.

Table 7.1   Realization of the experiencer argument: NP/ strong pronoun vs. *Pro* /Only clitics

Experiencer	*Pro*/ Only clitics	%	NP	%	Total
Nominative experiencer	2 368	72,5 %	898	27,5 %	3 266
MIHI EST (dative experiencer)	3 546	89,3 %	423	10,7 %	3 969

Table 7.2   Realization of the state noun: NP/ strong pronoun vs. Unrealized

State noun	Unrealized	%	Realized	%	Total
Nominative exp construction	0	0 %	3 266	100 %	3 266
MIHI EST construction	0	0 %	3 969	100 %	3 969

As evident from the tables above, the majority of the examples in the nominative experiencer construction contain *pro*, more precisely 72,5 %, and only 27,5 % have an overt experiencer subject. A comparable situation occurs in the examples containing the MIHI EST construction, where 89,3 % of the examples show no overt nominal experiencer, containing only the obligatory dative clitic. Observe also that in these examples, exactly as in the canonical structures with a nominative subject experiencer, it is the nominal experiencer that is left unexpressed in the majority of the cases, whereas the state noun is always present.

Given that clitics have a fixed position, being attached to the verb, and that they tend to be considered in Romanian as case markers (cf. Manoliu-Manea 1987), examining the nominal experiencer may yield a better insight into the preferred position of this argument in the clause. Therefore, for the purpose of this study, I focus on MIHI EST structures with doubled NPs or strong pronouns and on structures containing an overt subject in the nominative experiencer construction.

Depending on information structure, in Romanian canonical structures place their nominative experiencer subjects either preverbally, as in (187a), or postverbally, as in (187b).

(187) a. *Pe   timpul   lui,   românii   au   făcut foame   şi   frig*
         in   time.the   his.POSS   Romanians.the   have.3PL   made hunger   and   cold
         'In his time, Romanians forbore hunger and cold   (elenaudrea.ro)
      b. *Fiindcă   aşa   are   chef   guvernul*
         because   so   has   will.ACC   government.the
         'Because this is the will of the government'   (hotnews.ro)

The same holds for structures containing the MIHI EST construction. The experiencer NP or strong pronoun may either occur in clause initial (188a) or postverbal (188b) position, or more rarely between the verb and the postverbal state noun (188c), in which case it is also postverbal.

(188) a. *Tatii*   *i-*   *era*   *groază*   *de*   *apă*
          father.DAT   him.DAT=   was   terror   of   water
          'My father was terrified by the water'   (A. Vasile, *Despot Voda*, 1850)
       b. *I-*   *a*   *fost*   *foame*   *fetiţei*
          her.DAT   has   been   hunger   girl.the.DAT
          'The girl has been hungry'   (laptematern.ro)
       c. *Îi*   *era*   *omului*   *sete*
          him.DAT   was   man.the.DAT   thirst
          'The man was thirsty'   (M. Caragiale, *Craii de Curtea Veche*, 1929)

Table 7.3 and Table 7.4 illustrate these preferences in my dataset for nominative experiencer construction and the MIHI EST construction, respectively.

Table 7.3   The nominative experiencer construction – word order of the realized arguments

	Preverbal	%	Postverbal	%	Total
Experiencer (nom)	**846**	**94,2 %**	**52**	**5,8 %**	**898**
State noun (acc)	161	4,9 %	3 105	95,1 %	3 266

Table 7.4   The MIHI EST construction – word order of the realized arguments

	Preverbal	%	Postverbal	%	Total
Experiencer NP (dat)	**358**	**84,6%**	**65**	**15,4%**	**898**
State noun (nom)	71	1,8%	3 898	98,2%	3 266

The bar charts in these figures clearly show that, contrary to the state noun, in both constructions, the nominal experiencer tends to be preverbal (94,2 % in the nominative construction and 84,6 % in the MIHI EST construction), and only in an insignificant percentage postverbal (5,8 % in the nominative construction and 15,4 % in the MIHI EST construction). The graph in Figure 7.1 illustrates this similarity of the behavior of the two constructions.
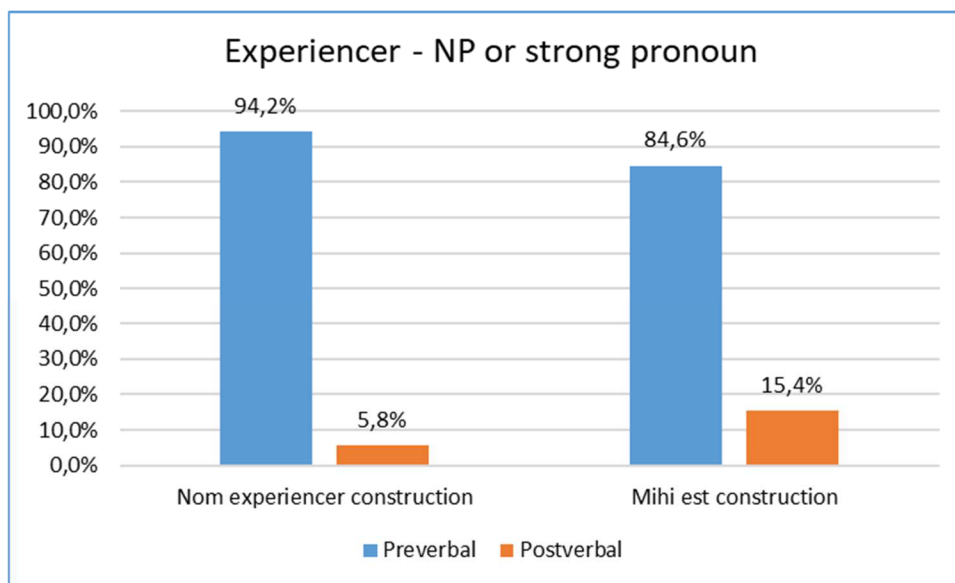
Figure 7.1   Nominativ vs. dative experiencers – preferred position in the clause

Pană Dindelegan (2016: 112) mentions that the pre-21st-century Romanian is characterized by more freedom with respect to word order, whereas the word order becomes more restricted in the 21st century. In order to get more insight into this claim, I have carried out a diachronic study of the two constructions. In this study, I first identified the proportion between the examples containing *pro* and those with a realized subject in the nominative experiencer construction, as well as the examples with nominal experiencers, with or without clitic doubling, in the MIHI EST construction. Second, I explored which position in the clause each of the two main arguments prefer in these constructions. Table 7.5 illustrates, in a diachronic perspective, the proportions between structures containing only a clitic, or a *pro* experiencer, and structures in which the experiencer is realized by an NP or a strong pronoun. These facts are illustrated also in the graphs in Figure 7.2 and Figure 7.3.

Table 7.5   The experiencer argument (16th-21st) – overt vs. *pro* /only clitic

		16th-18th	19th	20th	21st
Nom-Acc	NP experiencer	80 (27,3 %)	69 (33,2 %)	87 (28,3 %)	662 (26,9 %)
	Pro	213 (72,7 %)	139 (66,8 %)	220 (71,7 %)	1 796 (73,1 %)
	Total	**293 (100 %)**	**208 (100 %)**	**307 (100 %)**	**2 458 (100 %)**
Dat-Nom	NP experiencer	30 (18,9 %)	30 (9,0 %)	71 (10,3 %)	292 (10,5 %)
	Only clitic experiencer	129 (81,1 %)	305 (91,0 %)	620 (89,7 %)	2 492 (89,5 %)
	Total	**159 (100 %)**	**335 (100 %)**	**691 (100 %)**	**2 784 (100 %)**

Nominative experiencer construction
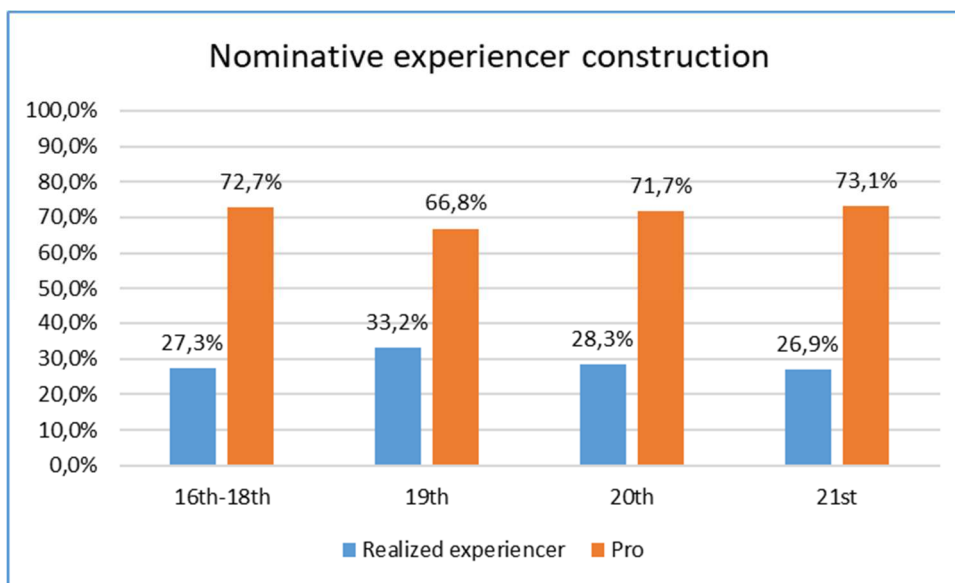
Figure 7.2   The nominative experiencer (16th-21st) – Overt vs. *Pro*

Mihi est construction
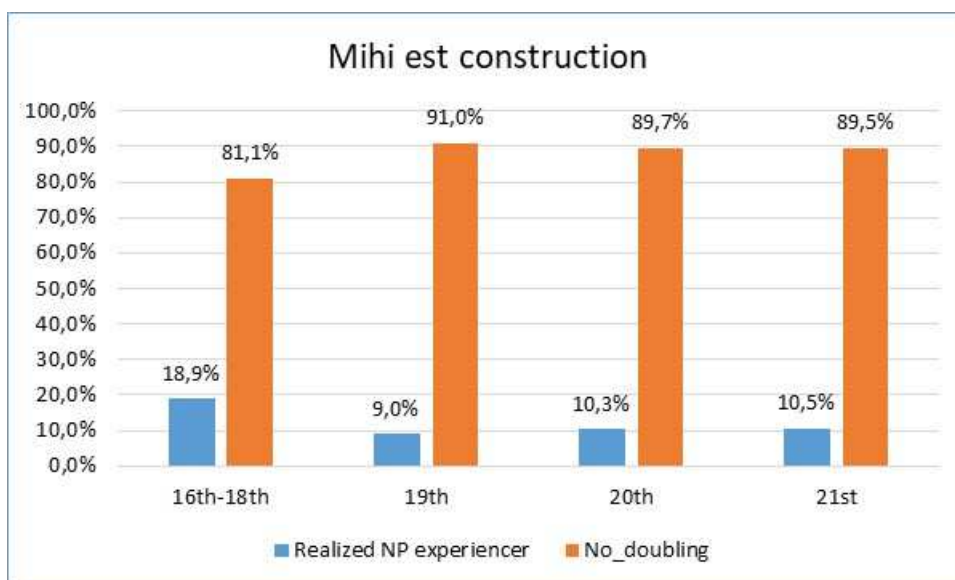
Figure 7.3   The dative experiencer (16th-21st) – Overt vs. No doubling

The two graphs above reveal a certain stability across the centuries in the proportion between canonical structures with *pro* and structures with a realized subject on the one hand, and between non-canonical structures with nominal dative experiencers with and without clitic doubling, on the other hand.

As for the position in the clause, preferred by the arguments of these constructions, this section presents only the outcomes concerning the experiencer argument in both structures, i.e., the nominative experiencer and the dative experiencer. The results regarding the state noun will be presented in Section 7.2.5, where the preferred position of the state noun is discussed. Consequently, Table 7.6 presents the situation throughout the centuries for the preferred word order of the overt experiencer, in both the

nominative experiencer construction and the MIHI EST construction. These facts are visualized in Figure 7.4 and Figure 7.5, respectively.

Table 7.6   The experiencer argument (16th-21st) – preverbal vs. postverbal

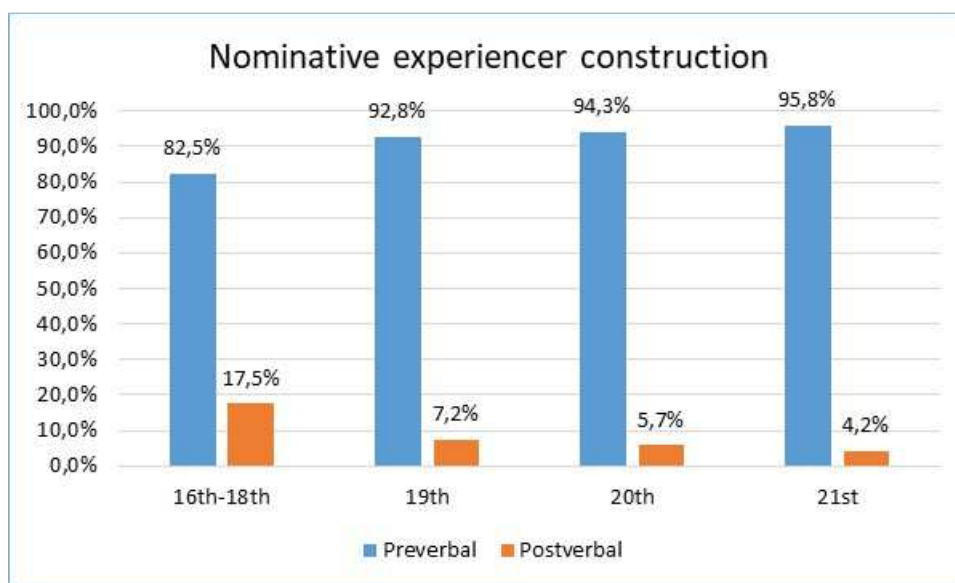|		16th-18th	19th	20th	21st
Nom-Acc	Preverbal	66 (82,5 %)	64 (92,8 %)	82 (94,3 %)	634 (95,8 %)
	Postverbal	14 (17,5 %)	5 (7,2 %)	5 (5,7 %)	28 (4,2 %)
	Total	**80 (100 %)**	**69 (100 %)**	**87 (100 %)**	**662 (100 %)**
Dat-Nom	Preverbal	20 (66,7 %)	23 (76,7 %)	62 (87,3 %)	253 (86,6 %)
	Postverbal	10 (33,3 %)	7 (23,3 %)	9 (12,7 %)	39 (13,4 %)
	Total	**30 (100 %)**	**30 (100 %)**	**71 (100 %)**	**292 (100 %)**

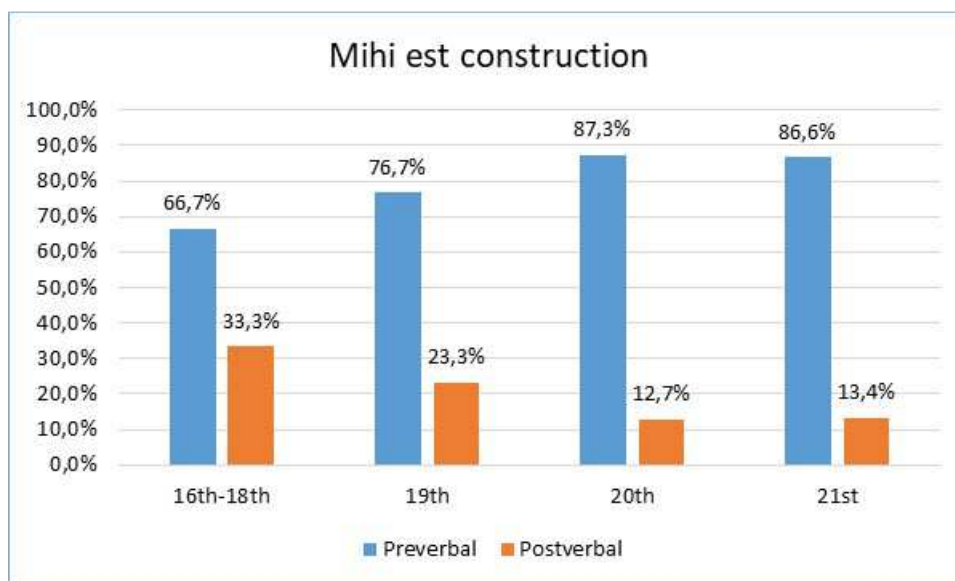Figure 7.4   The experiencer argument (16th - 21st) – preverbal vs. postverbal

Figure 7.5   The dative experiencer (16th-21st) – preverbal vs. postverbal

The bar charts in Figure 7.4 and Figure 7.5 above show that both constructions evolve toward a more constrained word order, that is, the preverbal position, which was already preferred, becomes even more frequent than the postverbal position. Observe that, in the case of the the nominative experiencer construction, this tendency is stronger than in the case of the MIHI EST construction. Although not in the same proportion as in the present-day language, this increased tendency toward the preverbal position of the experiencer existed already in the 16th-century Romanian.

In the view of what has been presented above, it is clear that the dative experiencer in the MIHI EST construction follows exactly the same behavioral pattern as the subject experiencer encoded in the nominative with respect to word order. In other words, the realized dative experiencer in the MIHI EST construction is being left unexpressed and leaves behind only the obligatory clitic in 72,5 % of the occurrences. This behavior is similar to that of the experiencer subject in the nominative, which is being left unexpressed in 89,3 % of the instances. These proportions remain more or less stable throughout the centuries. In addition, the dative experiencer, when realized by a nominal, occurs in preverbal position (84,6%) rather than in postverbal position, and so does the overt nominative experiencer (98,2%) in the canonical structure.

These facts represent remarkable evidence that Romanian tends to show SV word order, indeed supporting the position taken by Pană Dindelegan (2016: 112) and Giurgea (2017), and confirming that the similarities between Romanian and other SV languages (e.g. other Romance languages) are greater than the differences. This can be confirmed, based on the present study, for the structures having an experiencer as a subject, either canonical, or non-canonical. As for the contrasting position taken by other scholars (Dobrovie-Sorin 1993; Cornilescu 2000; Alboiu 2002), which argue that Romanian is a VS language this may possibly apply to other structures in Romanian which do not contain an experiencer as a subject. This is, however, material for further research.

The results presented above show that word order is an important diagnostic for subjecthood and that it aids in identifying the subject in Romanian, a subject which, as has been shown, prefers the clause-initial position.

It is generally known, however, that the preverbal position is typically associated with topics or foci. Nevertheless, besides the fact that the position of topics coincides with the preferred position of the canonical subject in the nominative experiencer constructions, one may argue that these facts corroborate the present claims, as is also shown in (Ilioaia & Van Peteghem, in prep.). These data show that the dative realized by a nominal experiencer in the MIHI EST construction tends to be topical, which is also a typical subject property (cf. also Giurgea 2017, among others).

Moreover, the following examples provided by Geber (2011: 91) confirm that the preverbal dative experiencer occurs in subject position and not in a focus position. Contrary to recipient datives (189b), the dative experiencer may contain or be realized by a quantifier (189a), which is evidence that it is not left-dislocated, as left dislocation is

impossible with quantifiers. Hence, the dative in this construction necessarily occupies the subject position, just like the nominative *nimeni* 'nobody' does in (189c).

(189) a. *Nimănui*   *nu*   *îi*   *place*   *muzica*
          nobody.DAT   not   him.DAT   pleases   music.the
          'Nobody likes music'

       b. **Nimănui*   *comitetul*   *nu*   *i-*   *a*   *dat*   *un*   *premiu*
          nobody.DAT   committee.the   not   him.DAT=   has   given   a   prize

       c. *Nimeni*   *n-*   *a*   *spus*   *adevărul*
          nobody.NOM   not   has   said   truth.the
          'Nobody said the truth'

Given these clarifications, the following conclusions can be stated. First, the study of word order with canonical nominative experiencer construction confirms that Romanian does not behave differently from the other Romance languages, showing SV word order in structures with an overt experiencer subject. Second, in the case of canonical structures containing an experiencer in subject position, the situation remains relatively stable between the 16th and the 21st centuries, but shows a tendency toward more precision in the dominant word order (the gap between preverbal and postverbal positions increases). Third, the comparison between the MIHI EST construction and the canonical nominative experiencer construction reveals that the dative experiencer patterns with the nominative experiencer with respect to their preference for preverbal position and also with respect to the argument which can be left unexpressed. Finally, this case study on word order confirms its validity as a subject test, since it can easily differentiate between subjects and objects in Romanian, in canonical and non-canonical structures containing an experiencer.

7.1.2   Binding

One of the most frequently used diagnostics for detecting the subject status of non-canonical subjects is binding of reflexives. However, as highlighted in Section 3.2.3, in Romanian, as well as in other languages (cf. Icelandic, Latin, and Greek), cases of object binding also exist. Whereas in other languages a strategy has been identified to distinguish between subjects and objects using the test of reflexive binding, in Romanian the relevance of this test is severely reduced, as argued in Section 3.2.3, in Chapter 3, since binding does not tease apart subjects and objects, as a result of a less strict applicability of binding rules, in this language. With other words, not only reflexives but personal pronouns too may be bound in structures where, in other languages, they are complementary.

In spite of this difficulty, Geber (2011: 90) provides evidence in favor of the reliability of the binding test in distinguishing between subjects and objects, with respect to constructions containing psych verbs such as *plăcea* 'please'.

(190) a. *Fiecărui   copil_i   îi   place   jucăria   sa*
      each.DAT   child   him.DAT   pleases   toy.the   his.POSS
      'Each child likes his toy.'
    b. **Fiecare   jucărie_i   îi   place   copilului   său_i*
       each   toy   him.DAT   pleases   child.the.DAT   his.POSS

Geber (2011) argues that the dative experiencers in these constructions behave like subjects. Indeed, in example (190) above, it is the dative experiencer, that controls the reflexive, and not the postverbal noun *jucăria* 'toy'. This also holds for the dative experiencer of the MIHI EST construction, as shown in (191a), where the reflexive pronoun *sine* is bound by the dative weak pronoun *îi*.

(191) a. *Îi_i   era   silă_j   de   sine_{i/*j}   (însuşi)*
      him.DAT   was   disgust   of   himself   himself
      'He was feeling disgust for himself'   (L. Rebreanu, Padurea spânzuraţilor, 1922)
    b. *Mara_i spune că   băiatului_j   îi   era   silă_k   de   sine_{i/j/*k}*
      Mara   says   that   boy.the.DAT   him.DAT   was   disgust   of   himself/herself
      'Mara says that the boy was feeling disgust for himself /her'

Furthermore, the reflexive pronoun is coreferential with the dative experiencer, in (191b), which occurs in the same clause, but also with Mara, which it is not situated in its clause, hence showing that *sine* 'himself' occurs in long distance reflexivization too.

However, I have shown in Section 3.2.3, in Chapter 3, that objects and subjects have similar behavior in binding contexts, including instances of reflexive, possessive or reciprocal binding, and even of quantifier binding (cf. 64b, from Section 3.2.3, *supra*, repeated here for convenience, under 192).

(192) *A   dat   fiecărui_i   student   propriul   său_i   exemplar*
      has   given   each.DAT   student   own.the   his.POSS   copy
    'S/He gave each student his own copy'

To conclude, dative experiencers in the MIHI EST construction behave like subjects in that they may bind anaphors. However, the binding test is less reliable, since recipient datives may also bind, to the same extent, reflexives and personal pronouns.

7.1.3   Control of the unexpressed subject of non-finite verb forms

Other fundamental subjecthood tests are related to control of the unexpressed subject of non-finite clauses, mostly infinitives (cf. Icelandic, see Sigurðsson 1992, 2004; Barðdal 2002, 2006; Barðdal & Eythórsson 2003, 2012; Eythórsson & Barðdal 2005), but also gerunds (cf. Russian, see Moore & Perlmutter 2000). As shown in Section 3.2.4, in Romanian both of these non-finite forms occur in control structures, as well as an additional one, the supine. Moreover, recent analyses point toward instances of control also with subjunctive clauses (cf. Section 3.2.4, in Chapter 3).

The two ways in which these tests are used will be discussed successively in what follows: (i) only subjects of non-finite verb forms can be controlled by an argument of the main clause (7.1.3.1); (ii) in certain syntactic contexts, only subjects may control PRO (7.1.3.2).

7.1.3.1   Only subjects may be a controlled PRO

Zaenen, Maling & Thráinsson (1985: 454-455) and many other scholars (Sigurðsson 1992, 2004; Barðdal 2000, 2002, 2006; Barðdal & Eythórsson 2003, 2012, Eythórsson & Barðdal 2005) have shown that in Icelandic, the unexpressed dative of the embedded infinitive clause may be controlled by the nominative subject of the main clause (193a). It may also be coreferential with an arbitrary controller (193b).

(193)   Icelandic

a. *Ég$_i$*   *vonast*   *til*   ___$_i$ *að*   *vera*   *hjálpa*
   I   hope   for   PRO.DAT   INF   be   helped
   'I hope to be helped'   (Barðdal 2002: 71, ex. 21a)

b. *Það*   *er*   *ekki*   *gott*   ___   *að*   *verða*   *óglatt*
   it   is   not   good   PRO.DAT   INF   become   queasy
   'It is not good to feel sick.'   (Barðdal & Eythórsson 2018: 262, ex. 3a)

In Section 3.2.4.2.1, I have shown that this test is reliable with respect to Romanian canonical subjects and that it applies also to accusative and dative experiencers in certain psychological constructions. Although these experiencers must occur as clitics and cannot be implicit, it is clear that they are controlled by the subject of the verb of the main clause with obligatory subject control. This represents compelling evidence of their subject behavior. The examples in (194)-(195) illustrate how this test applies to Romanian structures containing the MIHI EST construction in infinitive complement clauses (194a-b), and in subjunctive complement clauses (195a-b).[88]

---

[88] Given the progressive replacement of complement infinitives with subjunctives, the acceptability of certain examples with control infinitives or subjunctives may vary among native speakers of Romanian. These examples are preceded by a "?", in the text.

(194) a. ? *Clara_i*   *a*   *reuşit*   *a*   *nu*   ***(-i_i)**   *fi*   *frică /*

Clara   has   managed   INF   not   =her.DAT   be   fear /

***a**   *nu*   *fi*   *ei_i /*   *Clarei_i*   *frică*

INF   not   be   her.DAT /   Clara.DAT   fear

'Clara managed not to feel afraid'

     b. *Nu*   *e*   *bine*   *a*   **(-ţi)*   *fi*   *frică*   *din*   *orice*

not   is   good   INF   you.DAT   be   fear   from   anything

'It is not good to be afraid of everything'

(195) a. ? *Clara_i*   *a*   *reuşit*   *să*   *nu*   ***(-i_i)**   *fie*   *frică /*

Clara   has   managed   SUBJ   not   =her.DAT   be.3SG   fear /

***să**   *nu*   *fie*   *ei_i /*   *Clarei_i*   *frică*

SUBJ   not   be.3SG   her.DAT /   Clara.DAT   fear

'Clara managed not to feel afraid'

     b. *Nu*   *e*   *bine*   *să*   **(-ţi)*   *fie*   *frică*   *din*   *orice*

not   is   good   SUBJ   you.DAT   be.3SG   fear   from   anything

'It is not good to be afraid of everything'

These examples show that the dative experiencer of a MIHI EST construction patterns with canonical subjects in that it is controlled by the subject of the verb of the main clause when this is an obligatory subject control verb. Additional evidence that the dative experiencer is a controlled clitic is that it cannot be realized by an NP or a strong pronoun.

Recall that the omission of the dative clitic in non-finite clauses is not possible in Romanian when the clitics occur in experiencer constructions with a specific reading, as in (194)-(195). My dataset does not attest any case with clitic omission, and the examples above are considered ungrammatical by native speakers of Romanian without the dative clitic.[89] As explained in Section 3.2.4, in Chapter 3, the ungrammaticality of the examples above with clitic omission can be explained by the obligatoriness of clitics in specific experiencer constructions, in Romanian (cf. Geber 2011: 72). Dative clitics even co-occur obligatorily with dative NPs in clitic doubling constructions, reason for which they have been analyzed as case markers, or as the 'spell-out' of certain morpho-syntactic features of the verb (cf. Manoliu-Manea 1990; Monachesi 2000).

Nonetheless, the obligatoriness of clitics does not exclude the subject behavior of dative experiencers in Romanian, since they are obligatorily controlled by the subject of the main clause, nor does it contradict the evidence that these structures are obligatory control structures. Therefore, I argue that the control test that focuses on the oblique PRO obligatorily controlled by a subject is relevant and conclusive in Romanian, in spite of the obligatory presence of the dative clitic.

---

[89] It has to be noted that cases of the MIHI EST construction with clitic omission may also be possible in Romanian, but could not be retrieved in my dataset, most likely due to the very specific queries I created in order to limit noise. It is well-known that clitic omission was possible in certain experiencer constructions with psych verbs in pre-21st century Romanian, when the use of clitics was less constrained than in present-day Romanian (cf, among others, Rivero 2009).

7.1.3.2 Only subjects may control PRO

The second way of using control into non-finite clauses as a subjecthood test relies on configurations in which the dative experiencer is in the position of the controller. In languages such as French (cf. Legendre 1989), German (cf. Haspelmath 2001a), and Russian (cf. Moore & Perlmutter 2000), this test is used especially in adjunct infinitive or gerund clauses. In order for this test to be reliable, the unexpressed subject of the adjunct clause must be controlled by the oblique subject of the main clause.

As for Romanian, this test is used by Geber (2011: 91) in subjunctive clauses occurring as complements of psychological verbs as in (196a-b) and by Cornilescu (2009: 204) in adjunct infinitive clauses, as in (197).

(196) a. *(Lui Ion$_i$)* *îi$_i$*   place   să   danseze   ___$_i$
 DAT John him.DAT pleases SUBJ dance.3SG PRO
 '(John) / {S/he} likes to dance.'   (Geber 2011: 91, ex. 42)

 b. *(Lui   Ion$_i$)* *i$_i$* se   cuvine   să   câştige   ___$_i$
 DAT John him.DAT SE   ought.to SUBJ win.3SG PRO
 '(John) / {S/he} deserves to win'   (Geber 2011: 91, ex. 43)

(197) *Mi$_i$-*   e   destul   de   foame   pentru   a   mânca   ___$_i$
 me.DAT= is enough   of   hunger   for   INF eat   PRO
 toate   prăjiturile
 all   cakes.the
 'I am hungry enough to eat all cakes'   (Cornilescu 2009: 204)

Indeed, in configurations with a MIHI EST construction in the main clause, the dative experiencer controls the PRO of the adjunct clause, be it a prepositional infinitive (198a), a gerund (198b), or even a supine clause (198c).

(198) a. *Îmi$_i$*   era   teamă   **înainte**   **de**   **a**   intra   ___$_{i/*k}$
 me.DAT was fear   before   of   INF enter   PRO
 în   cabinet
 in   consultation_room
 'I was scared before I went into the consultation room'   (la-psiholog.ro)

 b. *De   ce   **mi$_i$-***   e   frig   ___$_{i/*k}$ **gândind**   la   tine?*
 of   what me.DAT is   cold   PRO   thinking   at   you
 'Why do I feel cold when I think of you?'   (zang.ro)

 c. *După   accident,   Roxanei$_i$ i$_i$-   era   groază   **de**   condus   ___$_{i/*k}$*
 after accident   Roxana.DAT   her.DAT= was terror   SUP driven   PRO
 'After accident, Roxana was terrified of driving'

The fact that the dative experiencer of the MIHI EST construction in these examples is the only possible controller of the PRO in the adjunct clauses constitutes strong evidence in favor of its subject status, and, hence, confirms the validity of the control test which focusses on the oblique experiencer as a controller of PRO.

7.1.4 Raising

In recent studies, Romanian has been claimed to have non-canonical dative subjects based on the subject-to-subject raising test (Rivero & Geber 2003, 2004; Geber 2006, 2011; Alboiu 2007; Rivero 2009; Cornilescu 2009). However, whereas in other languages subjects can raise out of infinitive clauses, this phenomenon is very infrequent in Romanian, because of the rareness of the infinitive clause (cf. Section 3.2.4.1). Nonetheless, in addition to the rare cases of raising from infinitive clauses, Romanian has been shown to allow the subject to raise out of subjunctive clauses (199a) (Rivero & Geber 2003, 2004; Geber 2006, 2011; Alboiu 2007; and Rivero 2009) and even out of indicative clauses (199b) (Geber 2011).[90]

(199) a. *Victor   poate /   trebuie /   începe   să   deseneze   foarte bine*
        Victor   can /   must /   begins   SUBJ   draw.3SG   very   well
        'Victor can / must / is beginning to draw very well'
    b. *Băieţii   par   că   au   călătorit   ieri*
        Boys.the   seem.3Pl   that   have.3Pl   traveled   yesterday
        'The boys seem to have traveled yesterday.' (Geber 2011:16, ex. 7)

It has been argued that the dative experiencer of psychological verbs can also raise, which is considered evidence of its subject status (Geber 2011: 73-75, among other scholars). However, it has to be noted that only dative NPs may raise, but not dative clitics. Clitics must remain in situ, attached to the embedded verb (200a-b), in specific experiencer constructions. Hence, subject-to-subject raising may only be observed in examples with clitic doubling of an NP or a strong pronoun, as in (200c) (cf. Geber 2011: 73-75).

(200) a. *Pare   să   -i   placă   muzica*
        seems   SUBJ   =him.DAT   please.3SG   music.the
        'S/he seems to like music'
    b. **îi   pare   să   placă   muzica*
        him.DAT   seem.3PL   SUBJ   please.3SG   music.the
    c. *Lui   Ion   pare   să   -i   placă   muzica*
        DAT   Ion   seems   SUBJ   =him.DAT   please.3SG   music.the
        'Ion seems to like music'

My dataset on the MIHI EST does not contain any example of subject-to-subject raising with nominal experiencers. However, examples that sound perfectly natural to native speakers can be found on the web instantiating raising both out of infinitive (201a) and of subjunctive (201b) clauses.

---

[90] Geber (2011: 14-16) has shown that Romanian patterns with Portugese in allowing the subject to raise out of indicative clauses, and it patterns with Greek when it allows the subject to raise out of subjunctives.

(201) a. *Micuților   nu   pare   a   le   fi   frică   de   el*
little.the.DAT   not   seems   INF   them.DAT   be   fear   of   him
'The little ones don't seem to be afraid of him'   (giurgiu-net.ro)

b. *Dă   -i   biberonul [...]   de   fiecare   dată   când*
give.IMPER.2SG =him.DAT baby_bottle.the   at   each   time   when
bebeluşului   pare   să   -i   fie   foame
baby.the.DAT   seems   SUBJ   =him.DAT   be.3SG   hunger
'Give him the baby bottle whenever the baby seems to be hungry'   (suntmamica.ro)

In order to eliminate a possible interpretation of these experiencers as involving focus, consider the following examples, where the dative experiencer in MIHI EST construction may be realized by a quantifier (202a), contrary to recipient datives (202b). Since quantifiers may not occupy a focus position, examples of this type exclude a focus analysis.

(202) a. *Nimănui   nu   pare   să   -i   fie   frică*
nobody.DAT   not   seems   SUBJ   =him.DAT   be.3sg   fear
'Nobody seems to be afraid'

b. **Nimănui   nu   pare   să   -i   fi   dat   un   cadou*
nobody.DAT not   seems   SUBJ   =him.DAT   be.3SG   given   a   gift

c. *Nimeni   nu   pare   să   înţeleagă   ceva*
nobody.NOM   not   seems   SUBJ   understand.3SG   something
'Nobody seems to understand something'

These examples show that the dative experiencer in the MIHI EST construction necessarily occupies the subject position, just like the nominative *nimeni* 'nobody' does in the canonical structure in (202c) above.

Hence, the examples given in this section can be considered strong evidence that the Romanian MIHI EST construction allows raising of the dative NP or strong pronoun experiencers, on the only condition of being doubled by a clitic, endorsing the subject status of the dative experiencer.

7.1.5   Deletion of the subject in telegraphic style

As shown in Section 3.2.8, the test of deletion in telegraphic style, applies to nominative subjects as well as to dative experiencers of psych verbs in Romanian, exactly like in Icelandic (Barðdal 2006: 55), without resulting in an ungrammatical structure.

The same can be claimed about the dative experiencer of a MIHI EST construction, which, in telegraphic style, would leave unexpressed its experiencer rather than the state noun, as illustrated in (203a-b):

(203) a. *Mi-   a   fost   foame*
me.DAT=   has   been   hunger
'I was hungry'

    b. *Fost*   **(foame)*
      been   hunger
      'I was hungry'

(204) a. *Am*   *trecut*   *examenul*
      have.1sg   passed   exam.the
      'I have passed the exam'

    b. *Trecut*   **(examenul)*
      passed   exam.the
      'I have passed the exam'

Observe that the dative experiencer in (203) behaves like the nominative subject of the canonical structure in (204), in that it is the experiencer and not the state noun that is left unexpressed in the telegraphic style. Moreover, the omission of the postverbal noun would make the understanding of the message much more difficult, if not impossible, in both the canonical and the non-canonical structure, as this would, basically, mean that the predicate has been left unexpressed. This constitutes strong evidence that the dative experiencer in the MIHI EST construction behaves like a subject, and also that the test of deletion of the subject in telegraphic style is relevant in Romanian.

7.1.6   Bare quantifiers in clause-initial position

The test of bare quantifiers in clause-initial position is applicable also to experiencers occurring in the MIHI EST construction. This test has been shown to represent evidence in favor of the subject status of accusative and dative experiencers in Romanian constructions with psych verbs by Rivero (2009, based on Dumitrescu & Masullo 1996). Given that quantifiers cannot occur in a focus position (cf. Rizzi 1982, 1986; Cardinaletti 2004; Geber 2011) but are very suitable in subject position (Rizzi 2005: 211), the following examples can only be analyzed as containing dative subjects. Indeed, the MIHI EST structure in (205) contains a dative experiencer realized by a bare quantifier, and can only have a neutral reading. In contrast, the fronted dative goal in (206) cannot be realized by a bare quantifier, the configuration being, hence, ungrammatical.

(205) *Nimănui*   *nu*   *-i*   *este*   *dor*   *de*   *şcoală*
      nobody.DAT   not   =him.DAT   is   longing   of   school
      'Nobody likes the book'

(206) ***Nimănui**   *nu*   *-i*   *a*   *dat*   *un*   *cadou*
      nobody.DAT   not   =him.DAT   has   given   a   gift

(207) **Nimeni**   *nu*   *înţelege*   *nimic*
      nobody   not   understands   nothing
      'Nobody understands anything'

The example in (205) shows that the dative experiencer realized by the bare quantifier *nimănui* 'to nobody', in the MIHI EST construction, patterns with the canonical subject

nimeni 'nobody' in (207). The grammaticality of the structure in (205) and its neutral reading clearly constitute evidence in favor of a subject analysis of the dative experiencer in the MIHI EST construction. Therefore, the test of bare quantifiers in clause-initial postion is a relevant subject test in Romanian.

7.1.7 Secondary predication

The possibility of an argument to take a depictive, secondary predicate represents one of the criteria in identifying a subject and has been shown (cf. Section 3.2.10, in Chapter 3) to be relevant with respect to Romanian accusative and dative experiencers in psych constructions. As evident from the examples (208a) and (209a) below, configurations containing the MIHI EST construction and a secondary predicate in agreement with its dative experiencer are entirely grammatical and very natural.

(208) a. *(Fetei)*   *Îi*   *era*   *frică*   ***pierdută***   *în*   *noapte*
       girl.DAT.FEM.SG   her.DAT   was   fear   lost.FEM.SG   in   night
       'The girl was afraid while lost in the dark'

      b. **Fetei*   *i-*   *a*   *arătat*   *drumul*   ***pierdută***   *în*   *noapte*
       girl.DAT.FEM.SG   her.DAT=   has   shown   way.the   lost.FEM.SG   in   night

(209) a. *Copiilor*   *le*   *era*   *frig*   ***îmbrăcaţi***   *subţire*
       children.the.DAT.MASC.PL   them.dat   was   cold   dressed.MASC.PL   thinly
       'Children had cold while insufficiently clothed'

      b. **Copiilor*   *le-*   *a*   *dat*   *haine*   ***îmbrăcaţi***
       children.the.DAT.MASC.PL   them.DAT=   has   given   clothes   dressed.MASC.PL
       subţire
       thinly

(210) *Fata*   *a*   *sunat*   *la*   *poliţie*   ***speriată***
      Girl.the.FEM.SG   has   called   at   police   scared. FEM.SG
      'The girl called the police while scared'

Indeed, the examples in (208a) and (209a), which contain secondary predicates bearing on the dative experiencer, pattern with (210). In this last example, the secondary predicate *speriată* 'scared' agrees with the canonical subject *fata* 'girl', in gender and number. The dative experiencers in the (a) sentences above occupy a higher position in the structure and contrast with the beneficiary dative in (208b) and the goal dative in (209b), which occupy a lower position in the structure and are, hence, unable to take secondary predicates. Based on this evidence, I claim that the ability of the dative experiencer to take secondary predicates represents a relevant criterion for distinguishing between dative subjects and other datives, such as beneficiaries or goals, in Romanian.

7.2 The status of the state noun

This section focuses on the status of the state noun in the MIHI EST construction. The state noun shows default nominative morphology, therefore, it is generally analyzed as a postverbal subject (Pană Dindelegan 2013a: 186) in Romanian grammars. At the same time, the construction is described as being impersonal (Cornilescu 2009: 236, Pană Dindelegan 2013a: 106-107). I argue that the state noun behaves like a predicate already in the 16th century, and that its behavior has become increasingly predicate-like. My claim is based on the following arguments: lack of determination of the state noun (7.2.1), its modification (7.2.2), its complementation (7.2.3), and word order (7.2.5).

7.2.1 Determination

It is generally acknowledged that subjects tend to be definite and referential (cf. Givón 1976: 154). This is also the case in Romanian, even though, like most other languages, Romanian allows definite as well as indefinite NPs in subject position. As shown in Table 7.7, more than 99 % of the state nouns occurring in the MIHI EST pattern in my dataset are used without an article, and only 0,7 % with an article.

Table 7.7   Determination of the state noun in the MIHI EST pattern

Period	Determined N	%	Bare N	%
16th–18th	0	0 %	159	100 %
19th	0	0 %	335	100 %
20th	4	0,6 %	687	99,4 %
21st	22	0,8 %	2 762	99,2 %
Total	26	0,7 %	3 943	99,3 %

When present, the article is always indefinite. It occurs either when the noun is modified by an adjective (211a) or by an intensifying PP (211b), but it can also occur with non-modified nouns, especially in exclamations (212a), where it often co-occurs with a consecutive clause with intensive meaning (212b). Remarkably, the presence of the article is not attested before the 20th century, which may indicate that it is a recent tendency.

(211) a. *Mi-*   *era*   *o*   *jenă*   *teribilă*
           me.DAT   was   an   embarrassement   terrible
           'I felt terribly embarrassed'   (phantasma.lett.ubbcluj.ro)
       b. *Mi-*   *era*   *o*   *foame*   *de*   *lup*
           me.DAT   was   a   hunger   of   wolf
           'I was terribly hungry (like a wolf)'   (mansardacubunatati.ro)

(212) a. *Uf,   mi-   e   o   ciudă   pe   lumea   asta   stereotipă!*
      ooh   me.DAT   is   an   anger   on   world.the   this   stereotypical
      'I feel so angry with this stereotypical world!'   (otilia-mantelers.urbankid.ro)
    b. *Mi-   era   o   foame   de   crăpam*
      me.DAT   was   a   hunger   that   cracked.1SG
      'I was so hungry I was going to die'   (corporatistu.ablog.ro)

The fact that the state noun is mostly bare argues in favor of its predicate status and against a subject status, since, as argued by Cornilescu (2009: 203), bare singular NPs are not appropriate as subjects in Romanian, being rather suitable as predicates.

7.2.2   Modification

Another key aspect closely linked to determination is modification. In the previous section, I showed that the presence of the article depends on the presence of a modifier of the state noun in the MIHI EST construction. Therefore, I examined the modification of the state nouns in the MIHI EST pattern, more particularly whether these nouns occur with typical nominal modifiers such as adjectives, or rather with adverbs or degree markers, which are considered verbal modifiers, since they usually modify predicative adjectives or verbs. In my dataset, both are found. However, adverbial modifiers ans degree markers outnumber adjectival ones: 291 out of 362 examples contain a modifying adverb (213a), i.e. 80 % of the modified state nouns, whereas only 54, i.e. 15 %, contain a modifying adjective (213b). The remaining 5 % are modified by other types of modifiers such as intensifying PPs illustrated in (213c), or intensifying clauses, as in (213d), the last two examples given in the previous section (7.2.1), under (211b) and (212b), and repeated here for convenience.

(213) a. *Îmi   era   **foarte** foame*
      me.DAT   was   very   hunger
      'I was very hungry'   (forum.7p.ro)
    b. *Îi   era   **mare** frică   de   tîlhari*
      he.DAT   was   big   fear   of   robbers
      'He was very afraid of robbers'   (A. Vasile, Muntele de foc, 1850)
    c. *Mi-   era   o   foame   **de**   **lup***
      me.DAT   was   a   hunger   of   wolf
      'I was terribly hungry (like a wolf)'   (mansardacubunatati.ro)
    d. *Mi-   era   o   foame   **de**   **crăpam***
      me.DAT   was   a   hunger   that   cracked.1SG
      'I was so hungry I was going to die'   (corporatistu.ablog.ro)

The most frequent adverbs in the datatset are scalar adverbs, expressing a certain degree of the modified state: *foarte* 'very', *mai* 'more', *prea* 'too (much)', *cam* 'rather', *tare* 'very', *atât de* 'so much' *îngrozitor de* 'terribly', *destul de* 'enough', *grozav de* 'terribly', or *cât de* 'how

much'. The data confirm, hence, that the state noun in the MIHI EST construction behaves like a predicative adjective and denotes property (cf. Cornilescu 2009: 203).

Seen from a diachronic perspective, the available data give a better insight into the behavior of the state noun with respect to its modification. Hence, the evolution of the modification of the state nouns throughout the centuries shows that the present situation has not always been the same. This evolution is represented in Figure 7.6.
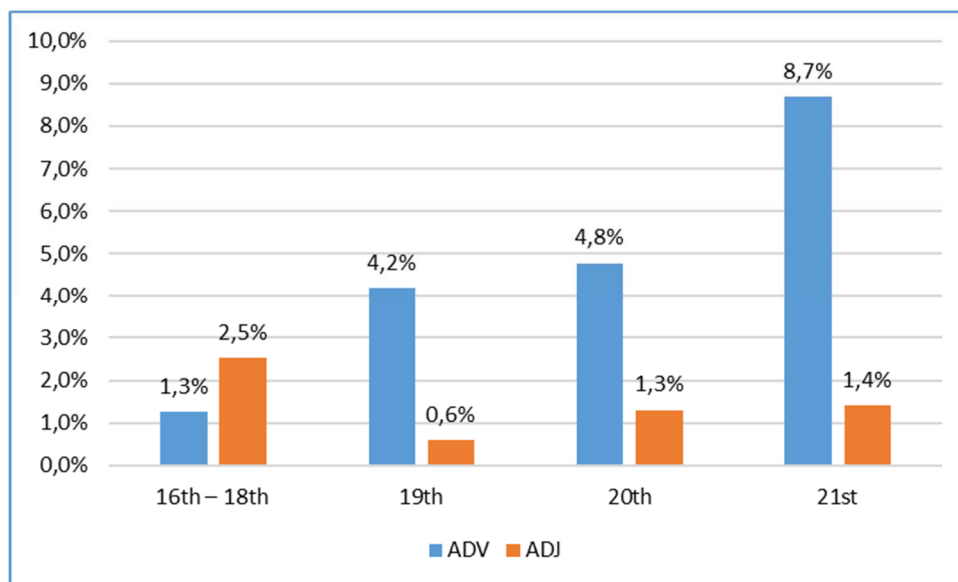

Figure 7.6   Adverbial vs. adjectival modification of the state noun

The graph reveals that adjectival modification was more frequent than adverbial modification between the 16th and 18th centuries, but tends to decrease in the coming centuries. By contrast, adverbial modification increases steadily and becomes much more frequent in the 21st century Romanian. These facts corroborate my hypothesis that the state noun behaves more and more like a predicate in present-day Romanian.

7.2.3   Complementation

The state noun of the MIHI EST construction can take a complement, which refers to the stimulus of the state. This complement may be encoded as a PP, composed of the preposition *de* 'of' + NP (214a), or it may be clausal. In the latter case, it may be non-finite and occur as an infinitive clause (214b) or a supine clause (214c).[91] It may also be a finite clause headed by the subjunctive particle *să* (214d) or the complementizer *că* 'that' (214e).

---

[91] Cf. Footnote 37.

(214) a. *Mi-*   *era*   *frică*   ***de***   ***moarte***
     me.DAT   was   fear   of   death
     'I was scared of death'   (B.S. Delavrancea, Teatru, 1878-1913)

   b. *Nu*   *mi-*   *e*   *jenă*   *de*   *a*   *-mi*   ***asuma***   ***naţionalitatea***
     not   me.DAT   is   uneasiness   of   INF   =me.DAT   assume   nationality.the
     'I don't have any problem to assume my nationality'

   c. *Îi*   *era*   *silă*   ***de***   ***râs***
     him/her.DAT   was   disgust   SUP   laughed
     'She felt disgusted of laughing'   (M. Caragiale, Craii de Curtea Veche, 1929)

   d. *Mi-*   *e*   *ruşine*   ***să***   ***dansez***
     me.DAT   is   shame   SUBJ   dance.1SG
     'I feel ashamed to dance'

   e. *Mi-*   *e*   *teamă*   ***că***   ***nu***   ***voi***   ***mai***   ***visa***
     me.DAT   is   fear   that   not   FUT   anymore   dream
     'I am afraid I will not dream anymore'

The graph in Figure 7.7 visualizes the evolution of the complementation of the state noun in my dataset. It does not include the supine, of which the dataset contains only one example in the MIHI EST construction.
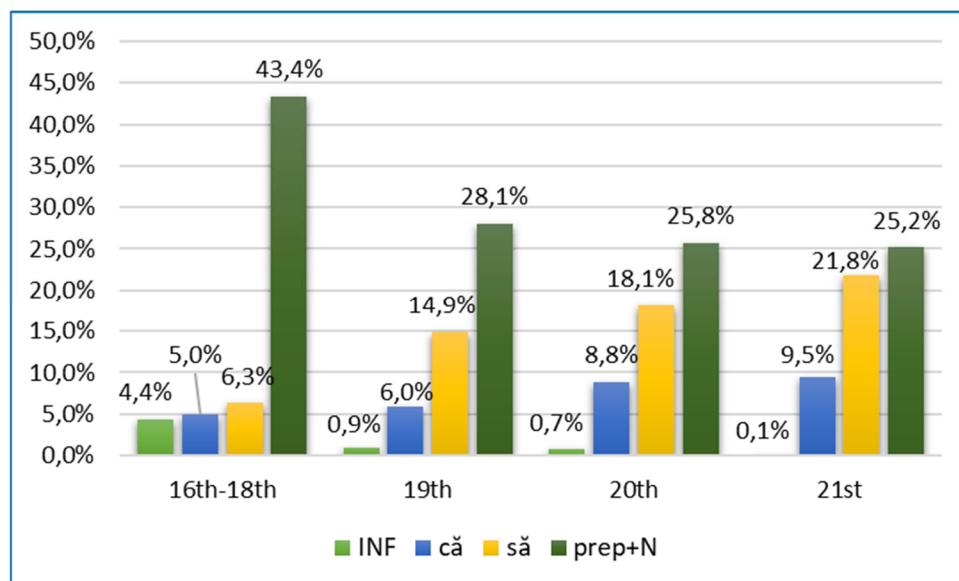

Figure 7.7   Evolution of the complementation of the state noun

Notice that the most frequent complementation across the centuries is the *de-PP*. However, this type of complementation tends to decrease in favor of the *să*-clause. The *că*-clause also becomes gradually more frequent, whereas the infinitive, which was already rare in old Romanian, is used even more rarely in present-day Romanian, due to its progressive replacement with the Balkan subjunctive (cf. Hill 2014; Hill & Mišeska-Tomić 2009), among others), as already mentioned in Section 3.2.4.1 above. Even though

all these types of complements can be used adnominally,[92] mostly with deverbal nouns, they typically occur with predicative categories, such as verbs and adjectives, fact that points toward the predicative nature of the state noun.

7.2.4   The state noun – a control predicate

As argued in several studies (Kempchinsky 1986; Terzi 1993; Cornilescu 2000; Alboiu 2007; Hill 2014), the infinitive, the supine as well as the subjunctive can give rise to control phenomena, but not the indicative. Indeed, with certain types of verbs, the implicit subjects of infinitives, of supines and of subjunctives must or may be coreferential with the subject of the verb in the main clause. When control is obligatory, the complement clause cannot be encoded as a *că*-clause, as shown in (215).

(215) *Mara* *se* *apucă* *a* *citi /* *de* *citit /* *să* *citească /*
       Mara   SE   grabs   INF read /   SUP read /   SUBJ read.3SG /
       *că *citeşte*
       that   reads
       'Mara starts to read'

Therefore, a closer examination of the complements of the state noun in the MIHI EST structure reveals how these nouns behave with respect to control. This investigation is based on the results of the previous section on the complementation of the state noun in the MIHI EST construction. In this section, I showed that the *de* + PP complementation, although still the most frequent, decreases significantly throughout the centuries, whereas subjunctive *să*-clauses and indicative *că*-clauses increase considerably. Since infinitives and supines, which are among the complements of the state noun, are very rare, I have restricted this research to the competition between *să* and *că-* clauses. The graph in Figure 7.8 shows that almost all nouns occur more frequently with the subjunctive *să*-clause than with the indicative *că*-clause.

---

[92] The PP headed by *de* has more adnominal uses than the *să-* and *că*-clauses. Within NPs, this type is in competition with the genitive, which is typically adnominal and never occurs with the state noun in the MIHI EST construction. The PP headed by *de* PP is used with bare nouns, whereas the genitive is used with DPs (cf. Mardale 2007).
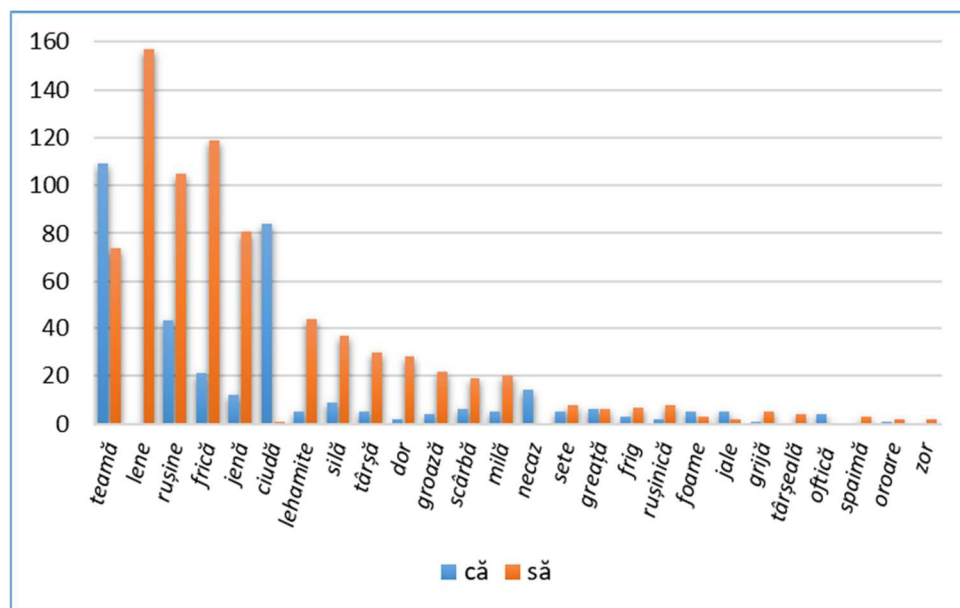
Figure 7.8   Competition between *că* and *să*

The graph indicates that only one noun, i.e. *lene* 'laziness', never occurs with a *că*-clause (216) and is, hence, an obligatory control predicate. The only predicate which occurs almost exclusively with *că* clauses in the dataset is *ciudă* 'envy', illustrated in (217a), in a non-obligatory control structure. However, examples can be found on the web, in which this noun occurs with a *să*-clause (217b). In other words, except for *lene* 'laziness', all nouns visualized in the graph show non-obligatory control, but tend to occur more frequently with *să*-clauses, which allow control, than with *că*-clauses, which exclude control.

(216) *Mi-*　　　*e*　　*lene*　　　*să*　　*citesc /*　　***că***　*citesc*
　　　me.DAT=　is　laziness　SUBJ　read.1SG /　that　read.1SG
　　　'I'm too lazy to read'

(217) a. *Mi-*　　*e*　　*ciudă*　　***să***/　***că***　*nu*　　*plec*　　　　*niciodată*　　*la*　　*timp*
　　　　me.DAT=　is　frustration　SUBJ/　that　not　leave.1SG　never　　　　at　　time
　　　　'I am frustrated that I never leave on time'

　　　b. *Mi-*　　*e*　　*ciudă*　　*să*　　*nu*　*ştiu*　　*limba*　　　　*ţării*
　　　　me.DAT　is　frustration　SUBJ　not　know.1SG　language.the　country.the.GEN
　　　　în　*care*　　*trăiesc*
　　　　in　which　　live.1SG
　　　　'I'm frustrated that I don't know the language of the country in which I live'
　　　　　　　　　　　　　　　　　　　　　　　　　　　　　(niche.snap.monster/Forum)

The increasing frequency of the subjunctive across the centuries confirms that the noun tends to behave more and more like a predicate in Romanian.

7.2.5   Word order

In Section 7.1.1 of this chapter, I presented a usage-based study, carried out with the aim to investigate and compare neutral word order in canonical structures containing a nominative experiencer in subject postion, and in structures containing the MIHI EST construction. In what follows, I present the findings with regard to the syntactic behavior of the state noun occurring in the MIHI EST construction in comparison with the situation in the nominative experiencer construction.

First, observe that the state noun is never left unrealized, both in canonical (218) and in non-canonical structures (219), as shown in Table 7.8, below. If the state noun were a subject, as it is considered by certain scholars, how can it be explained that it may never be replaced by a pronoun and may never be left unrealized in a pro-drop language like Romanian? In the example in (218), the nouns *foame și frig* 'hunger and cold' realize the syntactic direct object and cannot be absent.

(218) *Pe   timpul   lui,   românii   au   făcut *(foame   și   frig)*
       in   time.the   his.POSS   Romanians.the   have.3PL   made   hunger   and   cold
       'In his time, Romanians forbore hunger and cold   (elenaudrea.ro)

Similarly, in example (219) below, *foame* 'hunger' cannot be left unrealized without causing the sentence to become ungrammatical.

(219) *I-   a   fost   *(foame)   fetiței*
       her.DAT=   has   been   hunger   girl.the.DAT
       'The girl has been hungry'   (laptematern.ro)

Table 7.8   The state noun overt vs. covert

State noun	Unrealized	%	Realized		Total
Nom-**Acc** structure	0	0 %	3 266	100 %	3 266
Dat-**Nom** structure	0	0 %	3 969	100 %	3 969

The examples and the data presented above show that the state noun in the MIHI EST construction, traditionally analyzed as a subject (due to its encoding in the unmarked nominative case and to verb agreement), patterns in fact with the object of the canonical structure, given its permanent presence in the sentence.

The same behavior is observed when the data are analyzed from a diachronic perspective. Table 7.9 below shows that, in both the canonical nominative experiencer construction and the MIHI EST construction, the state noun cannot be left unrealized throughout the centuries.

Table 7.9   The state noun (16th-21st) – Realized vs. Unrealized

		16th-18th	19th	20th	21st
Nom-Acc	NP	293	208	307	2 458
	Pro	0	0	0	0
	Total	293	208	307	2 458
Dat-Nom	NP	159	335	691	2 784
	Pro	0	0	0	0
	Total	159	335	691	2 784

Subsequently, the investigation focuses on the position in the clause preferred by the state noun. The following examples illustrate the two distinguished positions, namely the preverbal (220) and postverbal (221) position.

(220) *Ruşine   mi-   i   a   -ţ   spune*
      shame   me.DAT=   is   INF   =you.DAT tell
      'I feel ashamed to tell you'   (1682, Dosoftei, Viaţa şi petrecerea sfinţilor)

(221) *Mi-   era   silă*
      me.DAT=   was   disgust
      'I felt disgusted'   (Delavrancea, Proza, 1878 – 1913)

Contrary to the experiencer, the state noun occurs mostly postverbally. As evident from Table 7.10 and Table 7.11, this position is preferred in 95,1 % of the occurrences in the nominative experiencer construction, and in 98,2 % of the occurrences in the MIHI EST construction. The state noun can also occupy a preverbal position, although very rarely, in both constructions (4,9 % and 1,8 %, respectively).

Table 7.10   The nominative experiencer construction – word order of the realized arguments

	Preverbal	%	Postverbal	%	Total
Experiencer (nom)	846	94,2 %	52	5,8 %	898
State noun (acc)	161	4,9 %	3 105	95,1 %	3 266

Table 7.11   The MIHI EST construction – word order of the realized arguments

	Preverbal	%	Postverbal	%	Total
Experiencer NP (dat)	358	84,6%	65	15,4%	898
State noun (nom)	71	1,8%	3 898	98,2%	3 266

A closer investigation of these facts from a diachronic perspective reveals that the preference for the postverbal position of the state noun has always been very high as opposed to the preverbal position. This is illustrated, for each historical period, in Table 7.12 for both the nominative experiencer construction and the MIHI EST construction. Figure 7.9 and Figure 7.10 give a visualization of the frequencies.

Table 7.12   Position of the state noun (16th-21st) – preverbal vs. postverbal

		16th-18th	19th	20th	21st
Nom-Acc	Preverbal	46 (15,7 %)	11 (5,3 %)	15 (4,9 %)	89 (3,6 %)
	Postverbal	247 (84,3 %)	197 (94,7 %)	292 (95,1 %)	2 369 (96,4 %)
	Total	**293 (100 %)**	**208 (100 %)**	**307 (100 %)**	**2 458 (100 %)**
Dat-Nom	Preverbal	5 (3,1 %)	2 (0,6 %)	8 (1,2 %)	56 (2,0 %)
	Postverbal	154 (96,9 %)	333 (99,4 %)	683 (98,8 %)	2 728 (98,0 %)
	Total	**159 (100 %)**	**335 (100 %)**	**691 (100 %)**	**2 784 (100 %)**
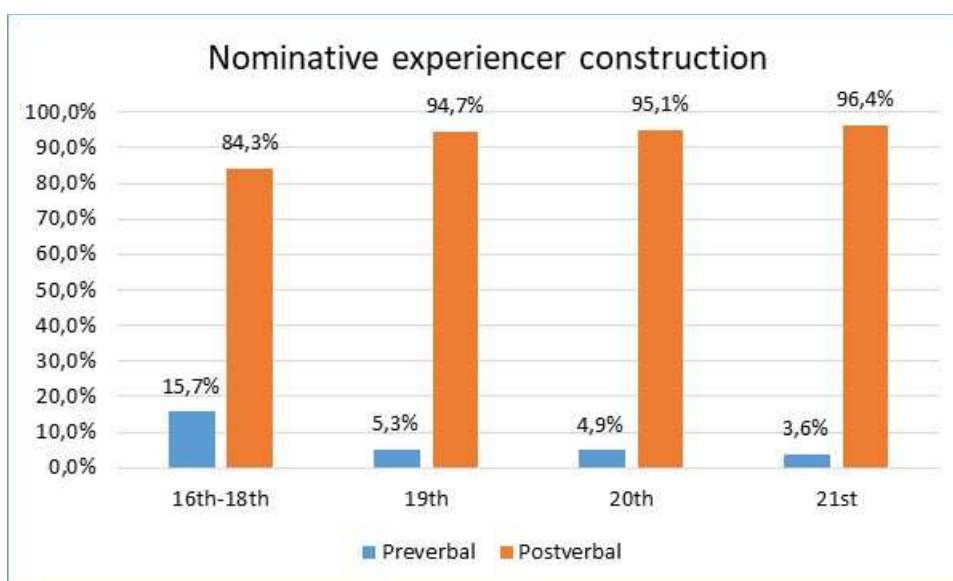
Figure 7.9   The position of the state noun (16th-21st) in the nominative experiencer construction
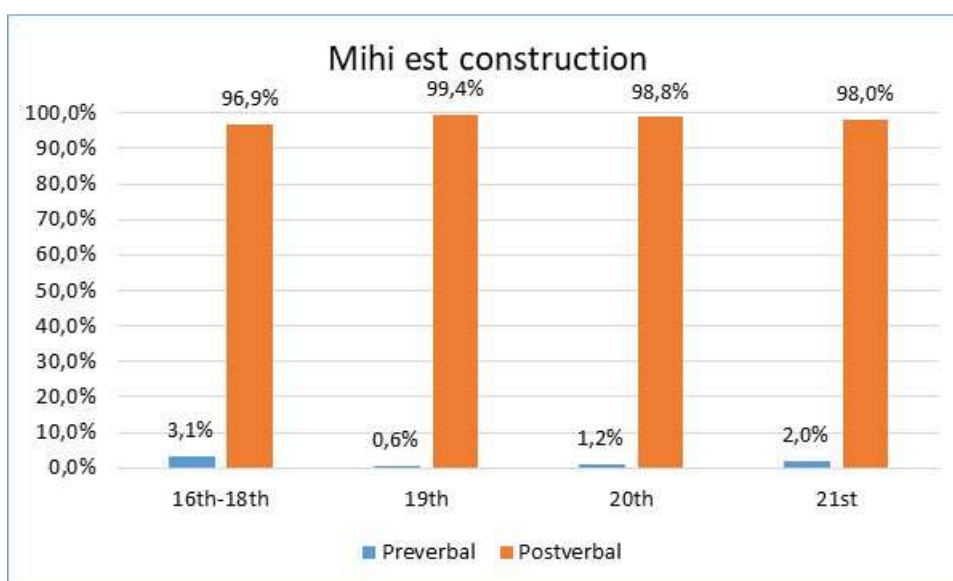
Figure 7.10 The position of the state noun (16th-21st) in the MIHI EST construction

These graphs show that, from the 16th century until now, the state noun occurs postverbally, and, in very few cases, preverbally. Moreover, a slight increase can be noticed throughout the centuries. This increase is more noteworthy in the canonical structure, from 84,3 % in old Romanian (16th-18th centuries), to 96,4 % in the 21st century, whereas in the MIHI EST construction the increase is considerably less noteworthy.

The findings of this case study show that the state noun does not behave like the subject of a canonical structure. More precisely, the state noun is always ralized and prefers the postverbal position, instead of the preverbal position, which, based on the analyzed data, is the dedicated subject position in canonical structures. These results suggest that the state noun is not the subject of the construction, and strengthen instead the predicate hypothesis.

7.2.6 Other arguments

In addition to what has been discussed in the previous sections, two other arguments have been shown to provide evidence in favor of the predicate-like behavior of the state noun in the MIHI EST construction. These involve substitutes and specific question words.

The syntactic position of the state noun in the MIHI EST construction can be occupied also by an adjective in Romanian, as in (222a-b). Dobrovie-Sorin & Giurgea (2013: 93-94) observe that the only other context in which adjectives and nouns can substitute each other is in predicate position (222b).

(222) a. *Mi-*   *e*   *foame /*   *Mi-*   *e*   *drag*
       me.DAT=   is   hungry /   me.DAT=   is   dear
       'I am hungry/ I like him (He is dear to me)'
    b. *Victor*   *e*   *elev /*   *Victor*   *e*   *intelligent*
       Victor   is   student /   Victor   is   intelligent
       'Victor is a student/ Victor is inteligent'

Moreover, as signaled by Cornilescu (2009), the state noun in the MIHI EST costruction has the ability to alternate, besides adjectives, also with adverbs such as *aşa*, 'so, such', as in (223b). Furthermore, the typical question words for this noun are *cum* 'how' and *ce* 'what'. The former is typical for adjectival predicates, illustrated in (223a), whereas the latter is a nominal one, but is much less frequent.

(223) a. *Cum*   *îţi*   *este?*
       how   you.DAT   is
       'How do you feel?'

b. *Aşa   mi-   a   fost   tot   timpul /   Mi-   a   fost   aşa*
  so   me.DAT=   has been all   time.the / me.DAT=   has been so
  tot   timpul
  all   time.the
  'I felt so all the time'   Cornilescu (2009: 203)

Hence, the fact that adjectives as well as adverbs can occupy the same syntactic position as the state nouns of a MIHI EST construction, and the fact that the typical question for these nouns is also typical for adjectival predicates, represent clear evidence in favor of the hypothesis that the state noun behaves like a predicate in this construction.

7.3   Conclusions

The aim of this chapter has been to identify the subject in the MIHI EST construction. In order to do this, I have focused the attention on both the dative experiencer and the state noun. The former makes a very good candidate for a subject through its syntactic behavior, but it fails with respect to the nominative case and verb agreement. The latter is traditionally seen as the subject of the structure, although it has no behavior of a canonical subject.

I argue, hence, that the dative experiencer occurring in the MIHI EST construction displays several relevant subject properties: word order, control, subject-to-subject raising, deletion of subjects in telegraphic style, bare quantifiers in clause-initial position, and the ability to take secondary predicates. As for binding, this diagnostic is less convincing due to less strict binding rules in Romanian. Tests such as deletion under coordination and deletion in imperatives, although they apply to Romanian examples, they are not conclusive due to certain specific features of Romanian. Should we conclude from this that the dative experiencer is not a genuine syntactic subject, but a syntactic indirect object that takes on only a few subject properties, as argued by Moore & Perlmutter (2000) for Russian, Haspelmath (2001a) for several Indo-European languages, and (Fanselow 2002) for German? Following Barðdal & Eythórsson (2003, 2018), I have emphasized that the negative results of Romanian with respect to certain subject tests are due to a general property of clitics, which are essentially case markers and, hence, obligatory in Romanian specific experiencer constructions. Therefore, tests that involve the omission of dative experiencers do not disconfirm the subject status of the dative experiencer, but are simply non-conclusive. All other tests that do not involve unexpressed subjects, such as word order, binding of reflexives, and raising, clearly support the hypothesis that the dative experiencer is a subject in the MIHI EST construction. I have also shown that the state noun of the MIHI EST construction behaves

like a predicate with respect to its modification, complementation, and word order, among others.

The diachronic analysis reveals that, from the earliest attested texts, the state noun behaves like a predicate and the dative argument like a syntactic subject and this tendency becomes increasingly stronger throughout the centuries. The dative experiencer tends more and more to occur clause-initially, whereas the state noun occurs postverbally, and behaves like a predicate, taking typical verbal modification and complementation.

Chapter 8   Productivity of the MIHI EST construction

The present chapter aims to investigate to which extent the MIHI EST construction expands in present-day Romanian. In order to address this research question, I have carried out two different studies: a survey meant to estimate the current tendencies in speaker's usage, and a diachronic corpus study, as a means of observing the evolution of this construction throughout the centuries.

The notion of productivity was initially used in the field of morphology, but in more recent studies it has also been applied to syntax (Goldberg 1995, 2019; Barðdal 2008; Barðdal et al. 2012; Barðdal & Gildea 2015; Zeldes 2012). One of the first definitions of morphological productivity is the one proposed by Schultink (1961). He identifies it with "the possibility for language users to unintentionally coin (uncountably) new words by using an existing word-formation rule" (Schultink 1961: 113). More recent definitions describe it as the property of a word formation process used by speakers to coin "new/potential words" (cf. Aronoff 1976: 38) or as "the possibility that a pattern will apply to new forms" (Bybee 1995: 430; Bybee & Thompson 1997: 384).

As for syntactic productivity, a generally accepted definition does not really exist yet. From a usage-based perspective, syntactic productivity is favored by high *type frequency*, that is, by a high number of different items attested in the relevant slot of a construction (Goldberg 1995, 2019, Bybee & Thompson 1997). Barðdal (2008: 1) identifies it with "an argument structure construction's ability to attract new or existing lexical items, i.e. a construction's extensibility".

The first section of this chapter presents a survey involving native speakers of Romanian, by means of which I investigate to which extent speakers tend to allow the occurrence of new nouns in the MIHI EST construction, nouns which are synonymous with nouns already occurring in this construction (cf. Chapter 6). As for the second section of this chapter, it provides another perspective on productivity, based on a corpus. By means of an integrated study containing quantitative as well as qualitative measures, the changes in productivity of the MIHI EST construction are observed throughout the centuries. This study is based on a dataset extracted from the two corpora mentioned above, for the five documented centuries of Romanian.

8.1 Productivity in the speakers' usage – a pilot study

In this section, one of the aspects of the productivity of the MIHI EST construction is studied on the basis of a questionnaire survey. More precisely, I estimate to which extent native speakers of Romanian accept in the MIHI EST construction nouns which are not attested yet in this construction, but are synonymous with the ones occurring frequently in this construction (cf. Chapter 6). After briefly describing the methodology, I present and discuss the findings of this study, followed by a set of preliminary conclusions.

8.1.1 Methodology

8.1.1.1 Materials: Survey

For the purpose of this study, an online questionnaire survey has been created, using Surveyhero.com and administered in six Romanian universities, namely University Taransylvania Braşov, Universtity of Bucharest, Universtity Babeş-Bolyai Cluj, Craiova University, University Al. I. Cuza Iaşi, and Western Universtity of Timişoara. The survey was launched on 9th of May 2018 and remained available for one month, until the 9th of June 2018. The last completed survey dates from the 5th of June 2018. Before starting the survey, the participants were requested to provide a series of personal data such as name, e-mail address, age range, city of origin, form of education and specialization, the institution to which they were affiliated, and their native language. Note that providing one's name was optional. The *Surveyhero* tool collected information about the moment when the survey was started, and took note of the status of the completion. Each participant has been automatically labeled with a random participant ID number.

8.1.1.2 Participants

A number of 185 respondents have completed the survey, out of a total of 332 participants who started it. These participants are ranging in age from 15 to 85 years. Among them, 133 answered all questions. Hence, the description and analysis of the results of this survey is based on the responses from these 133 participants. The majority of the respondents are aged 15–25 years, and are students (bachelor 78 %, master 14 %). A smaller number of participants were either doctoral students, assistants or young professors, aged 36–45 years (6 %). A very small percentage of the respondents (2 %) were aged above 45 years. Of the total number of participants, 124 were females (93 %) aged 15–85, and nine were males (7 %) aged 15–45 years.

Participants with different educational background were targeted, without them having any specific knowledge in the field of linguistics. However, since the contact

professors teach languages at the universities involved, it is not surprising that 83 % of the respondents study languages, such as Romanian, English, French, Spanish, German. A specific background in Romanian language and literature applies for 26 % of the participants.

8.1.1.3 Procedure

The participants were given the link to the questionnaire by their university professors by means of an e-mail and could choose if and when to complete it.[1] The respondents were encouraged to participate by highlighting the positive impact of their contribution and the importance of the survey to the study presented in this dissertation.

The questionnaire aims to investigate the acceptability of 15 nouns in the MIHI EST construction (cf. Table 8.1, below). These nouns had been selected based on their synonymy with nouns from the gathered list presented in Chapter 6. Note that the nouns selected for the survey are not as frequent as their synonyms, among which some are very frequent, as is the case for *teamă* 'fear', *ruşine* 'shame', *frig* 'cold', *milă* 'pity', *poftă* 'appetite', *ciudă* 'rancor', *lene* 'laziness', and *jale* 'grief'. A small number of nouns are less frequent: *durere* 'pain', *invidie* 'envy', *linişte* 'tranquility', *nelinişte* 'anxiety', *jelanie* 'grief, sorrow', *furie* 'anger' and *deznădejde* 'despair'. As a control mechanism, the questionnaire also contains examples of five other nouns, further referred to as *fillers*, or control nouns. Two of the control nouns are very frequent in the MIHI EST construction (*frică* 'fear' and *jenă* 'embarrassment'), the other three being less frequent but attested more than once in the MIHI EST construction (cf. *oroare* 'horror', *necaz* 'rancor' and *târşă* 'disgust, horror'). Table 8.1 on the next page displays the list of the selected nouns (test nouns, followed by control nouns), and their synonymous correspondents from my dataset.

Each noun was used in four sentences: in three sentences containing the verb *fi* 'be', i.e., in the MIHI EST constructions, and in one sentence containing the verb *avea* 'have', i.e. in the HABEO construction. This yields a total of 60 test sentences.

Table 8.1   Test nouns and control nouns in the survey

Noun	Translation	Synonym of
Test nouns		
amărăciune	grief, sorrow	*jale*
chin	anguish, pain	*durere*
îndurare	pity	*milă*
pizmă	envy, jealousy	*invidie*
ranchiună	rancor	*ciudă*
tihnă	tranquility	*linişte*
mânie	anger	*furie*
răceală	cold	*frig*
sfială	shame	*ruşine*
râvnă	appetite, zeal	*poftă*
zbucium	anxiety	*nelinişte*
temere	fear	*teamă*
tângă	grief	*jelanie, jale*
descurajare	despair, demoralization	*deznădejde*
indolenţă	laziness	*lene*
Control nouns		
frică	fear	-
jenă	embarrassment	-
oroare	horror	-
necaz	rancor	-
târşă	disgust, horror	-

Since the aim of this study is to observe the acceptability of the selected nouns in the MIHI EST construction, in which they have not yet been attested, the three sentences in which they occur in the MIHI EST construction, are artificially constructed by me for this purpose. As for the occurrence of the selected nouns in the HABEO construction, these are, in most cases, attested. The control nouns too have been employed in four sentences each, in the same way as the test nouns. With these 20 control-sentences, the total size of the questionnaire amounts to 80 sentences.

Among the three sentences with the MIHI EST construction, one contains a PP stimulus (224), whereas the other two contain instantiations of control in infinitive (or subjunctive) (225) and in gerund clauses (226).

(224) *Mi-*   *e*   *temere*   *de*   *moarte*
       me.DAT=   is   fear   of   death
       'I'm afraid of death'

(225) *Îi*   *era*   *oroare*   *înainte*   *de*   *a*   *intra*   *la*   *examenul*   *oral*
       him.DAT   was   horror   before   of   INF   enter   at   exam.the   oral
       'He was horrified before taking the oral exam'

(226) *Îi   fu   îndurare   văzând   atâția   copii   nenorociți*
      him.DAT   were   pity   seeing   so_many   children   distressed
      'S/He felt pity seeing so many unfortunate children'

The sentences were randomly ordered and the participants were asked to rate each sentence on a 7-point Likert acceptability scale, with 1 standing for "This is impossible in Romanian", 4 standing for "I don't know (I can't decide)", and 7 standing for "This is exactly how I would say it", following the methodology used in Barddal (2008). Hence, judgments 1–3 represent unacceptability while judgments 4–7 represent acceptability or felicitousness. This entails that the acceptability level is set at 4.

The responses have been extracted from the online platform and further annotated in an excel document. From the personal information collected during the questionnaire with the consent of the participants, some information has been annotated as extralinguistic variables in the Excel file, namely Gender, Age, Region (Center, East, South, West Romania, and Republic of Moldova, henceforth, Moldova), Education level (Bachelor, Master, PhD, and Other) and Professional specialization (Romanian, Modern Languages, or Other field, such as political studies, journalism, history, international relations, etc). Furthermore, a series of linguistic variables was added, such as Etymology of the noun (Lat., Sl., Fr., or Other, namely Gr., Hu., or unknown etymology), Derivation (derived vs. non-derived), Polarity (positive vs. negative connotation), and Stimulus type (finite clause, non-finite clause, PP, or no stimulus). These variables have been annotated in the excel document, during the preparation of the data for the analysis.

8.1.2   Statistical analysis

The data analysis was performed using SPSS (SPSS Statistics 26), and a part of it in R (R Core Team 2020).[2] I start by reporting some basic descriptive statistics. Then I present my findings based on an elementary bivariate analysis, which aims to evaluate the effect of the linguistic and extralinguistic variables on the acceptability rating.

8.1.3   Results

The analysis of the collected responses reveals interesting insights into the dynamics of the MIHI EST construction in speakers' usage. The acceptability judgements of the nouns under scrutiny are based on the average rating per question. This has first been calculated

---

[2] The statistical analysis in R reported on in this chapter has been performed in collaboration with Ludovic De Cuypere, the specialist in Statistics in our department, and under the guidance of the consultants from FIRE UGent (Fostering Innovative Research based on Evidence), to whom I am deeply indebted.

for the whole dataset, i.e. for the total number of participants. Then an average rating has been calculated per participant, which has allowed to identify possible outliers among the participants. Subsequently, the average rating per question has been calculated for each of the linguistic variables (Etymology, Derivation, Polarity and Stimulus type), as well as for the extralinguistic variables (Gender, Age, Region, Education level and Specialization).

Recall that the survey has been initiated with the intention to verify whether or not a series of nouns can occur in the MIHI EST construction. Since the selected nouns are not attested in this construction, I have relied on sentences artificially created by me through introspection. The aim of the analysis is to investigate whether these unattested sentences are rated as acceptable by native speakers.

It is worth mentioning that Keller, Lapata & Ourioupina (2002) and Keller & Lapata (2003) have detected, in their research, a positive correlation between frequency and acceptability judgments of native speakers. They show that the more frequent a lexical-syntactic combination is in a specifc corpus (their study is based on the British National Corpus), the higher it is rated in acceptability by native speakers. Conversely, infrequent lexical-syntactic combinations are judged less acceptable by native speakers. It thus becomes clear that the threshold of 4 is rather high for the nature of the present survey, and that the acceptability judgements need to be nuanced and cannot be described in tones of black and white.

In order to calculate the mean rating for each question, the scores given by each participant for a specific question have been added up, and then divided by the total of 133 participants. The obtained result is a decimal number between one and seven. Table 8.2 contains all the sentences considered acceptable with an average rating above 4. In this table, "***" indicates sentences containing a control noun. For the translation of all the sentences used in the present survey I refer the reader to Appendix 1, at the end of this dissertation.

Table 8.2   Sentences above the acceptability threshold of 4

Q_Nr	Sentence	Mean
Q28***	*Mi-era foarte* frică *înainte de a intra în sala de examen.*	6,74
Q78***	*Îi era şi mai* frică *văzând că sala era plină.*	6,63
Q55***	*Îi e* frică *de ea însăşi.*	6,57
Q19***	*Îi era* jenă *văzând că toţi o privesc.*	6,55
Q33***	*Mi-era* jenă *înainte de a-mi mărturisi greşeala.*	6,24
Q23***	*Am mare* oroare *de insecte.*	6,17
Q64	*Am un mare* chin *în suflet.*	5,79
Q52	*A avut mare* îndurare *de noi!*	5,77
Q71	*Am o* amărăciune *în suflet, de neînţeles!*	5,69
Q21***	*Avea* necaz *pe fratele său.*	5,38
Q51***	*Mi-e* oroare *de accidente.*	5,33
Q1***	*Îi era* jenă *de ea însăşi.*	5,22
Q27***	*Îmi era şi mai* necaz *văzând că el nu venise.*	5,13
Q62***	*Am mare* frică *să nu se simtă cumva deranjat.*	5,05
Q66	*Am mare* zbucium *în suflet.*	4,83
Q46	*Am* ranchiună *în suflet când îi privesc.*	4,71
Q69	*Am avut mare* mânie *în suflet la aflarea veştii.*	4,67
Q34	*Când intru în casa ta, am mare* tihnă *în suflet.*	4,65
Q9***	*Îi era* târşă *de aşa oameni.*	4,64
Q37	*Am o mare* răceală *în suflet.*	4,61
Q43	*Are mare* pizmă *pe fratele ei!*	4,56
Q60***	*Îi era* oroare *înainte de a intra la examenul oral.*	4,32
Q47***	*Mi-era* târşă *auzindu-l cum striga.*	4,24
Q20***	*Mi-era deja* târşă *înainte de a-l vedea.*	4,10

As expected, the most acceptable sentences contained control nouns, followed by sentences containing some of the non-control nouns in a HABEO construction. Based on the average rating per question, all sentences containing a MIHI EST construction with a test noun have been rated under the acceptability level of 4. However, what is interesting is that even the sentences with control nouns have not all reached the acceptability threshold of 4. For instance, Table 8.3 below shows that, with the exception of *frică* 'fear', each of the control nouns has received a poor rating in at least one sentence. Moreover, for two of the selected control nouns (cf. *târşă* 'disgust, horror' and *jenă* 'embarrassment') the sentence containing the HABEO construction has been rated under the treshold of 4, whereas the MIHI EST construction is rated higher.

Table 8.3   Control sentences below the acceptability level

Q_Nr	Sentence	Mean
Q14***	*Îi era* necaz *pe sine însuşi.*	3,95
Q25***	*Am mare târşă de oameni ca el.*	3,87
Q72***	*Ne fusese* oroare *văzând atâtea nedreptăți.*	3,70
Q3***	*Am mare* jenă *de vecina mea.*	2,95
Q2***	*Mi-era* necaz *înainte să intru în sala de examen.*	2,93

The bimodal distribution of the overall acceptability ratings shown in Figure 8.1 indicates that the test sentences provoked strong responses of the respondents. More than half of all ratings are equal to either 1 (n = 4263, 40%) or 7 (n = 2424, 23%). Recall that 1 equals "totally unacceptable", while 7 equals "totally acceptable". Additionally, Figure 8.1 shows a difference between the control and test nouns. Overall, as expected, filler nouns tend to receive a higher rating than the test nouns. This difference is significant, based on a Wilcoxon-Mann-Whitney test (p-value < 0.0001).
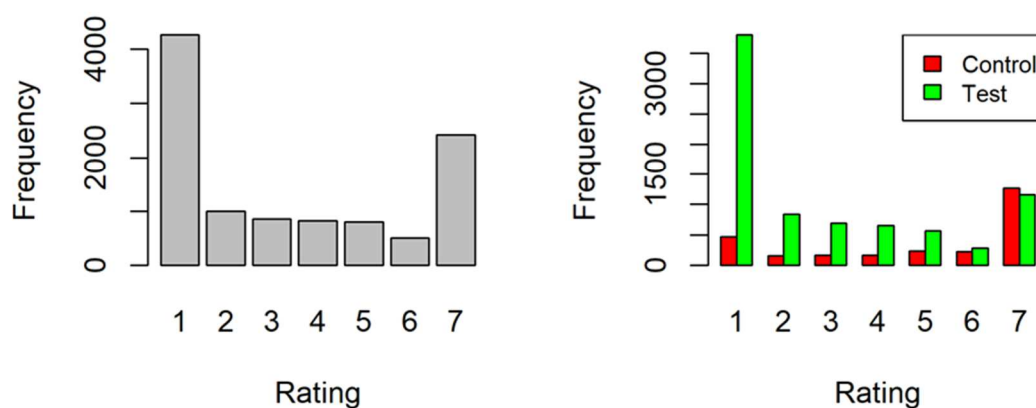
Figure 8.1   Acceptability ratings (left) and Ratings by Control vs. Test noun (right)

8.1.3.1   Test and control nouns

Zooming in on the ratings per noun, Table 8.4 presents overall summary statistics for each noun, while Figure 8.2 displays the distributions of the ratings for each noun by means of histograms. Table 8.4 presents the following statistics: Minimum (received score), Mean of the scores per noun, followed, between brackets, by SD (Standard Deviation), Median (i.e. the middle value, separating the total number of ratings in two equal halves), and Maximum (received score).

Table 8.4 Bivariate descriptive statistics of Rating by Noun

Noun	Min	Mean (SD)	Median	Max
Filler nouns	**1**	**5 (2.4)**	**6**	**7**
Frică	1	6.2 (1.7)	7	7
Jenă	1	5.2 (2.2)	6	7
Oroare	1	4.9 (2.4)	2.4	7
Necaz	1	4.3 (2.4)	5	7
Târşă	1	4.2 (2.5)	5	7
Test nouns	**1**	**2.9 (2.2)**	**2**	**7**
amărăciune	1	3.7 (2.4)	3	7
Chin	1	3.6 (2.5)	3	7
îndurare	1	3.3 (2.5)	2	7
pizmă	1	3.3 (2.3)	3	7
ranchiună	1	3.2 (2.3)	2	7
Tihnă	1	3.1 (2.3)	2	7
mânie	1	2.9 (2.3)	2	7
răceală	1	2.8 (2.2)	2	7
Sfială	1	2.8 (2.2)	2	7
râvnă	1	2.8 (2.1)	2	7
zbucium	1	2.7 (2.2)	1	7
temere	1	2.6 (2.0)	2	7
Tângă	1	2.2 (1.7)	1	7
descurajare	1	2.0 (1.8)	1	7
indolenţă	1	1.9 (1.6)	1	7

From Table 8.4 it can be observed that the mean and median ratings of the control nouns are all notably higher than those of the test nouns. Futhermore, the table indicates that all nouns received the lowest as well as the highest possible rating (cf. columns Min and Max in Table 8.4). The summary statistics suggest that the test nouns tend to be rated as not acceptable, as all means and medians are lower than 4, the acceptability threshold. Figure 8.2 on the next page illustrates the overall distribution of the ratings for each noun.

Notice again the large variability in the ratings (all nouns received the lowest as well as the highest possible rating), as well as the highly skewed distribution, which suggests that the nouns provoke extreme ratings.
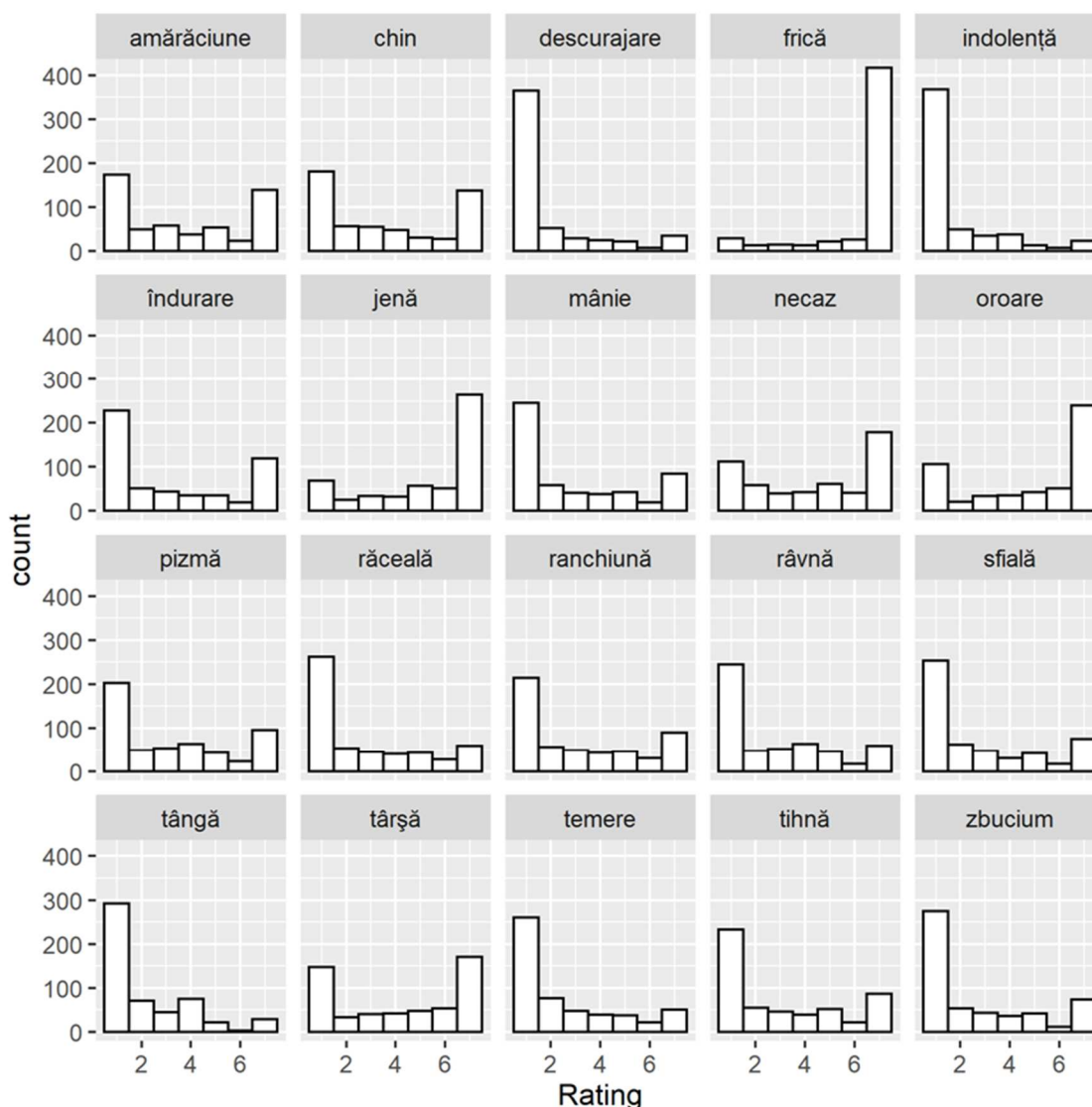
Figure 8.2 Distributions of the ratings per noun

8.1.3.2 MIHI EST construction vs. HABEO construction

In what follows, I split the collected data in function of the construction in which the nouns occur, namely the MIHI EST (with the verb *fi* 'be') or the HABEO (with the verb *avea* 'have') construction. More precisely, I investigate how the sentences containing these nouns ocurring in one of the two constructions have been rated in average.

The data show that the construction in which the nouns occur is crucial, since it has the highest impact on the ratings, and reflects, at the same time, a certain preference of the speaker for one of the two competing constructions. Figure 8.3 gives an overview of the ratings received by each of the two constructions. The test nouns are represented apart from the control nouns.
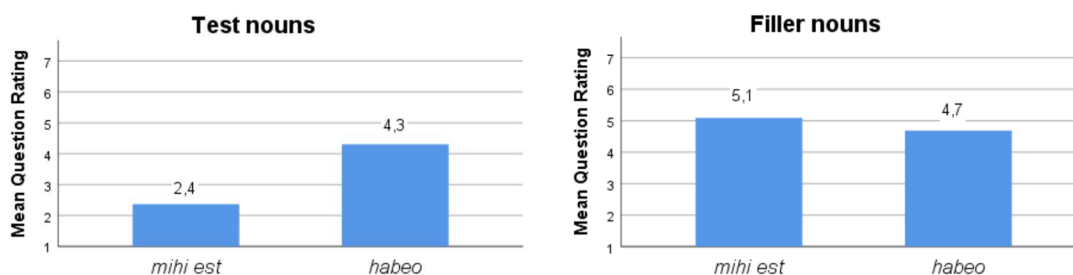
Figure 8.3   Mean rating per question in the MIHI EST and the HABEO constructions

Based on these graphs, there is a clear difference between the way the test nouns and the control nouns have been rated with regard to each of the two constuctions. The ratings received by the test nouns tend to be very low when they occur in the MIHI EST construction, compared to the scores these nouns receive when occurring in the HABEO construction. As for the control nouns, the situation is reversed, their preferred construction being MIHI EST. The tendencies are even clearer when we correlate the construction in which the nouns occur, with the spread of the scores (cf. Figure 8.3).

In Figure 8.4 below, the spread of the scores is illustrated through boxplot charts. The advantage of a boxplot is that it shows how the data are distributed. The length of the box tells us about how much variation exists in the data, and the line across the box, the median, marks the middle value of the data. Half of the scores are greater than or equal to this value and half are smaller. The comparative length of both whiskers (lower and upper side) taken with the position of the median (the line inside the box) gives us an idea of the distribution of the data. Such a graph has the benefit of the special marking of outliers (a circle in the graph) and of extreme values (a star in the graph). Outliers, as well as the extreme values, are isolated values situated outside the wiskers.
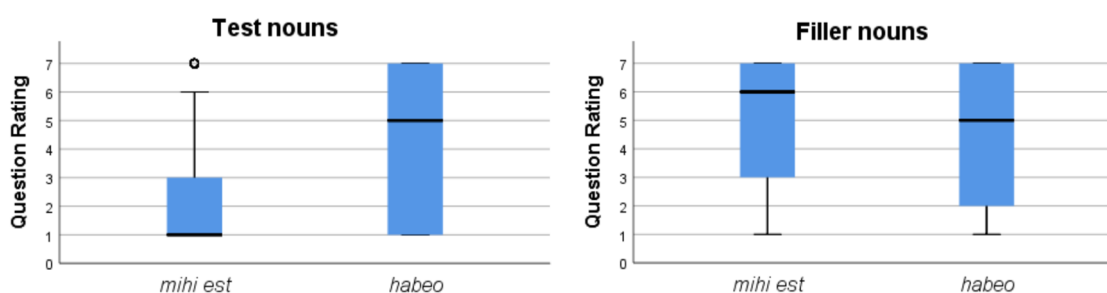

Figure 8.4   Spread of the scores in the two constructions: MIHI EST VS. HABEO

For instance, in Figure 8.4, in the left-hand side graph, the boxplot for the MIHI EST construction shows that half of the total ratings for the test nouns occurring in this construction are scores of 1. The other 50 % of the total ratings contains scores between 1 and 7. As for the box of the boxplot, it shows that, from the upper 50 % of the ratings, the first 25 % (in the box above the median) are scores between 1 and 3, and the remaining

25 % (the wisker) are scores between 3 and 6. The difference between the data included in the box and the data comprised in the wiskers is that the most frequent values (= received scores) occur in the box, whereas the wiskers contain the least frequent values. The circle situated on 7 shows that the test nouns occurring in the MIHI EST construction have also been scored with 7, the maximum score, although very isolated.

The representation of the data for the fillers (the right-hand side graph, in Figure 8.4, above) shows that they too received the minimum rating, regardless of the construction in which they occur. Moreover, when these nouns occur in the MIHI EST construction, the median of the scores tends to be higher than when they occur in the HABEO construction. This illustrates the competition between the two constructions, and may imply that the more acceptable a noun is in the MIHI EST construction, the less acceptable it becomes in the HABEO construction.

Let us now turn our attention to the selected nouns in order to observe how these have been rated in the two competing constructions. Table 8.5 displays the average rating per question for each noun, per construction.

Table 8.5   Bivariate descriptive statistics of Rating by Noun by construction

Noun	Min	Mean (SD)		Median		Max
		MIHI EST	HABEO	MIHI EST	HABEO	
Fillers						
frică	1	6.6 (1.2)	5.1 (2.3)	7	6	7
jenă	1	6.0 (1.7)	3.0 (2.0)	7	3	7
oroare	1	4.5 (2.5)	6.2 (1.6)	5	7	7
târşă	1	4.3 (2.5)	3.9 (2.4)	5	4	7
necaz	1	4.0 (2.4)	5.4 (2.1)	4	7	7
Test						
amărăciune	1	3.0 (2.2)	5.7 (2.0)	2	7	7
chin	1	2.9 (2.1)	5.8 (2.0)	2	7	7
pizmă	1	2.8 (2.1)	4.6 (2.4)	2	5	7
ranchiună	1	2.7 (2.0)	4.7 (2.4)	2	6	7
sfială	1	2.5 (2.1)	3.7 (2.4)	1	3	7
tihnă	1	2.5 (2.0)	4.6 (2.4)	1	5	7
îndurare	1	2.4 (2.0)	5.8 (2.0)	1	7	7
mânie	1	2.4 (1.9)	4.7 (2.4)	1	5	7
râvnă	1	2.4 (1.9)	4.0 (2.3)	1	4	7
temere	1	2.2 (1.8)	3.7 (2.3)	1	3	7
răceală	1	2.1 (1.7)	4.6 (2.3)	1	5	7
tângă	1	2.1 (1.6)	2.4 (1.9)	1	1	7
zbucium	1	2.0 (1.6)	4.8 (2.4)	1	5	7
indolenţă	1	1.7 (1.5)	2.2 (1.9)	1	1	7
descurajare	1	1.5 (1.2)	3.2 (2.4)	1	2	7

Figure 8.5 offers a visualization of these facts, by means of bar charts: the MIHI EST in blue in the graphs, and the HABEO construction in orange in the graphs. The chart in (a) contains the control nouns, whereas the chart in (b), the test nouns.
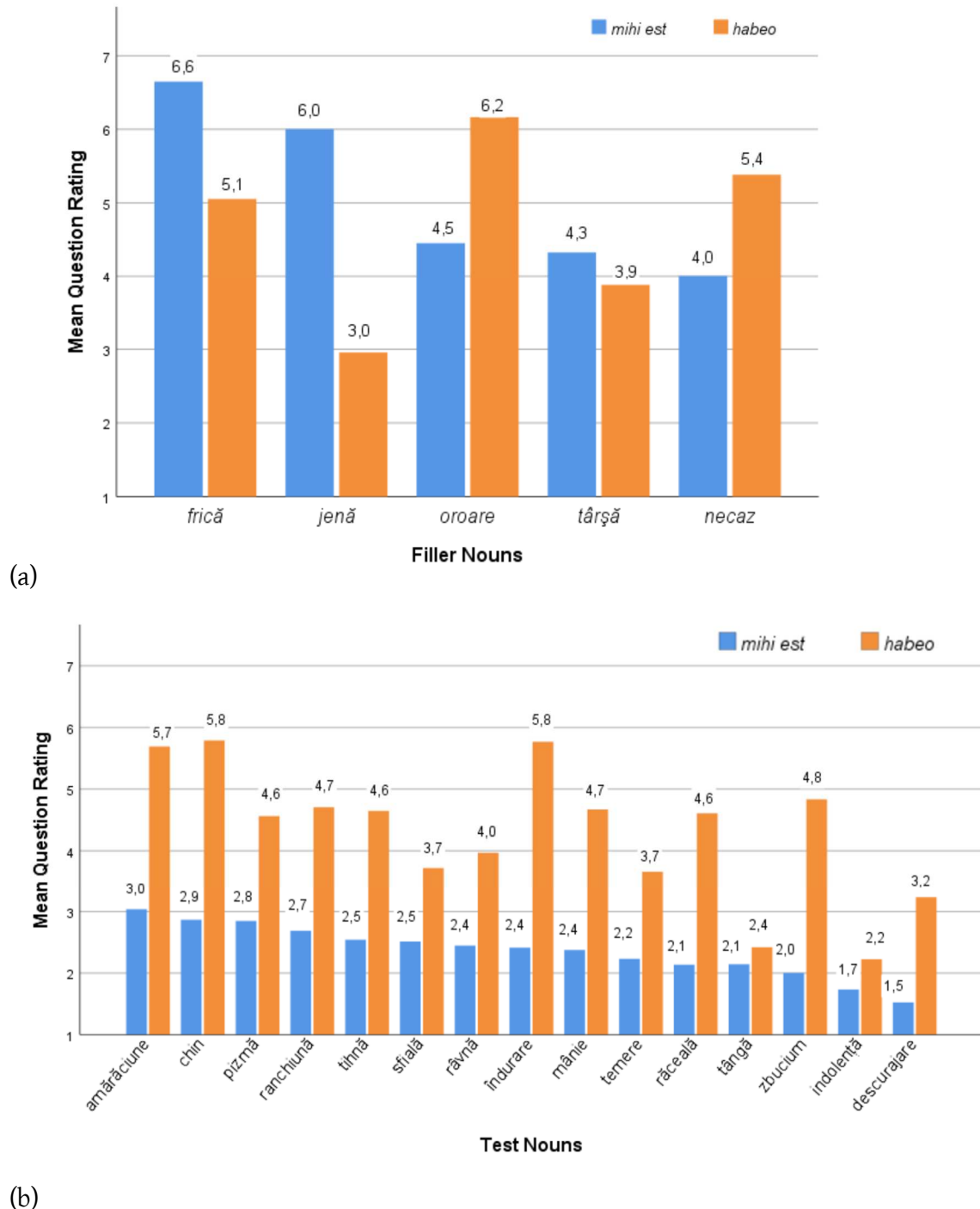

(a)

(b)

Figure 8.5   Nouns ordered in function of their average acceptability rating

As expected, the control nouns show the highest average rating (cf. Table 8.5 above, and Figure 8.5a). As for the test nouns, Figure 8.5b displays them in the order of their acceptability in the MIHI EST construction, *amărăciune* 'grief, sorrow' being the most

acceptable, and *descurajare* 'despair, demoralization' the least acceptable for the participating native speakers (cf. also, Table 8.5).

A closer look at the spread of the scores for each noun reveals how often a certain score has been received by a certain noun. In Figure 8.6 below, the spread of the scores for the test nouns is illustrated through boxplot charts, the nouns being ordered in function of their average rating when occurring in the MIHI EST construction. Observe, for instance, the first boxplot, the one in blue, for the highest ranked test noun in the MIHI EST construction, i.e. *amărăciune* 'grief, sorrow' (cf. the (a) graph). This boxplot shows that half of the total ratings are scores of 1 and 2, whereas the other 50 % of the total ratings of this noun are scores between 2 and 7. The box of the boxplot shows that, from the upper 50 % of the ratings of *amărăciune* 'grief, sorrow', the first 25 %, located in the box above the median, are scores between 2 and 5, and the remaining 25 % (the wisker, hence, the least frequent values) are scores between 5 and 7. In the same chart, in Figure 8.6a, the orange boxplot of *amărăciune* 'grief, sorrow', occurring in the HABEO construction, has the median 7, what means that 50 % of the received ratings are 7 for this noun, with no variation. As for the lower 50 % of the ratings, they are more spread. The most frequent 25 % of the ratings, contained in the box in the boxplot, are scores between 5 and 7, whereas the remaining 25 % of the ratings, the least frequent ones, are spread between 2 and 5. The circle situated on the line denoting 1 shows that, isolated, this noun has also been rated with 1.
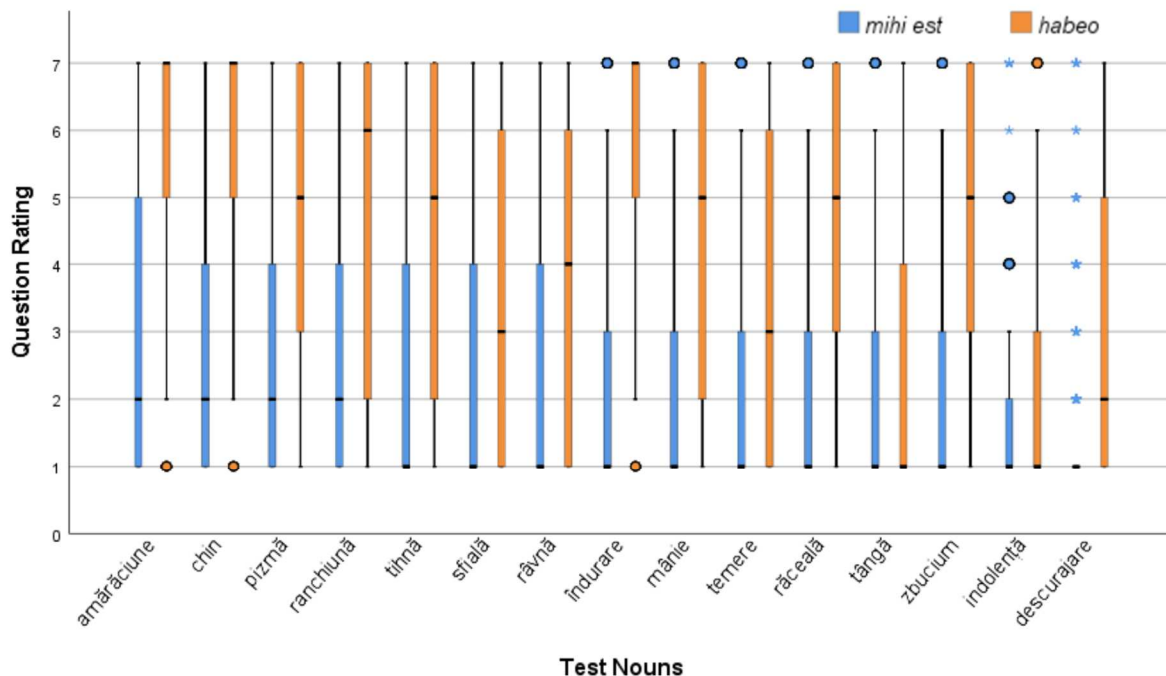
(a)
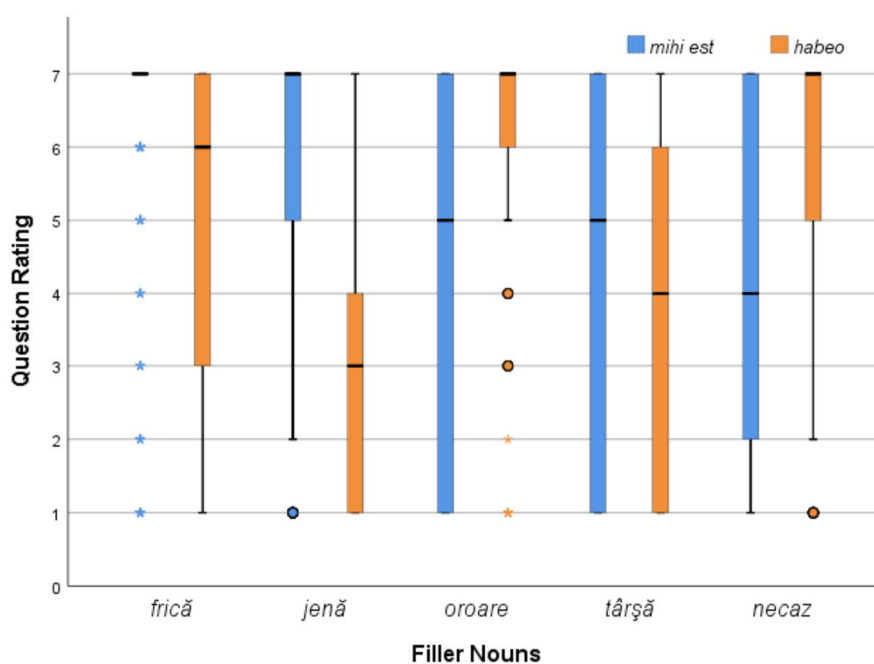
(b)

Figure 8.6   Spread of the ratings. Test nouns (a) vs. Control nouns (b)

Notably, in the (b) chart in Figure 8.6, the data for *frică* 'fear' in the MIHI EST construction have a very strange distribution. The (blue) boxplot has no box, and no wiskers, but only the median, situated on 7. This shows that the sentences containing the control noun *frică* *'fear'* in the MIHI EST construction have mostly been rated with the maximum score, namely 7, and only a few participants have rated them with scores from 1 to 6, but these remain extremely low-frequent ratings.

To conclude, higher mean ratings are witnessed when the nouns occur in the HABEO construction, with the exception of two control nouns (cf. *târşă* 'disgust, horror' and *jenă* 'embarrassment'), which seem to have the MIHI EST structure as the preferred construction. However, the distribution of the scores shows that the test nouns can also be rated as fully acceptable at least in certain sentencial contexts and by certain participants.

In what follows, I turn to establishing which linguistic contexts and which extralinguistic features related to the participants yield a higher acceptability rating.

8.1.3.3   Effect of the linguistic variables on the acceptability rating

In this section, my attention focuses on linguistic variables related to the test nouns. More particularly, I examine the impact of these variables on the competition between the MIHI EST construction, with the *fi* 'be', and the HABEO construction, with the verb *avea* 'have'. These variables are Etymology, Derivation, Polarity and Stimulus type on the acceptability rating. Table 8.6 presents descriptive statistics of the ratings of the selected linguistic variables, as well as on the construction in which the nouns occur. The reported

frequencies (Freq) represent the number of sentences in which a specific element occurs (for instance, five out of the 20 nouns are of French etymology, and each noun is rated in four different sentences; this gives 5 x 4 = 20 sentences in which the verb *fi* 'be' occurs. These 20 sentences represent 25 % of the total of 80 sentences of the survey). The figures in the N column refer to the total ratings this specific element has received (for instance, the 20 sentences containing the verb *fi* 'be' have been rated by 133 participants, that is, 20 x 133 = 2 660).

Table 8.6 Descriptive statistics for rating conditional on the linguistic variables (MIHI EST vs. HABEO)

| Variable and levels (Freq; %)	Min	Mean (SD)		Median		Max	N
		MIHI EST	HABEO	MIHI EST	HABEO		
Etymology							
French (20; 25%)	1	3.3 (2.5)	3.9 (2.5)	2	4	7	2 660
Latin (24; 30%)	1	2.4 (1.9)	4.9 (2.3)	1	6	7	3 192
Slavic (24; 30%)	1	2.7 (2.1)	4.1 (2.4)	2	4	7	3 192
Other (12; 15%)	1	4.6 (2.5)	4.9 (2.4)	6	6	7	1 596
Derivation							
derived (32; 40%)	1	2.4 (1.9)	4.4 (2.4)	1	5	7	4 256
non-derived (48; 60%)	1	3.5 (2.5)	4.4 (2.5)	3	5	7	6 384
Polarity							
negative (60; 75%)	1	3.0 (2.3)	4.5 (2.5)	2	5	7	7 980
positive (20; 25%)	1	3.2 (2.4)	4.2 (2.4)	2	5	7	2 660
Stimulus type							
no stimulus (25; 31%)	1	2.5 (2.0)	4.6 (2.5)	1	5	7	3 325
finite clause (7; 9%)	1	2.1 (1.8)	4.0 (2.5)	1	4	7	931
non-finite clause (28; 35%)	1	3.3 (2.5)	3.8 (2.3)	2	4	7	3 724
prepositional phrase (20; 25%)	1	3.5 (2.5)	4.6 (2.4)	3	5	7	2 660

As evident from this table, when the nouns occur in the MIHI EST construction, higher mean ratings seem to be associated with nouns with etymology "Other" (that is, Greek, Hungarian, or unknown etymology), as well as with non-derived nouns, with contexts containing a positive event and with structures taking a PP as a stimulus.

The first two variables, namely Etymology, and Derivation concern the nouns selected for the survey. Among the test nouns, three of them have French etymology, six come from Latin, five have a Slavic origin, and one comes from Hungarian. As for the control nouns, two have French etymology, one comes from Greek, one from Slavic, and for one noun the etymology is unknown. Since the nouns with Greek, Hungarian, and unknown etymology are very few, I grouped them under the label "Other".

The data show that etymology has a certain impact on the average acceptability rating. Higher mean ratings can be observed for the nouns with etymology "Other", either occurring in the MIHI EST or in the HABEO construction. Again, the sentences containing the HABEO construction tend to receive the highest ratings in the case of the test nouns

(cf. Figure 8.7a). As for the control nouns (Figure 8.7b), those of French etymology as well as the ones labeled with "Other" show considerably higher ratings in the MIHI EST construction than in the HABEO construction. The competition between the two constructions is represented in each of the following graphs. The blue bar gives the data for the MIHI EST construction, whereas the orange bar illustrates the data for the HABEO construction.
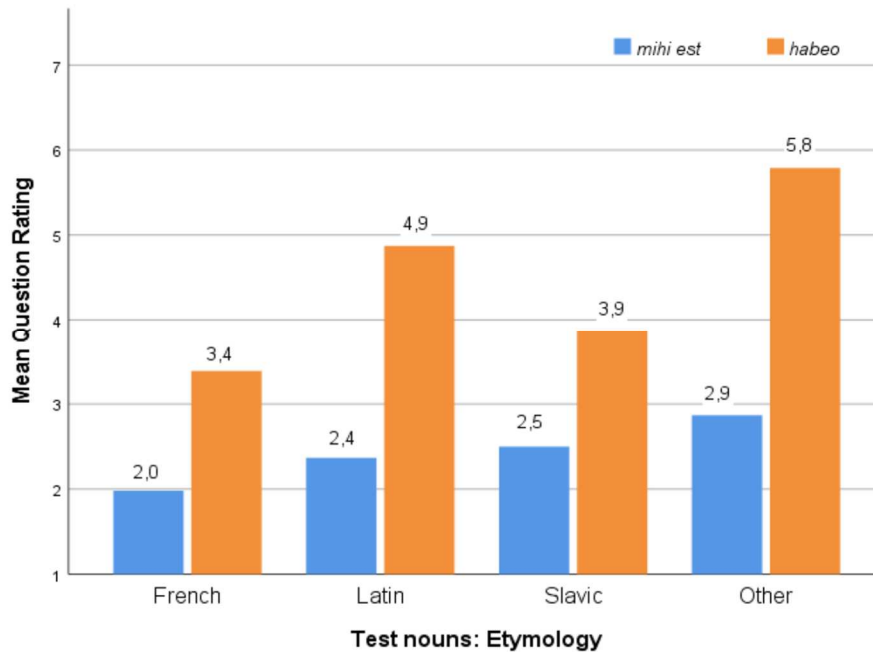

(a)
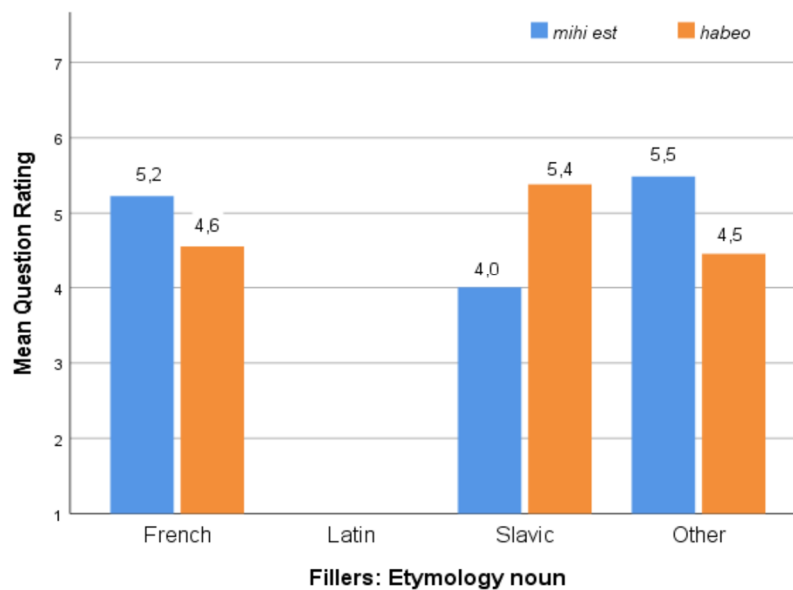
(b)

Figure 8.7   Etymology of the noun and its impact on the average acceptability rating. Test nouns (a) vs. Control nouns (b)

As for the Derivation variable, the list of nouns selected for the survey contains non-derived, as well as derived nouns. In order to identify which of the nouns I selected are derived, I relied on the information given in the DEX online, the online Romanian Explanatory Dictionary. Among the test nouns, eight of them are derived nouns (*amărăciune* 'grief, sorrow', *descurajare* 'despair, demoralization', *îndurare* 'pity', *răceală* 'cold', *râvnă* 'appetite, zeal', *sfială* 'shame', *temere* 'fear', *tihnă* 'tranquility'), and seven nouns are non-derived (cf. *chin* 'anguish, pain', *indolență* 'laziness', *mânie* 'anger', *pizmă* 'envy, jealousy', *ranchiună* 'rancor', *tângă* 'grief', *zbucium* 'anxiety'), whereas the control nouns are all non-derived.[3]
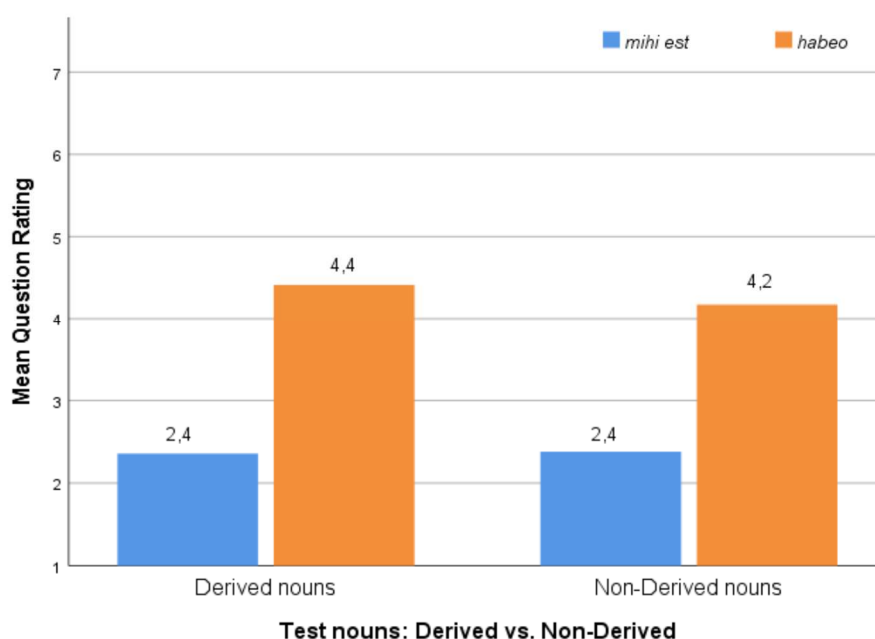

(a)
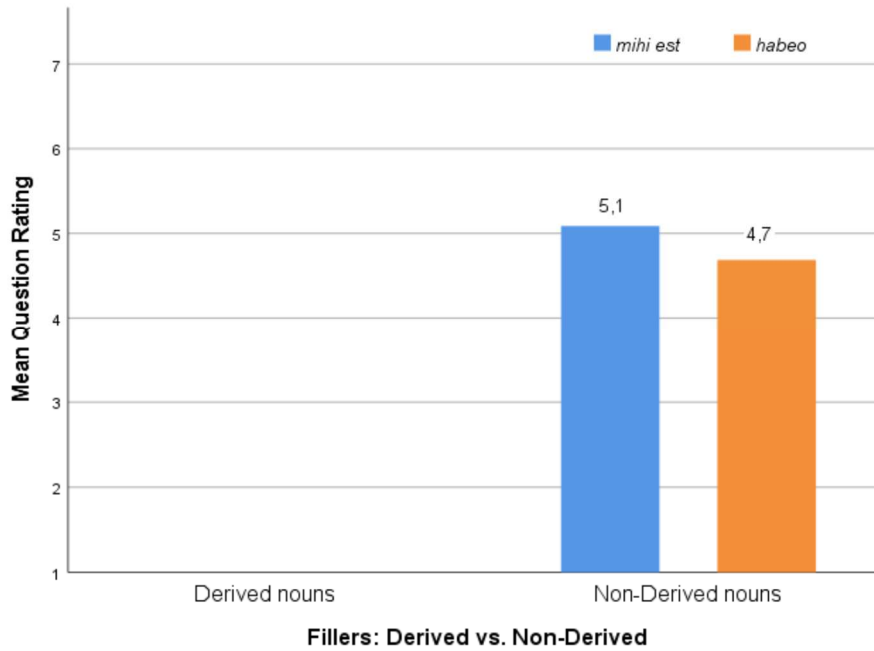
(b)

Figure 8.8   Derivation and its impact on the average acceptability rating. Test nouns (a) vs. Control nouns (b)

Overall, the test nouns are higher rated in the HABEO construction than in the MIHI EST construction, whether they are derived or non-derived (cf. Figure 8.8a). At a closer look at the test nouns occurring in the MIHI EST construction, the data show that the derived and the non-derived nouns receive approximately similar average ratings, the difference between them being rather minor.

The other two linguistic variables are Polarity of the event and Stimulus type. Polarity identifies whether the event refers to a negative or a positive state. Based on the collected data, slightly higher ratings have been given to the test sentences containing a MIHI EST construction with a positive Polarity. The same holds also for the filler sentences. Notwithstanding, in the control sentences containing the HABEO construction, negative polarity was higher rated. As for the competition between the two constructions, higher ratings are associated with the HABEO construction in the test sentences (cf. Figure 8.9a), but with the MIHI EST construction in control sentences (cf. Figure 8.9b).
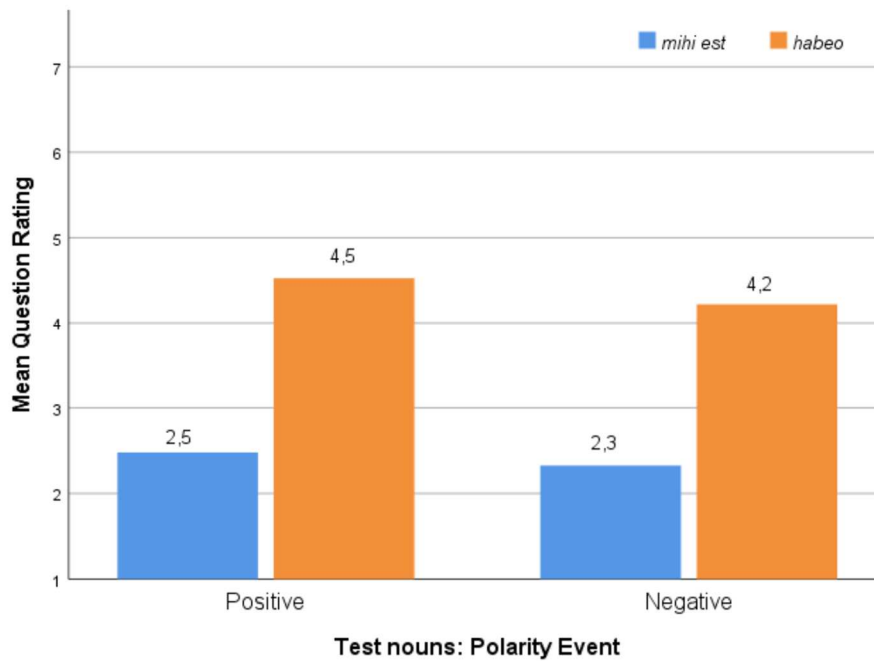
(a)
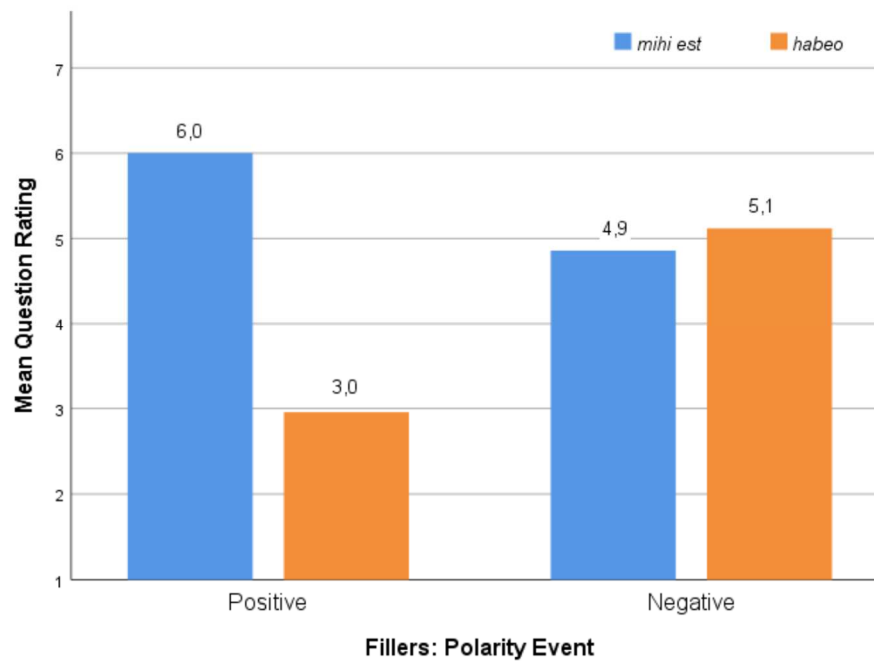
(b)

Figure 8.9   Impact of Polarity of the event on the average acceptability rating. Test nouns (a) vs. Control nouns (b)

The extent to which the presence and the type of the stimulus influence the acceptability rating has also been examined. In Figure 8.10 below, the impact of the presence and of the type of stimulus on the average acceptability rating is plotted first for the test sentences (a), then for the control sentences (b).
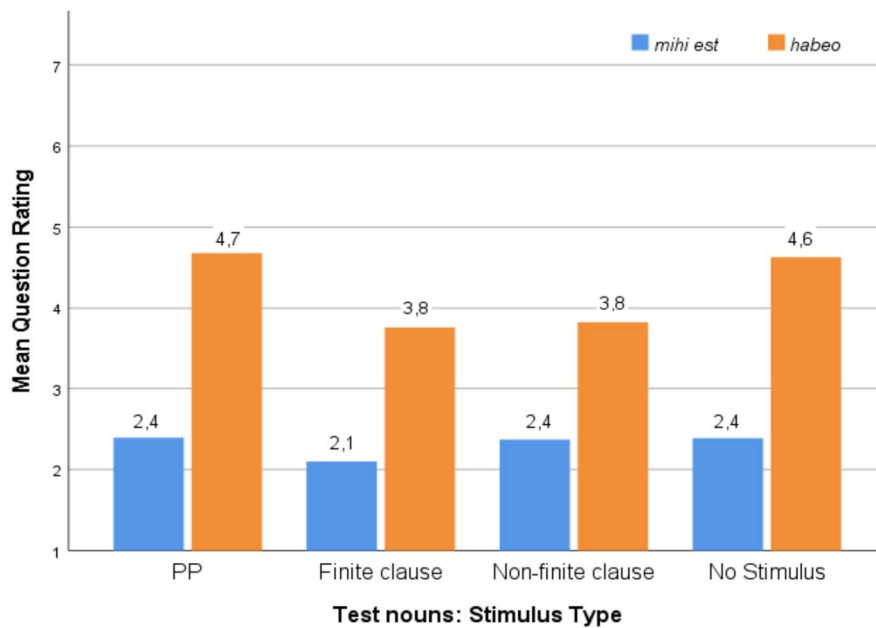
(a)
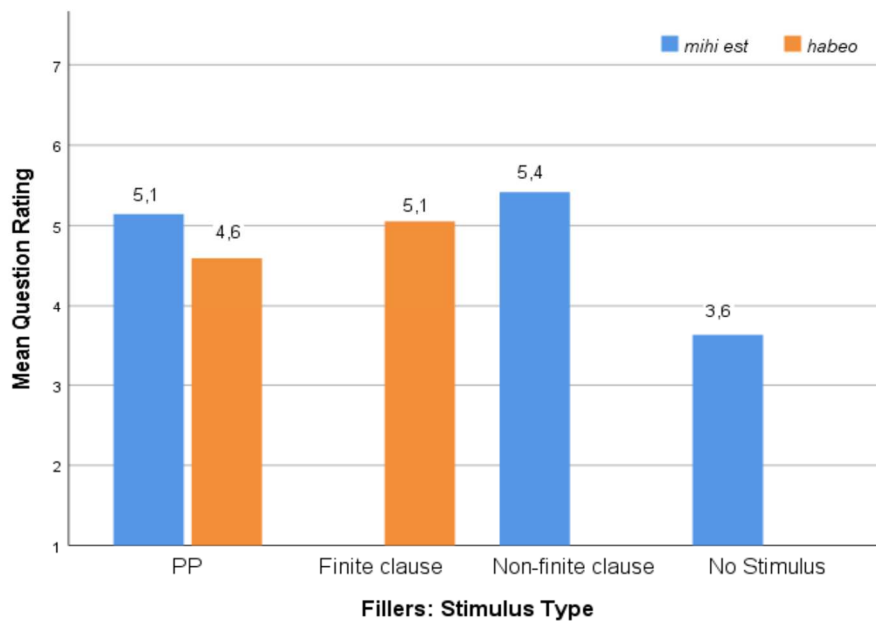
(b)

Figure 8.10 Stimulus type and its impact on the average acceptability rating. Test nouns (a) vs. Control nouns (b)

These data show that, in the case of the test sentences, the sentences without a stimulus are as highly rated as the ones with a stimulus (cf. Figure 8.10a), whereas control sentences with a stimulus are higher rated than the ones without a stimulus (cf. Figure 8.10b). Furthermore, test sentences containing the MIHI EST construction tend to be much lower rated when they occur with a finite clause, whereas the sentences containing the HABEO construction receive higher ratings when the stimulus is realized by a PP, or when it is not at all realized.

To conclude, in this section I investigated the impact of the linguistic variables on the acceptability ratings. In general lines, the linguistic variables yield small but important differences in mean ratings, with respect to sentences having the MIHI EST construction. Based on the collected data, higher ratings are associated with nouns with etymology "Other", that is, Greek, Hungarian, or unknown etymology, and with those expressing a positive event. While derivation seems not to have any significant influence on the average ratings, the stimulus realized by a finite clause causes considerably lower average ratings in test sentences containing the MIHI EST construction.

### 8.1.3.4   Effect of the extralinguistic variables on the acceptability rating

In this section, I investigate whether the acceptability rating is affected by a series of extralinguistic variables, related to the participants: Gender, Age, Region, Education level and Specilization. Table 8.7 presents some descriptive statistics of rating depending on these extralinguistic variables, and on the two constructions: MIHI EST and HABEO. The reported frequencies (Freq) are for individual participants, while N refers to the total number of observations related to the levels describing each variable (for instance, there are 124 female participants who gave a total number of 9 920 ratings; each participant rated 80 sentences, thus resulting 124 x 80 = 9 920).

Overall, the mean and median ratings for both constructions are situated around the acceptability threshold of 4. More precisely, the ratings for the sentences containing the MIHI EST construction are just below the acceptability level, whereas the ratings for sentences containing the HABEO construction are just above this level. In the MIHI EST construction, which is of interest in this study, higher mean ratings seem to be associated with males and PhD-level participants. Note, however, that these two groups contain only ten participants (nine males, of which one has a PhD, and one female PhD). In what follows, I will take a closer look at the correlation between some of the extralinguistic variables and the acceptability rating.
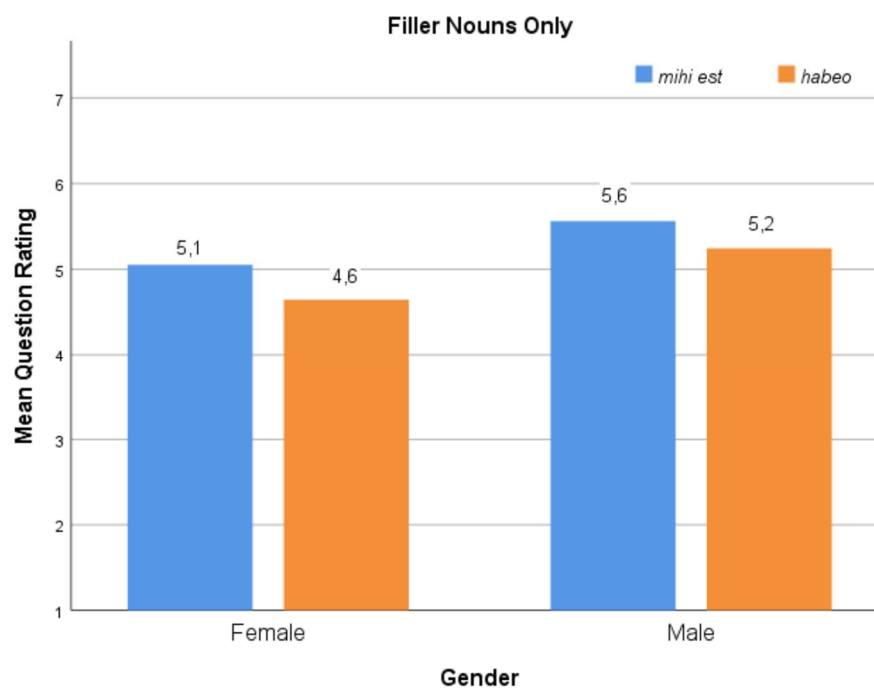
Table 8.7   Descriptive statistics for rating conditional on the extralinguistic variables (MIHI EST VS. HABEO)

Variable and levels (Freq , %)	Min	Mean (SD) MIHI EST	Mean (SD) HABEO	Median MIHI EST	Median HABEO	Max	N
Gender							
Female (124, 93%)	1	3.0 (2.3)	4.4 (2.5)	2	5	7	9 920
Male (9, 7%)	1	3.6 (2.4)	4.8 (2.2)	3	5	7	720
Age							
15-25 (104, 78%)	1	3.1 (2.3)	4.4 (2.4)	2	5	7	8 320
26-35 (18, 14%)	1	2.7 (2.3)	4.1 (2.5)	1	5	7	1 440
36-45 (58, 6%)	1	2.8 (2.5)	4.9 (2.5)	1	6	7	640
45+ (3, 2%)	1	2.4 (2.3)	3.8 (2.6)	1	4	7	240
Region							
Center (21, 16%)	1	2.8 (2.3)	4.1 (2.5)	1	4	7	1 680
East (67, 50%)	1	3.1 (2.4)	4.3 (2.5)	2	5	7	5 360
South (35, 26%)	1	3.0 (2.3)	4.6 (2.4)	2	5	7	2 800
West (7, 5%)	1	3.1 (2.5)	4.4 (2.6)	2	4.5	7	560
R. Moldova (3, 2%)	1	3.3 (2.4)	4.7 (2.4)	4	4	7	240
Education level							
Other (16, 12%)	1	2.6 (2.4)	4.5 (2.5)	1	5	7	1 280
Bachelor (95, 71%)	1	3.1 (2.4)	4.5 (2.4)	2	5	7	7 600
Master (20, 15%)	1	2.9 (2.3)	3.5 (2.5)	2	3	7	1 600
PhD (2, 2%)	1	3.3 (2.0)	5.1 (1.9)	3	6	7	160
Specialization							
Other fields (78, 59%)	1	3.1 (2.4)	4.4 (2.4)	2	5	7	6 240
Other languages (21, 16%)	1	3.1 (2.3)	4.4 (2.5)	2	5	7	1 680
Romanian language (34, 26%)	1	3.0 (2.4)	4.4 (2.5)	2	5	7	2 720

The two graphs in Figure 8.11 below illustrate the impact of Gender on the way the sentences in this survey have been rated. The (a) graph presents the test sentences, whereas the (b) graph illustrates the control sentences. Each of the following graphs illustrate also the competition between the two constructions: the blue bar gives the data for the MIHI EST construction, whereas the orange bar gives the data for the HABEO construction.
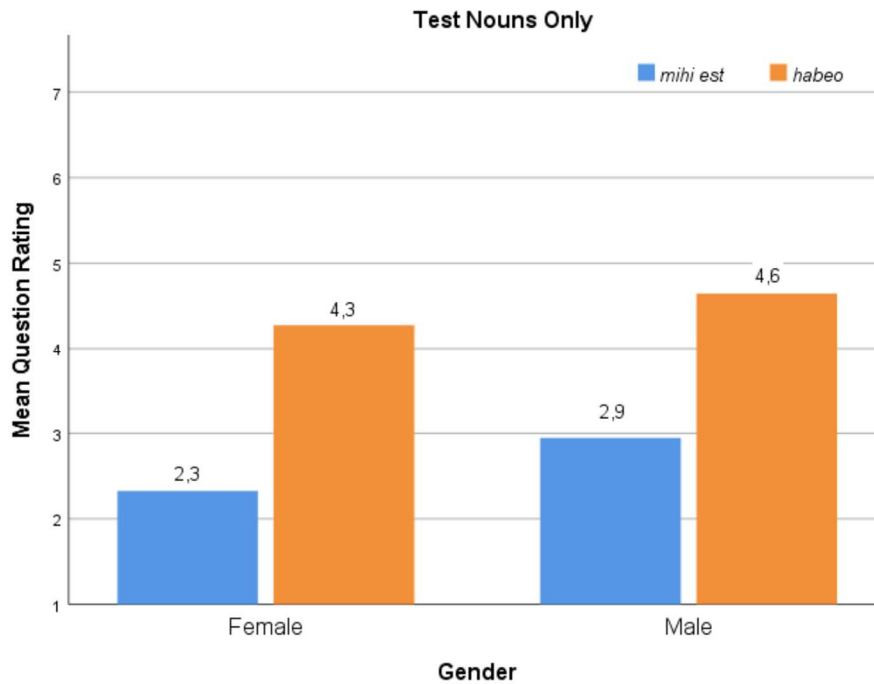
(a)

(b)

Figure 8.11 Impact of Gender on the acceptability rating. Test nouns (a) vs. Control nouns (b)

Note that these graphs are in line with the descriptive statistics presented in Table 8.7 above, where males seem to be associated with higher ratings than females. Furthermore, the competition between the MIHI EST and the HABEO construction is not affected by the gender variable. Males, as much as females, rate higher the HABEO construction in test sentences, whereas in control sentences the MIHI EST construction receives the highest ratings.

Figure 8.12 illustrates, by means of bar charts, how the average rating per question changes depending on the age of the participant, and this, for each of the two constructions of interest: the MIHI EST (blue bars), and the HABEO construction (orange bars).

(a)
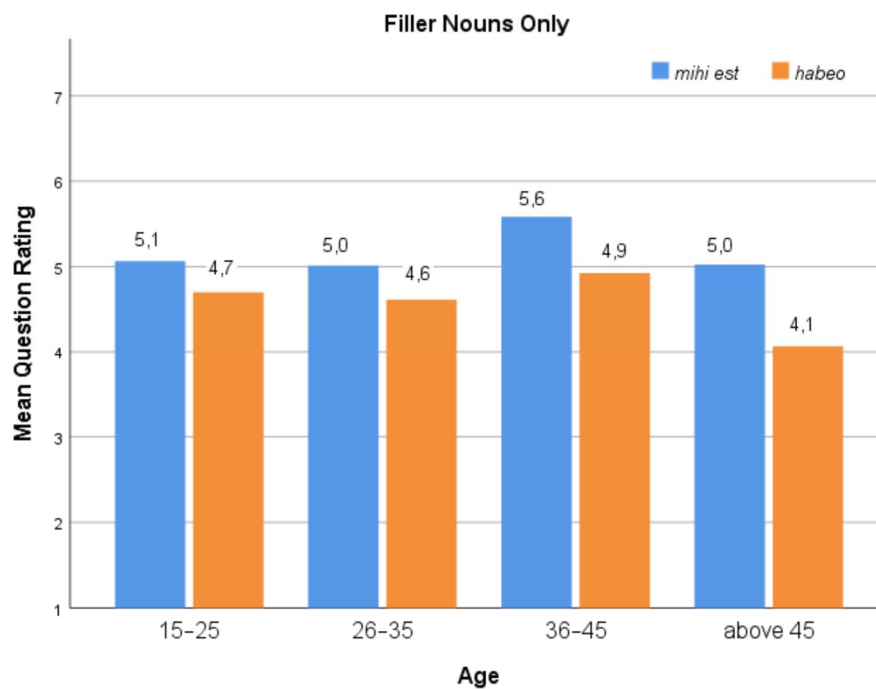
(b)

Figure 8.12 Impact of Age on the acceptability rating. Test nouns (a) vs. Control nouns (b)

As the first graph shows (cf. Figure 8.12a), the rating received by the test sentences containing a MIHI EST construction indicates a certain tendency for younger participants to be more permissive and to give more nuanced ratings, whereas the older participants are stricter in their rating. As for the control sentences, those containing the MIHI EST construction receive higher ratings than the ones with the HABEO construction, regardless of the age of the participants, as evident from the (b) chart in Figure 8.12.

Another interesting aspect to be investigated is whether there is any correlation between the acceptability rating and the region of provenance of the respondent. As described above, the participants originate from four different regions in Romania, namely Eastern-Romania, Southern-Romania, Western-Romania and Central-Romania, and a few of them are from the Republic of Moldova, also native speakers of Romanian.

The data show that three of these regions – although not all of them geographically adjacent – pattern together with respect to their high average ratings, namely Western-Romania (in the West), and Eastern-Romania, and the Republic of Moldova (in the North-East), the last one having also the highest ratings. Furthermore, the participants originating from Western-Romania have been the most generous in their rating for the test sentences containing the MIHI EST construction. Although in a relatively small number, namely seven (5,3 % of the total number of respondents), the participants representing Western-Romania have given three test sentences scores above the acceptability level of 4, as displayed in Table 8.8, below. The same tendency can be observed for the respondents from Moldova, which have given ten test sentences, containing the MIHI EST construction, scores equal with, or above the threshold of 4. Table 8.8 on the next page displays the situation for both regions, Western-Romania and Moldova. Mean refers to the average rating per question.

As for the other two regions, namely Southern-Romania, and Central-Romania, they show very similar ratings for the test sentences containing the MIHI EST construction, which are rather conservative. Among them, the participants originating from Central-Romania seem to be the strictest in their rating, with the lowest average ratings.

Table 8.8   Test sentences rated above the acceptability level of 4 by participants from Western-Romania and Moldova

	Region	Mean
	Western-Romania	
Q18	*Mi-e* tihnă *lângă tine.*	5,71
Q44	*Îi fu* îndurare *văzând atâţia copii nenorociţi.*	4,86
Q56	*Mi-e* amărăciune *în suflet fără a avea un motiv întemeiat.*	4,14
	(Republic of) Moldova	
Q49	*Mi-e* zbucium *deja de-o lună.*	6
Q32	*Mi-e* ranchiună *când îi văd împreună.*	5
Q54	*Mi-e* temere *de moarte.*	5
Q56	*Mi-e* amărăciune *în suflet fără a avea un motiv întemeiat.*	5
Q13	*Mi-a fost mare* chin *înainte de a-l revedea.*	4
Q15	*Mi-e* tângă *fără să ştiu de ce.*	4
Q18	*Mi-e* tihnă *lângă tine.*	4
Q29	*Mi-era o* mânie *grozavă înainte să văd raportul de vânzări.*	4
Q39	*Mi-e* ranchiună *auzind cum îl laudă părinţii.*	4
Q65	*De obicei mi-e* tihnă *înainte de a intra în sala de examen.*	4

In the graphs in Figure 8.13 below, these tendencies are illustrated by means of bar charts. The competition between the two constructions involved, MIHI EST (blue bars) and HABEO (orange bars), is also represented. As opposed to the situation for the test sentences containing the MIHI EST construction described above, the test sentences with the HABEO construction have received, as expected, higher ratings from all the participants.


(a)

Filler Nouns Only
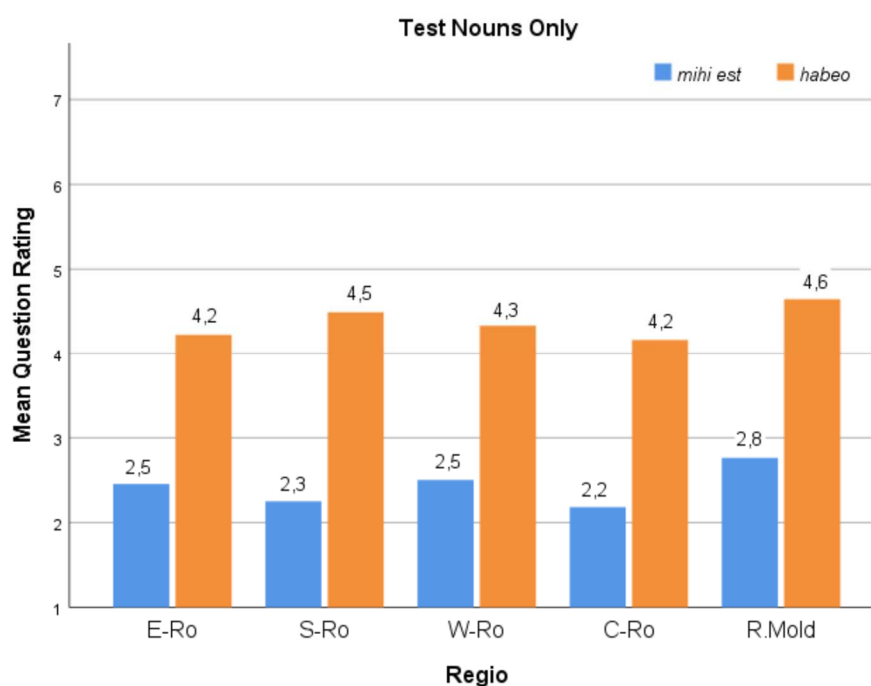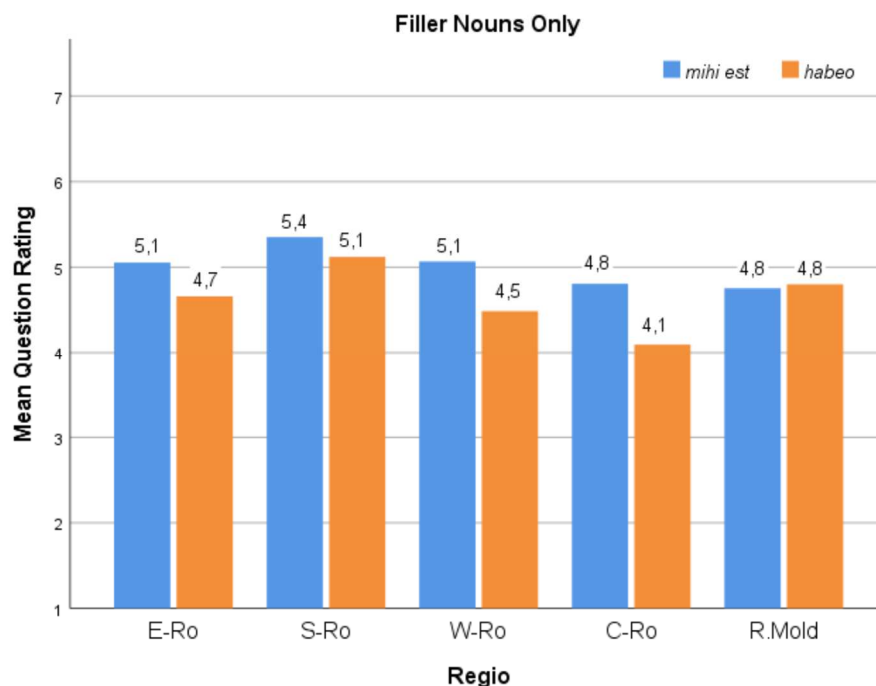
(b)

Figure 8.13 Region and its impact on the average acceptability rating. Test nouns (a) vs. Control nouns (b)

As for the control sentences, the ones containing the MIHI EST construction are preferred by the respondents from all the regions of Romania, with the highest ratings received from the participants from Southern-Romania. Note that the respondents from Moldova give the control sentences with the MIHI EST construction a slightly lower score (4,75) than the ones with the HABEO construction (4,8), which may suggest that not all the selected control nouns are fully acceptable in the MIHI EST construction; in this region, *necaz* 'rancor' and *târşă* 'disgust, horror' show up with very low ratings.

Another factor that may control the acceptability rating is the Professional specialization of the participants. As mentioned in Section 8.1.1.2 above, participants with different educational background were targeted, without any specific knowledge in the field of linguistics. Therefore, among the respondents, 26 % have specific knowledge of Romanian language and literature, 16 % have specific knowledge of modern languages, such as English, French, Spanish, or German, whereas 59 % of the participants have a different educational background such as, for instance, international relations, cultural studies, translations, biochemistry, sociology, or theology. The interaction between the specialization of the participants and the average acceptability rating has been plotted in Figure 8.14.

(a)
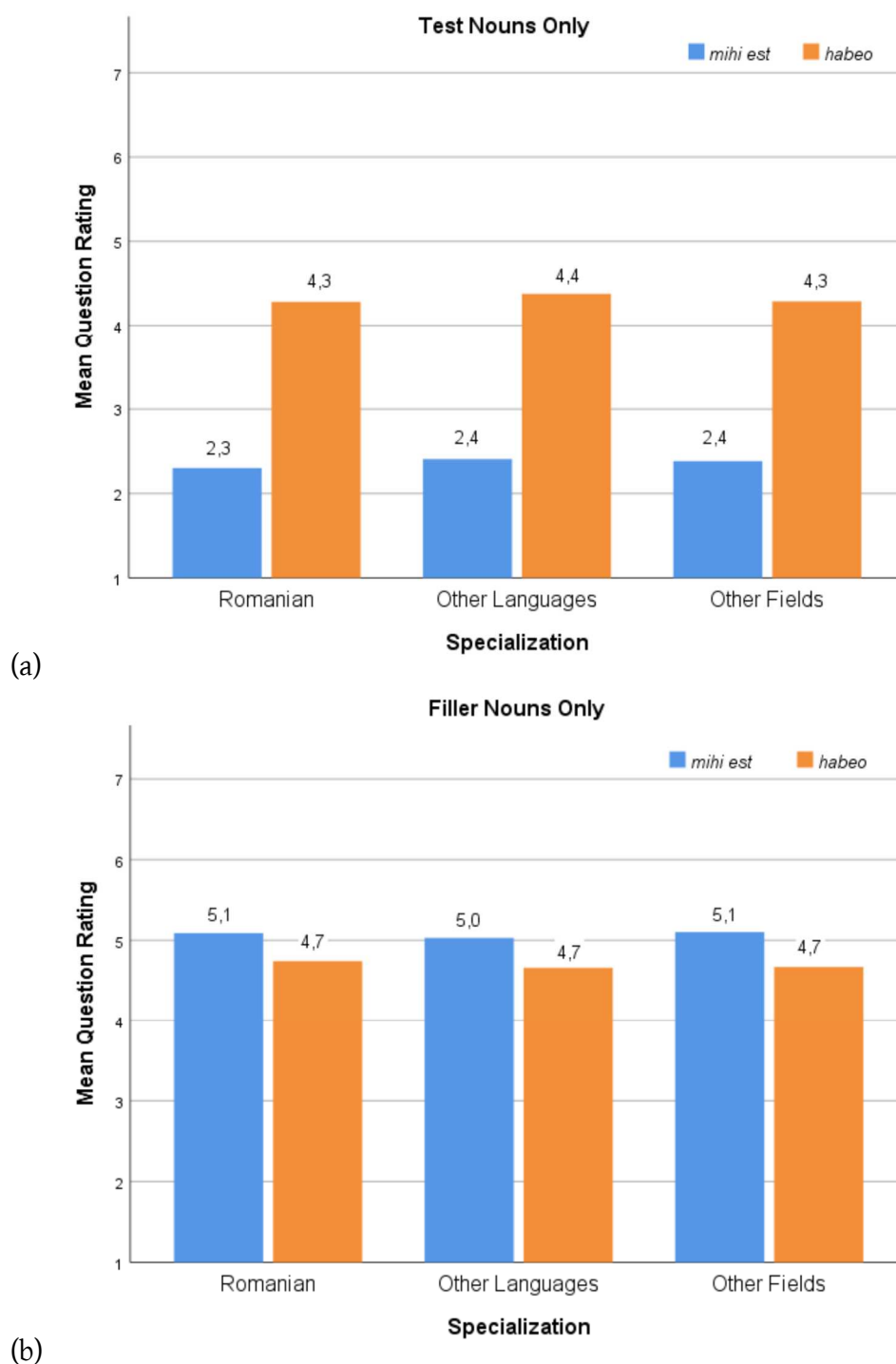
(b)

Figure 8.14 Specialization of the participant and its impact on the acceptability rating. Test nouns (a) vs. Control nouns (b)

Based on the collected data, the graphs in Figure 8.14 show that, in general lines, the educational background of the participants has a limited impact on the acceptability rating. As in the case of other extralinguistic variables, the test sentences containing the MIHI EST construction are lower rated than the test sentences containing the HABEO construction. However, at a closer look, a tendency can be observed, namely that the respondents studying Romanian language on any level, bachelor, master or doctorate, seem to be slightly stricter than other participants in their rating for the test sentences

containing the MIHI EST construction (cf. Figure 8.14a). However, statistically, there is no significant difference between the average ratings of participants with different educational backgrounds. As for the control sentences, the ones containing the MIHI EST construction receive slightly higher ratings than those with the HABEO construction from all the participants.

To conclude, the examined extralinguistic variables seem to have, to some extent, an impact on the acceptability rating of the test nouns occurring in the MIHI EST construction. Based on the collected data, higher acceptability ratings for the test nouns occurring in the MIHI EST construction are associated with males and younger respondents, which tend to be more permissive and to give a more nuanced rating than older participants. Similarly, the occurrence of the test nouns in the MIHI EST construction is rated higher by participants from Western- and Eastern-Romania and from Moldova, whereas respondents from Central-Romania tend to be stricter in their rating. As for the Professional Specialization of the respondents, the data show that those studying Romanian language tend to discard the occurrence of the test nouns in the MIHI EST construction, giving lower ratings than the rest of the participants.

8.1.4 Preliminary conclusions

In Section 8.1, I have reported on the results of the survey carried out among Romanian native speakers, with the intention to verify whether a series of nouns that was not attested in the MIHI EST construction is considered as acceptable in this construction by native speakers of Romanian. As described in the methodology section, the acceptability level has been established at 4, based on a 7-point Likert acceptability scale.

Bearing in mind that the sentences containing the construction under scrutiny with test nouns are not attested in spoken or written sources, the most realistic expectation is that all the test sentences containing the MIHI EST construction will be rated unanimously with 1, hence, as totally unacceptable. This expectation is not borne out, since the distribution of the scores for the nouns entering in the MIHI EST construction shows that every single test noun has been rated by at least one participant with 7, the maximum score, and the mean acceptability ratings for these nouns are situated between 1,5 for the least acceptable noun, *descurajare* 'despair, demoralization', and 3,0 for the most acceptable noun, *amărăciune* 'grief, sorrow'. Moreover, following Keller, Lapata & Ourioupina (2002) and Keller and Lapata (2003), who show that infrequent lexical-syntactic combinations are judged less acceptable by native speakers, it becomes clear that the results of the present survey study go far beyond my initial expectation.

Furthermore, the data show that the control nouns occurring in the MIHI EST construction have not all been rated above the threshold of 4, but have also been rated, at least once, with the minimum as well as with the maximum score, with average

acceptability ratings between 4,0 and 6,6. Hence, each of the selected nouns provoked extreme ratings. In order to understand what could influence the decision of the speakers in their rating, a bimodal analysis of linguistic and extralinguistic factors has been performed.

The impact of a selection of linguistic variables on the acceptability rating has been investigated, among which Etymology, Derivation of the Noun, Polarity of the event, and Stimulus type. The test nouns are in general, and as expected, considered less acceptable when they occur in the MIHI EST construction, than when they occur in the HABEO construction.

Zooming in on the test nouns occurring in the MIHI EST construction, the acceptability ratings tend to be considerably higher among nouns with etymology "Other", that is, Greek, Hungarian, or unknown etymology, and among those expressing a positive event. Derivation does not have a noteworthy influence on the average ratings, whereas test sentences with the MIHI EST construction containing a stimulus realized by a finite clause receive much lower average ratings compared to the ones containing other types of stimulus, or even with no stimulus at all.

Similarly, a selection of extralinguistic variables related to the respondents has been studied, among which Gender, Age, Region, Education level and Professional specilization. These variables have been examined for the test sentences as opposed to the control sentences relative to the construction in which the nouns occur. Overall, the test sentences containing the MIHI EST construction have received lower scores than the ones containing the HABEO construction. A closer investigation of the sentences containing the MIHI EST construction reveals that higher acceptability ratings correlate with participants younger than 45 years, and with respondents originating from Western- and Eastern-Romania and from the Republic of Moldova. Stricter in their ratings are those participants specialized in Romanian language, regardless of the level of their education, although this is not statistically significant, and the participants aged more than 45 years.

This study reveals a tendency which, to some extent, is in line with the claim that the MIHI EST construction is expanding in present-day Romanian (Ilioaia 2020), namely that even in speaker's usage, there is a propensity toward allowing new – in this case, synonymous – nouns in the MIHI EST construction. The fact that the average ratings obtained by means of this survey do not pass the specified acceptability threshold, indicates, however, that the tendency toward innovation of the MIHI EST construction in present-day language is conditioned by several factors. As shown in this study, a tendency toward innovation can be observed with younger speakers, having preferably, other specializations than Romanian language, and originating from any region of Romania, except Central-Romania. This tendency should be verified through a much broader study, involving a more balanced number of respondents, of different ages and from different backgrounds.

8.2 Productivity in the corpus

This section focuses on the dynamics of the usage of the MIHI EST construction in Romanian across the centuries. The productivity of the MIHI EST construction is defined by the number of nouns entering into this construction and by their frequency. Two important research questions are addressed: (i) how has the degree of productivity of the MIHI EST pattern evolved since the first attested Romanian texts?, and (ii) is the MIHI EST pattern expanding or retracting in productivity in present-day Romanian?

The research presented in this section is based on a dataset, extracted from the corpus described in chapter 5, containing data from both present-day and pre-21st century Romanian. Note that the dataset for present-day Romanian contains a substantial number of relevant examples of the MIHI EST construction, whereas the dataset for pre-21st century Romanian consists of a more restricted number of examples due to the scarcity of texts available for these periods. In the past years, research has shown that, when dealing with small corpora, a multidimensional approach containing quantitative (based on measurable data) as well as qualitative (based on observation, descriptive) methods is necessary in order to measure and evaluate the degree of productivity of a pattern (Baayen 2009; Zeldes 2012). This is, indeed, the way I investigate the Romanian data. A quantitative analysis, based on three aspects of productivity outlined by Baayen (2009),[4] is to be complemented by a qualitative analysis along the lines of Barðdal (2008).

In what follows, Sections 8.2.1 and 8.2.2 introduce the two approaches that are used, pointing out a few methodological peculiarities. Section 8.2.3 discusses the findings of this research. Finally, Section 8.2.4 concludes by formulating an answer to the two research questions.

8.2.1 Defining the rate of use

The present study aims to estimate the productivity of a syntactic structure, more specifically, the MIHI EST construction. In order to estimate syntactic productivity, which does not have yet a generally accepted definition, several methods have been proposed in the literature. For Barðdal (2008: 1), for instance, syntactic productivity represents a construction's extensibility, that is, its ability to attract new or existing lexical items. Barðdal's (2008) proposal stems from Goldberg (1995) and Bybee & Thompson (1997),

---

[4] The three different aspects of productivity he introduces – realized, expanding, and potential productivity – are defined according to the measuring method used to estimate each of them.

which consider that syntactic productivity is favored by a high number of different items attested in the relevant slot of a construction, hence, by a high *type frequency.*

Given the size of the corpus available for Romanian, a combined approach containing quantitative as well as qualitative methods is applied in order to estimate the degree of productivity of the MIHI EST construction. The quantitative analysis, which is presented in this section, is inspired by Baayen's (2009) insightful study, in which he identifies three different aspects of productivity. These are: realized productivity, representing the extent of use of a specific pattern (Baayen 1993); expanding productivity, indicating the rate at which a pattern expands by means of attracting new members (Baayen 1993); and potential productivity, defined as the rate at which the vocabulary increases (Baayen & Lieber 1991; Baayen 1993).

The qualitative analysis is based on Barðdal (2008), who suggests that syntactic productivity, understood as extensibility, is a function of type frequency, semantic coherence, and an inverse correlation between the two.

8.2.2   Methodology

8.2.2.1   The dataset

For the present study of productivity, a separate dataset has been gathered, different from the one used in Chapter 7. The nature of the present study and the specificity of the measurements used in this study required the compilation of a separate dataset, more suitable for the selected methodology. Therefore, using the advanced query option in Sketch Engine, examples containing the verb *fi* 'be', preceded by a dative clitic and followed by any noun, with one to two lexical items allowed to occur between the verb and the noun, have been retrieved.[5] Such structures have been collected from the corpus for the 21st century, *roTenTen16*, as well as from the corpus for the pre-21st century Romanian, created by me on the Sketch Engine platform, as reported on in Section 5.3, in Chapter 5.

This query gave 247 001 hits for present-day Romanian, which I have restricted to a random sample of 10 000 examples. The same query, executed in the corpus of the pre-21st century Romanian, returned 1 360 hits for the 16th-18th centuries (old Romanian), 1 053 hits for the 19th century (modern Romanian), and 1 351 hits for the 20th century (contemporary Romanian), hence a total of 3 764 examples, which were all preserved. After the elimination of doubles, the dataset contains 9 996 examples for the 21st century and a total of 3 662 examples for the pre-21st century periods, taken together. Out of these totals, I have been able to create four sets containing only the instances of MIHI EST, found

---

[5] Note that other possible nouns occurring between the verb and the noun are excluded by the CQL-query.

for each historical period. Table 8.9 gives an overview of the sets for each historical period, followed by their percentage related to the total number of returned examples for each period.

Table 8.9   Overview sets of relevant examples per century (16th-21st centuries)

Period	Examples	%
16th–18th	130	10%
19th	359	34%
20th	616	49%
21st	4 955	50%
Total	6 060	44%

The annotation phase of these sets consisted in deciding which structures are an instance of the MIHI EST construction and which not. All the other structures were simply labeled as *noise*, since their more specific identification was not relevant for the present stage of this research. Other variables such as case, period, semantic field, and polarity were annotated in separate columns.

8.2.2.2   Measurements

In order to generate answers to the research questions addressed in this section, namely whether non-canonical marking of core arguments expands in present-day Romanian, and what the shifts are to which this construction has had to adapt across the centuries, a set of productivity measurements are applied to the gathered dataset.

For this purpose, the following tests are applied to the Romanian data. First, token frequency is calculated. As stated in Croft (2001: 28) and in Goldberg (2006: 93), it helps us understand the degree of entrenchment of a specific construction in a given language. Next, type frequency is considered. The type-counting approach, in spite of its criticism (Aronoff 1976: 36), has been directly associated to productivity by several scholars (Goldberg 1995: 134; Anshen & Aronoff 1999; Croft & Cruse 2004: 308-313; Bybee 2006; Barðdal 2006b; Barðdal 2008). Moreover, type frequency is utilized by Baayen (2009) in his multidimensional approach as a measurement for realized productivity, which is, alongside with expanding and potential productivity, one of the three different aspects of productivity identified by him.

In Baayen's view, realized productivity can be gauged by counting the number of types in different historical periods, as suggested by Anshen & Aronoff (1999), or through the structural type distribution measurement proposed by (Baayen 2001). This second method is not part of the present study. Instead, potential productivity is estimated by means of two interdependent tests: potential productivity and global productivity (Baayen & Lieber 1991, Baayen 1993). As for the third aspect of productivity, expanding

productivity, it has been excluded from this study for practical reasons related to feasibility.[6]

Given the availability and the nature of the Romanian data, an exclusively quantitative approach may lead to incomplete or even incorrect conclusions. Therefore, the outcomes of the quantitative analysis of the Romanian data available for this study will be checked in a qualitative analysis based on the approach suggested by Barðdal (2008).

In her book on syntactic productivity, Barðdal (2008) suggests that the productivity of a syntactic construction should be analyzed as a function of its type frequency, its semantic coherence, and an inverse correlation between the two. Type frequency and semantic coherence (also identified as the opposite of the concept of variability put forward by Goldberg 1995) have both been previously mentioned in the literature as affecting productivity (Bybee 1995: 430; Goldberg 1995; Bybee & Thompson 1997; Clausner & Croft 1997; Barðdal 2001; Barðdal 2006b; Clausner 2002; Croft & Cruse 2004, amongst others). However, Barðdal (2008) is the first scholar who suggested a systematic link between the two as a predictor for productivity.

8.2.2.3 Methodological difficulties

The above-mentioned measurements, although designed for morphological processes, have been successfully applied to the field of syntax (Barðdal 2008; Zeldes 2012). Nevertheless, when the application of these methods interacts with certain factors, important methodological issues arise.

The first issue is related to using datasets of different sizes. As pointed out by Bauer (2001:149), and generally acknowledged in the field, the productivity index of a process drops as a function of increasing sample size. Note that the gap between the sizes of the datasets for the pre-21st century and the 21st century is significant in Romanian, the latter dataset being approximately 4.5 times larger than the former. Therefore, I rely on normalized values of frequencies calculated per million words, in order to support any comparative remarks between the different centuries. For the normalization of the values, I divide the type-count by the word-count of the whole dataset (noise included), then I multiply by one million.

Another way to solve this problem is working with equally sized samples for each of the historical periods. Although this solution is widely suggested in the literature, it is not

---

[6] Expanding productivity, known as the *hapax-conditioned degree of productivity* (Baayen 1993), is estimated by calculating the ratio between the total number of hapax legomena occurring in the construction and the total number of hapax legomena in the whole corpus. Because of the nature of my corpora, finding the total number of hapax legomena in the whole corpus requires an enormous amount of manual work. For this reason, I chose to exclude this measurement from my analysis.

feasible with the present dataset, since for the oldest period of Romanian I work with a considerably small dataset containing not more than 130 relevant examples.

Yet, these solutions are not able to cover the enormous gap between the amount of data for the pre-21st century Romanian and for present-day Romanian. Multiple other factors are involved, such as heterogeneity of the data with regard to text genre or the medium of spreading these texts. As a matter of fact, the data for the pre-21st century Romanian come from printed literary, administrative or religious texts, whereas the data for the present-day language come from informative, technical and commercial texts, as well as literary or religious texts spread via internet, or from blogs and forums.

Therefore, a quantitative analysis will be conducted for the four historical periods, without necessarily contrasting the results for the first three historical periods with the ones for present-day Romanian, since a comparison between the two stages does not seem to be constructive. Nevertheless, the two periods will be compared by means of a qualitative approach.

8.2.3 Findings

This section presents the results obtained when applying the measurements described in Section 8.2.2.2 to the available Romanian data. In the first part of this section, I observe the fluctuations in the degree of usage of the MIHI EST pattern across the centuries from a quantitative perspective, whereas the last part projects a qualitative outlook on these facts.

8.2.3.1 Toward a quantitative analysis of the MIHI EST construction

In this section, changes in productivity of the MIHI EST construction are observed across the centuries. I attempt to find an answer to the first research question of this study, namely how has the degree of productivity of the MIHI EST pattern evolved since the first attested Romanian texts. Hence, I first apply the token-count test on the Romanian data, then the focus shifts toward realized and potential productivity, as defined by Baayen (2009).

8.2.3.1.1 Token frequency

The importance of token frequency lies in identifying the storage or the entrenchment of a specific construction in the lexicon (cf. Croft 2001: 28, Goldberg 2006: 93). The application of the token frequency measurement to the Romanian data has yielded the results presented in Table 8.10. The figures have been normalized per million words in order to obtain comparable results. The table shows the absolute and the normalized token frequency in the dataset, for each period.

Table 8.10   Token frequency of the MIHI EST construction (16th-21st centuries)

	Token frequency absolute	Token frequency normalized/mln
16th–18th	130	1 886
19th	358	6 278
20th	616	8 792
21st	4 988	10 526

Both the absolute and the normalized figures show that this construction has increased in token frequency throughout the centuries. Put differently, the pattern has become more and more entrenched in Romanian. Although the differences in frequency are considerable, a z-test used to check the significance of the difference between two independent proportions shows that the increase in token frequency of the MIHI EST construction between the oldest period and the 20th century is strongly significant, based on both the one-tailed and the two-tailed P-values (Z = -17,62; one-tailed P-value < 0,0001; two-tailed P-value < 0.0002).[7] The graph in Figure 8.15 below, visualizes this growth in degree of entrenchment of the MIHI EST pattern throughout the centuries.
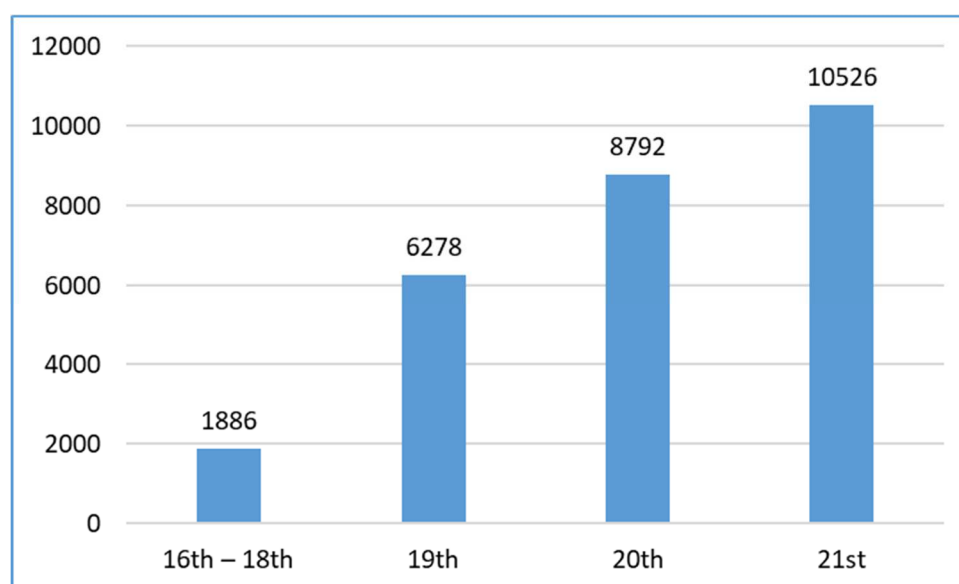
Figure 8.15 Token frequency MIHI EST construction (16th-21st centuries)

---

[7] An online z-test has been performed using the tool on the following website: [12 March 2020] http://vassarstats.net/propdiff_ind.html. The z-test for independent proportions compares two independent proportions. Also known as the *t-test for independent proportions*, or as the *critical ratio test*, it is used in medical research to calculate the risk difference. The z-test is identical to the chi square test, except that it estimates the standard normal deviate (z). Put differently, the square of the test statistic (z^2) is identical to the Pearson's chi square statistic X^2. When the focus is on the size of the difference between two proportions, the z-test is preferred to the chi square test (https://influentialpoints.com/Training/z-test_for_independent_proportions_use_and_misuse.htm).

Furthermore, the increase in token frequency between the 20th and the 21st centuries has also been shown to be strongly significant (Z = -4,243; two-tailed P-value < 0.0002). A comparison between the pre-21st century period and the 21st century has to be considered cautiously, since, as explained in Section 8.2.2.3, there are significant differences in size between the datasets for the two stages in Romanian, and, perhaps more importantly, great differences in text types.

8.2.3.1.2 Realized productivity or type frequency

The type-count measurement has first been criticized as an indicator for morphological productivity by Aronoff (1976: 36), who pointed out that, in a given language, for some of the affixes there might be too few bases available in order to form new types. Nonetheless, a few years later, the same scholar proposes the type-count approach as a reliable way to gain insight into the productivity of a structure (Anshen & Aronoff 1999). Moreover, this measurement has been validated also by Baayen (2009), among others, who recommends it as a way of gauging realized productivity.

Also called *extent of use*, the realized productivity estimated by type-count is restricted to *past achievement* in Baayen's view (Baayen 1993). Anshen & Aronoff (1999) used the type-count approach in their study to measure the extent of use of a morphological category C, by taking into account the number of types in a corpus with N tokens V(C,N) at a given point in time.[8] This measure is applied here to Romanian data to estimate the past achievement of the MIHI EST construction for each historical period. In order to generate comparable figures, frequencies are normalized.

Therefore, I have counted the different nouns occurring in the MIHI EST construction in the Romanian dataset for each documented historical period. The type-counts have then been normalized per million words, and are given in Table 8.11, besides the absolute type-counts for each historical period of Romanian.

Table 8.11   Type frequency of the MIHI EST construction (16th-21st centuries)

	Type frequency Absolute	Type frequency normalized/mln
16th–18th	13	189
19th	21	368
20th	24	343
21st	42	89

---

[8] The following abbreviations have been used in the remainder of this chapter: C = the category under scrutiny (the MIHI EST construction); N = the number of tokens in my dataset (i.e. the total number of examples instantiating the MIHI EST pattern, noise included); V = number of types of the nouns filling the state slot in the construction (i.e. the number of distinct nouns entering in the MIHI EST construction); V(1,C,N) = the number of the *hapax legomena*.

The normalized figures in Table 8.11 reveal an increase in type frequency in the 19th century compared to the oldest period, followed by a slight decrease in the 20th century. Whether this decline is statistically significant or not, is difficult to determine, since the proportions to be compared represent relatively small figures, 189, 368 and 343 per million, respectively. For more precision, a z-test has been conducted to verify both whether the extent of use of this construction has significantly increased from the first written texts until the 20th century, and whether the decrease revealed in the 20th century is statistically relevant. The z-test shows that the extent of use of the MIHI EST construction has increased between the earliest period and the 20th century, but not significantly, based on the two-tailed P-value (Z = -1,759; one-tailed P-value < 0,039; two-tailed P-value < 0.079). As for the slight decrease in extent of use of this construction between the 19th and the 20th century, although not surprising, the z-test clearly reveals that the difference is not statistically significant (Z = 0.243; one-tailed P-value < 0,404; two-tailed P-value < 0.808).
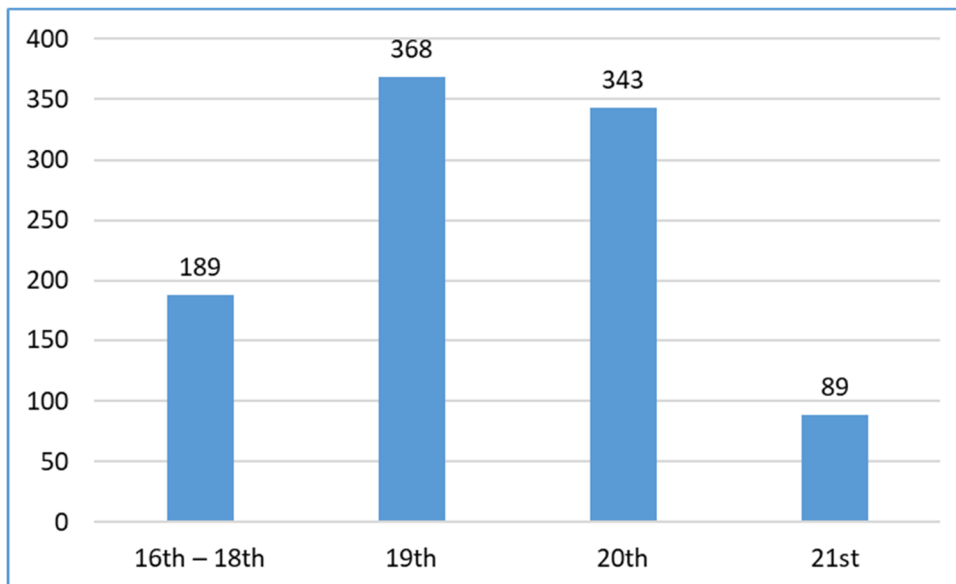
Figure 8.16 Realized productivity MIHI EST construction (16th-21st centuries)

The results of the statistical tests are illustrated by the graph in Figure 8.16. This graph indicates that the realized productivity increases in the 19th century but shows a slight decrease in the 20th century. As explained in Section 8.2.2.3, I have chosen not to compare the results of the measurements for the pre-21st century period with the ones for the 21st century, due to significant differences between the datasets. Hence, the substantial decline suggested for the 21st century in Figure 8.16 should be interpreted cautiously.

8.2.3.1.3   Potential productivity

Potential productivity, another aspect of productivity in Baayen's (2009) approach, can be estimated by means of two different measurements: potential productivity and global

productivity, proposed by Baayen & Lieber (1991) and Baayen (1993). The two measurements must be used together, as recommended by the two scholars, since none of them provides independently enough information on the potential productivity of a structure.

The first measurement of potential productivity, also called *category-conditioned degree of productivity* in Baayen (1993), is calculated by dividing the number of the hapax legomena in the dataset V(1,C,N) of the category under scrutiny (C), the MIHI EST construction in my study, by the total number of tokens of the specific construction N(C) in the same dataset (Baayen & Lieber 1991). The obtained ratio estimates the growth rate of the vocabulary of the construction itself: P = V(1,C,N)/N(C).

In order to calculate the potential productivity index of the MIHI EST construction in Romanian, the number of hapax legomena are identified for each period. These numbers are then divided by the already calculated token frequencies. The obtained ratio is a value between zero and one, whith one as the highest potential productivity index. It has to be noted that the process of normalizing the figures, which is the first step in neutralizing the differences in sample size, has not changed the index of potential productivity. Table 8.12 presents the absolute figures for the hapax-count (V) and the token-count (N), as well as the estimated potential productivity index (P), for each period.

Table 8.12   Potential productivity MIHI EST construction (16th–21st centuries)

	Hapax legomena (V)	Token frequency (N)	Potential productivity index (P)
16th–18th	4	130	0,031
19th	1	358	0,006
20th	7	616	0,011
21st	21	4 988	0,004

As shown in Table 8.12, this formula estimates considerable differences in potential productivity among the four historical periods. Based on these figures, no tendency can be identified. The highest potential productivity index (0,031) is observed for the oldest period of Romanian (16th–18th centuries), followed by a strong decrease in the 19th century (0,006). For the following period, the 20th century, the potential productivity index increases to 0,011, but remains much lower than the productivity index for the first period. The index for the 21st century is the lowest among the four stages of Romanian (0,004). Since two of the figures used to determine the potential productivity index are smaller than five, the z-test cannot be calculated in order to validate the statistical relevance of these figures. Figure 8.17 visualizes the variations between the four historical periods.
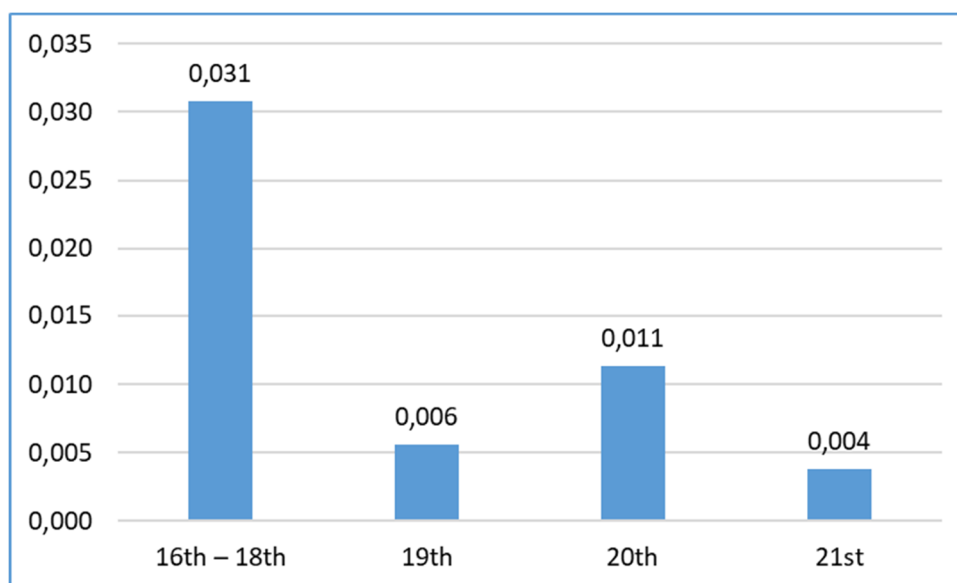
Figure 8.17 Potential productivity (P) MIHI EST construction (16th–21st centuries)

Recall that Baayen & Lieber (1991) point out that the potential productivity measurement must be accompanied by the global productivity measurement (P*),[9] which correlates the hapax-token ratio to the type frequency of the construction under scrutiny. These researchers recommend representing the global productivity on a P-V plane, with the degree of potential productivity (P) on the horizontal axis, and the extent of use, i.e. type frequency (V), on the vertical axis. Table 8.13 gives an overview of the fluctuations in global productivity of the Romanian MIHI EST construction throughout the centuries.

Table 8.13   Global productivity (P*) MIHI EST construction (16th–21st centuries)

	Type frequency (V)	Potential productivity (P)	Global productivity P*(x;y)
16th–18th	13	0,031	(0.031;13)
19th	21	0,006	(0.006;21)
20th	24	0,011	(0.011;24)
21st	42	0,004	(0.004;42)

The graph in Figure 8.18, based on the data in Table 8.13, visualizes the changes in global productivity of the MIHI EST construction over the four historical periods of Romanian. It has to be noted that Baayen & Lieber (1991) expect globally more productive processes to have large values for both type frequency (V) and potential productivity (P), which would, then, be positioned on the upper-right corner on the graph. Similarly, they expect globally unproductive processes to have low values for both type frequency (V), and

---

[9] The term global productivity was initially labeled P* by Baayen himself. Later, he confusingly used the same label for *the hapax-conditioned degree of productivity* (Baayen 1993).

potential productivity (P), and, thus, to be positioned on the lower-left corner on the graph.
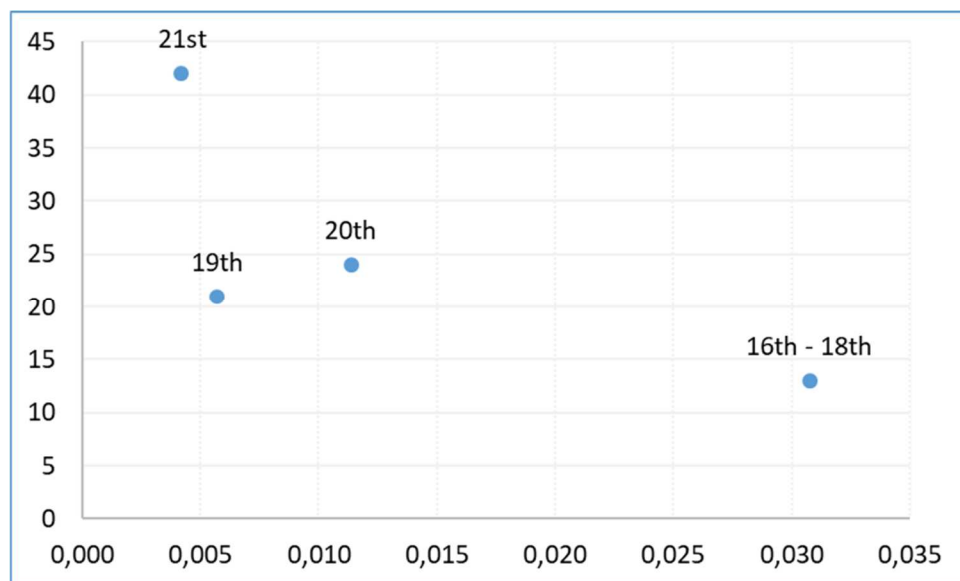

Figure 8.18 Global productivity (P*) MIHI EST construction (16th–21st centuries)

As emphasized by the two scholars, the drawback of this measurement is that, when considering only type frequency (V) or only potential productivity (P), it is impossible to estimate which process is more productive. In the Romanian dataset, the MIHI EST construction has, in the oldest period (16th–18th centuries), a reasonably high potential productivity index on the X-axis, namely 0.031, but a relatively low number of types (V) on the Y-axis, namely 13, what positions it on the right extremity of the X-axis, hence on the lower-right corner on the graph. By contrast, in the 21st century, the construction has scored a very low potential productivity value, namely 0.004, but a fairly high number of types on the Y-axis, namely 42, hence its position on the upper-left corner on the graph. This makes it difficult to draw any accurate conclusions on the potential productivity of the MIHI EST construction throughout the four historical periods of Romanian. Nevertheless, the qualitative discussion that follows in the next section is meant to shed more light on the interpretation of these outcomes.

In this section, I observed the Romanian data through a quantitative approach in order to find an answer to one of the research questions of this study, namely how has the degree of productivity of the MIHI EST construction evolved since the first attested texts until present-day Romanian. By means of an integrated quantitative analysis based on methods which combine three essential pillars for productivity, namely token frequency, type frequency and the hapax-count, I have shown that this construction has a very dynamic productivity pattern throughout the centuries. These indicators have been integrated in a multidimensional approach on productivity suggested by Baayen (2009). In his view, all these aspects are interdependent and, at the same time, each of them contributes to a better understanding of this phenomenon.

The fluctuations observed through the four historical periods of Romanian, characterizing each of these aspects of productivity, cannot be easily explained, since multiple factors interact. As a result, depending on the aspect being measured, contrasting tendencies are observed. Hence, whereas the normalized token frequency suggests a clear increase across all periods, the figures for type frequency show an opposite trend for the last two periods, with a slight decrease in the 20th century, but a significant drop in the 21st century. The potential and global productivity show less clear tendencies. The next section is expected to clarify these contradictory results by observing the productivity of the MIHI EST construction in the most recent texts by means of a qualitative analysis.

8.2.3.2   A qualitative approach to productivity

This section aims to give an answer to the second research question addressed in this chapter, namely whether the MIHI EST construction is expanding or retracting in productivity in present-day Romanian. In order to answer this question, it is important to consider the facts of the present-day language in tight relation with the fluctuations in degree of productivity of this construction throughout the previous centuries. In the preceding section, the variations in productivity of the MIHI EST construction across the four historical periods of Romanian have been observed and commented on diachronically. This section focuses on the present-day language and adopts the approach suggested in Barðdal (2008). In this approach, the productivity of a construction is predictable based on its type frequency, its coherence, and an inverse correlation between the two.

Syntactic productivity is visualized in this analysis on a graph with type frequency on the Y-axis and semantic coherence on the X-axis, as in Figure 8.19.[10] The productivity cline illustrates the inverse correlation between these two, with, at one end, the highest type frequency, and, at the other end, the highest degree of semantic coherence. In other words, full productivity and analogical formations are situated at the opposite ends of the productivity cline, representing "two sides of the same coin" (Barðdal 2008: 3). Barðdal (2008) points out that there are no extensions of non-productive processes, but only different levels of schematicity, influenced by differences in type frequency and degree of entrenchment of the schema. She further explains that there is a tight correlation between the type frequency and the semantic coherence of a given construction. The higher the type frequency of a construction, the lower the degree of semantic coherence is required for a construction to be productive. Conversely, the lower

---

[10] The coherence of a construction is defined as the internal morphological, phonological, or semantic consistency found between all the members of each construction or category. In the specific situation of this kind of syntactic constructions, semantic coherence is considered the most relevant (Barðdal 2008).

the type frequency of a construction, the higher degree of semantic coherence is necessary for a construction to be productive. As for the notion of entrenchment, the concept goes back to Langacker (1987) who postulates a continuous scale of entrenchment in cognitive organization. Every use of a structure has a positive impact on its degree of entrenchment. Units are variably entrenched depending on the frequency of their occurrence, hence, their token frequency.
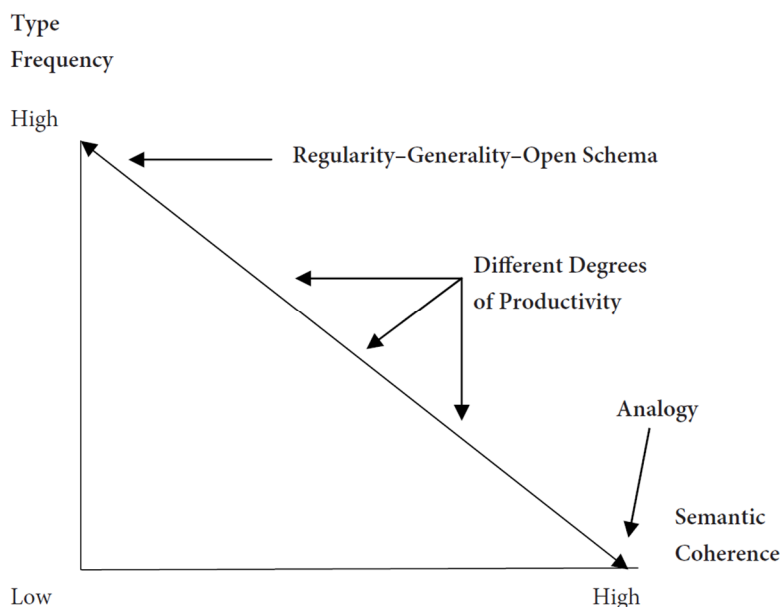
Figure 8.19 Different aspects of the cline of productivity (Barðdal 2008: 38)

In view of the outcomes of the quantitative analysis presented in the previous section, the Romanian MIHI EST construction is to be situated on the lower extremity of the productivity cline based on the data from the first three historical periods, due to its high semantic coherence and its relatively low type frequency (based on Barðdal 2004 and Barðdal et al. 2012). That is, the semantic domain of the nouns occurring in this construction during this period is restricted to a limited number of semantic classes such as emotion and bodily states. As for the type frequency, the quantitative analysis shows that the absolute type-count increased from 13 types in the oldest period to 24 types in the 20th century. It has to be noted, however, that the degree of inovation differs from one period to another. For instance, in each of the three periods, the proportions of frequent and unfrequent nouns are different. Table 8.14 gives an overview of these proportions for all periods. In this table, "Corpus" refers to all hits, noise included, "Tokens" refers to all relevant examples containing the MIHI EST construction, "Frequent types" are considered the types occurring more than twice in the dataset, and "Dislegomena" refers to all types occurring twice in the dataset. The relative figures (%) are calculated by dividing the absolute number of types to the total number of types for each historical period.

Table 8.14   Proportion of frequent and unfrequent nouns – MIHI EST construction

Period	Corpus	Tokens	%	Types	Frequent types	%	Dislegomena	%	Hapax legomena	%
16th–18th	1 355	130	10%	13	6	46%	3	23%	4	31%
19th	1 050	358	34%	21	17	81%	3	14%	1	5%
20th	1 257	616	49%	24	15	63%	2	8%	7	29%
21st	9 996	4 988	50%	42	19	45%	2	5%	21	50%

As evident from Table 8.14, after the earliest period of Romanian, the focus shifts to the entrenchment of the construction from its innovation for two centuries, the 19th and the 20th. This trend can be perceived in the token-count, which nearly doubles with each period, from 130 to 358 in the 19th century, and from 358 to 616 in the 20th century. Moreover, whereas in the earliest period the frequency of the examples containing the MIHI EST construction (token frequency) represents 10 % of the total hits in my corpus, this proportion increases steadily in the following centuries: in the 19th century the tokens represent 34 % of the total hits, whereas in the 20th century the tokens represent already 49 % of the total hits.

These trends continue in the 21st century, where the relevant tokens represent 50 % of the total hits. What is striking, however, is the degree of innovation observed in the present-day language: among the 42 types, 21 are hapax legomena and two types are dislegomena, which means that 55 % of the types are newly entered nouns in this construction.

In the view of these details, I conclude that, throughout the centuries, innovation and entrenchment succeeded each other in the evolution of the MIHI EST construction, having as an outcome a dynamic, slowly increasing productivity degree.

Nevertheless, if the figures for type and token frequency are rather fluctuating across the centuries, one thing is certain, namely that for the first three periods, the semantic coherence is high and stable, the construction allowing combinations almost exclusively with nouns denoting psychological or physiological states. In contrast, in the present-day language, semantic coherence seems to start dissolving and the construction seems to achieve a higher degree of schematicity. This observation is made based on the higher number of types entering this construction, and on the variety of semantic domains from which these nouns are selected, such as time (*noapte* 'night', *toamnă* 'autumn'), elements of nature (*foc* 'fire'), weather (*ger* 'freezing'), matter (*vomă* 'vomit'), and events (*vis* 'dream'). A few cases of such new types are illustrated in the examples in (227)–(231).

(227) *I-*   *era* *mare* *foc* *de* *mine*   (element of nature)
   him.DAT= was big fire of me
   'He cared very much about me'   (1929 - Caragiale, Mateiu - Craii de Curtea Veche)

(228) *Mi-*   *e* *noapte* *[...]* *şi* *mi-* *e* *ger* (time/weather)
   me.DAT= is night [...] and me.DAT= is freezing
   'I feel like night and i am freezing'   (www.escoala.ro)

(229) *[...]* *te*   *mai*   *aud*   *până* *să*   *-ţi*   *fie*   *vis*   (event)

       [...] you.ACC still hear.1SG until SUBJ =you.DAT be.3SG dream

       'I still hear you until you will be dreaming'   (www.cuplari.ro)

(230) *Vomă*   *mi-*   *e*   *de*   *toţi:* *judecători,* *[...]*   (matter)

       vomit   me.DAT= is   of   all   judges[...]

       'They all make me sick: judges, [...]'   (www.hotnews.ro)

(231) *Mi-*   *e*   *toamnă* *iubito,*   *căci*   *[...]*   (time)

       me.DAT= is   autumn   beloved   because   [...]

       'I feel like autumn my love, because [...]'   (www.intelepciune.ro)

The repartition of the variety of semantic fields throughout the four historical periods of Romanian has been shown in Figure 8.20, whereas Figure 8.21 zooms in on the newly accepted semantic classes, showing that these all cluster into the last historical period, namely present-day Romanian.
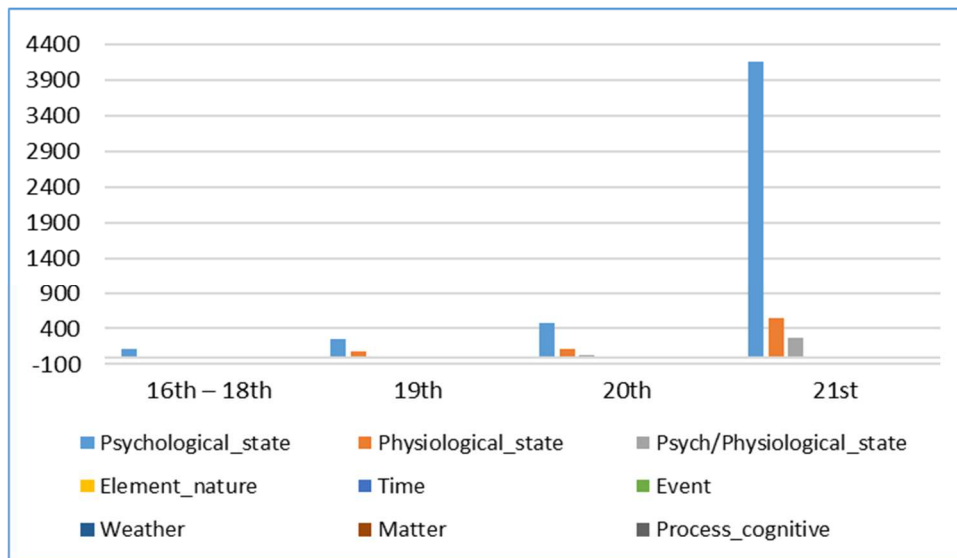

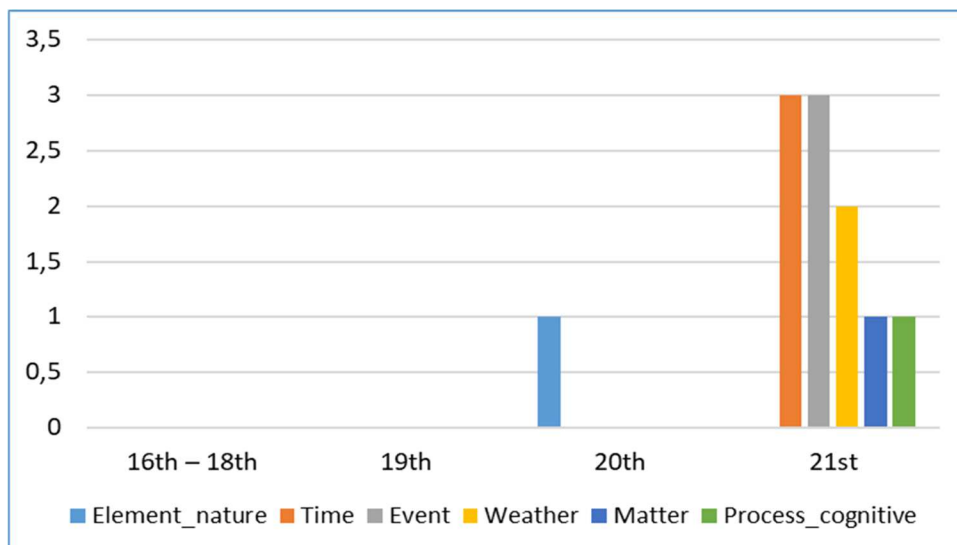Figure 8.20 Semantic fields of nouns in the MIHI EST construction (16th–21st centuries)

Figure 8.21 Zoom in the less frequent semantic classes (16th–21st centuries)

However, a closer look at these classes gives compelling insights into this predisposition toward innovation. Note that, throughout all periods, the semantic class of psychological states remains the predominant one and increases with each historical period. Moreover, the new types entering the MIHI EST construction are all used metaphorically and refer to emotional states. This shows that the construction has not extended its meaning, but only its lexical inventory. Hence, it is not to be positioned higher on the productivity cline, and has not achieved a higher degree of schematicity neither, contrary to what was expected. Nevertheless, the expansion withnessed in the present-day language is a clear sign of a tendency toward productivity, since the construction is attracting new lexical items.

8.2.4   Preliminary conclusions

In this second study, a number of productivity tests were applied to the Romanian MIHI EST construction, on the one hand, in order to observe possible changes in the degree of productivity of this construction throughout the centuries, and, on the other hand, to assess whether the use of this construction is expanding or retracting in present-day language.

The diachronic study has revealed that the degree of productivity of the MIHI EST construction has been very dynamic throughout the centuries, showing fluctuating tendencies for the two aspects of productivity considered here: realized and potential productivity. Put differently, throughout the pre-21st century period, innovation of the lexical inventory and entrenchment succeed each other, resulting in a dynamic, slowly increasing productivity degree of the MIHI EST construction. This tendency is validated for the 21st century as well, by the expansion of the lexical inventory revealed for this last century through the qualitative analysis that has been carried out. These outcomes are pertinent to the first research question, that is, how has the degree of productivity of this construction evolved since the first Romanian texts.

As for the second research question, whether the MIHI EST construction is expanding or retracting in productivity in present-day Romanian, the answer has to be very nuanced. After three periods of stability, when only nouns expressing physiological or psychological states were admitted in the MIHI EST construction, the situation changes in present-day Romanian, since the construction allows original combinations with nouns coming from different semantic fields such as events, acts, time, meteorological phenomena, and elements of nature. This innovation confirms the tendency toward productivity of the construction, which, due to its high semantic coherence, can coerce nouns from other semantic fields into the initial physiological and psychological interpretation. However, the expectation that the degree of schematicity will also increase as a consequence of this innovation, is not borne out, since the metaphorical use

of these nouns allows the construction to continue to express physiological or psychological states. Nevertheless, this remains a sign of expansion. Whether the expansion of this specific construction can be seen as a significant indication for a propensity in Romanian toward non-canonical marking of core arguments, which goes against the tendency in the SAE languages toward canonical marking (cf. Haspelmath 2001a, Seržant 2013), is difficult to say. Further research including other types of predicates, such as adjectives, adverbs or verbs that occur with non-canonical arguments is required in order to build enough foundation for this claim.

Chapter 9   Summary and conclusions

The aim of this thesis has been to investigate potential oblique-subject constructions with nominal predicates in Romanian, paying special attention to the MIHI EST construction. The ultimate goal of this research, manifested in the two research questions put forward in Chapter 1, was twofold. On the one hand, my aim has been to assess whether the dative subject-like experiencer in the MIHI EST construction displays subject properties, and, on the other hand, to investigate whether potential non-canonical subject marking in Romanian is expanding or regressing. This is particularly relevant in the context of the claimed tendency of central Standard Average European languages toward canonical marking of core arguments (cf. Haspelmath 2001a, Seržant 2013).

9.1   Summary

The first chapter presents the object of this study, i.e. the MIHI EST structure, and defines the main objectives of the present research. It also explains the motivation of this research and its place within the FWO project by which it was financed. The present study was inspired by the classification proposed by Haspelmath (1998, 2001a, and 2001b) – based on Bossong (1998) – in which Romanian is allegedly isolated from central Standard Average European (SAE) languages, and is, instead, classified together with peripheral language families such as East Slavic and Baltic, due to their marking of core arguments. The two research questions, also reiterated above, are followed by a brief description of the structure of the thesis.

Chapter two aims to properly position Romanian among the SAE languages with respect to non-canonical case marking of subjects. It first reflects on the definition of subject, giving an overview of the difficulties encountered by linguists when trying to grasp the nature of this concept. It turns out that the definitions proposed in the scholarship highly depends on theoretical frameworks. One of the earliest definitions,

provided by Keenan (1976: 312), describes the subject as an NP that shows "a clear preponderance of the subject properties" from his list. Keenan (1976) distinguishes between coding, behavioral, semantic, and pragmatic properties of subjects. Among these, only the coding and the behavioral properties are considered in the literature on subjects, since these only provide clear criteria for distinguishing between subjects and objects. As opposed to Keenan's approach, more formal theoretical frameworks define the subject as the DP immediately below the sentence node (cf. Lexical Functional Grammar), or the DP in the specifier position of IP (cf. Minimalism) (see Farrell 2005). Less formal approaches define the subject as a language-specific relation, as does Dryer (1996, 1997), who views using the same labels (like *subject* or *object*) for grammatical relations in different languages as problematic, and considers it a terminological issue.

Among the less formal views, the prototype approach proposed by Seržant (2013), envisages a scalar notion of the subject. Referring to Keenan's (1976) approach, Seržant (2013) defines the canonical subject, which he labels *prototypical subject*, as the argument with the maximal set of subject properties found in a specific language. This definition leaves room for non-canonical subjects, which are defined as arguments that are not encoded in the nominative and do not trigger verb agreement, but do show other subject properties. Non-canonical subjects are seen as less prototypical subjects, since they do not show the maximal set of subject properties in a language. This prototype approach describes less-prototypical subjects as being in continuous evolution varying diachronically from *subject-like obliques* to *non-canonical subjects*. A *subject-like oblique*, according to Seržant, displays very few behavioral subject properties. By acquiring most of the behavioral properties, it evolves into a *non-canonical subject*, and eventually into a *prototypical subject*, when it has the maximum set of behavioral as well as coding subject properties (Seržant 2013). Yet, other scholars, adopting a properties-independent approach of the subject, view the subject as a non-scalar notion, defined by its leftmost position in the argument structure of a specific event (Eythórsson & Barðdal 2005, Barðdal 2006a, and Barðdal & Eythórsson 2018).

Other related issues are addressed in this chapter, such as the universality of the concept of subject. The great variety of constructions containing a subject in each language has led scholars such as Eythórsson & Barðdal (2005: 827, among others) to suggest that a universal concept of subject cannot be maintained, and that all definitions of the subject are *language-specific* or even *construction-specific*, as well as the properties that characterize it.

The aim of the third chapter is to identify the properties of subjects in Romanian. The first part describes the coding properties of the canonical subject, i.e. nominative case and verb agreement, while the second part focuses on the behavioral properties of subjects and examines whether accusative and dative experiencers behave like nominative subjects. In this second part, three types of predicates are taken into consideration: (i) psychological verbs of the *preoccupare* and *piacere* classes, as defined by

Belletti & Rizzi (1988), (ii) pain verbs with an accusative, and (iii) *complex predicate constructions*, which combine an experiencer with a light verb and a state, thus acquiring an experiential meaning. The following behavioral properties are examined: word order, reflexive binding, control, raising-to-subject, conjunction reduction, deletion in imperatives, deletion in telegraphic style, and quantification.

Several of the applied tests are shown to be relevant and conclusive in Romanian. This is the case for word order, raising-to-subject, control, deletion in telegraphic style, bare quantifiers in clause-initial position, and secondary predication. Other tests, such as deletion of the subject in imperatives, conjunction reduction, and binding, although applicable, are not conclusive in Romanian, due to certain language-specific constraints such as pro-drop, obligatoriness of dative clitics in specific experiencer contexts, less strict binding rules and the limited use of the infinitive, replaced by the Balkan subjunctive. It has to be mentioned that these specific criteria do not provide any evidence against the subject status of the dative or accusative experiencers under scrutiny. Only one of the syntactic properties under scrutiny totally fails to distinguish between subjects and objects in Romanian (raising-to-object). Considered from a theory-independent perspective, the non-canonical subject clearly patterns in Romanian with canonical subjects.

In chapter four, I situate the Romanian MIHI EST pattern among the complex predicate constructions. Two main types of complex predicate constructions are of interest: *the determined state noun type* and *the bare state noun type.* The first type is represented by the *CAPIO inchoative* construction (cf. *Mă apucă foamea* lit. me.ACC seizes hunger.the 'I start feeling hungry') and by its aspectual opposite, the *cessative* construction *Mi-a trecut somnul* lit. me.DAT has passed sleep.the 'I don't feel sleepy anymore'). As for the second type of construction with a bare state noun, it is represented by the MIHI EST construction (cf. *Mi-e dor de casă* lit. me.DAT is longing of home 'I miss home') and by its inchoative variant, the *VENIO inchoative* construction (cf. *Îmi vine somn* lit. me.DAT comes sleep 'I begin feeling sleepy').

Aside from the MIHI EST construction, I also discuss the so-called *verbal* MIHI EST construction (Bauer 2000: 180), in which MIHI EST combines with a finite or a non-finite clause. The construction then carries a modal meaning. Two types of modal structures are distinguished: the pattern with a deontic modal *fi* 'be', which can express an obligation, a physical or moral constraint, or predestination, and the pattern with a dynamic modal *fi* 'be', expressing willingness, ability, an unreal situation or a dream. These patterns were initially analyzed as raising structures (Avram 1999, Cornilescu 2009, and Ionescu 2013, cited in Barbu 2015). However, this analysis has been revised by Barbu (2015), who suggests a raising analysis only for the deontic modal *fi* 'be' and a control analysis for the dynamic modal *fi* 'be'. On the basis of my data, I challenge this recent analysis, and I suggest that, when they occur with a dative, an approach in terms of control is more appropriate for both the deontic and the dynamic modal *fi* 'be'.

After exploring the verbal MIHI EST construction, I examined the complex predicate constructions under scrutiny and I traced them back to their origin. Some of them clearly go back to Indo-European (e.g. MIHI EST and *VENIO inchoative* constructions), whereas for others the origin remains unclear due to scarcity of data (*CAPIO inchoative* and the *cessative* constructions). A larger part of this chapter is dedicated to the origin of the MIHI EST pattern. The traditional view on the origin and the evolution of this pattern – which puts forward the predicative possession construction as its ancestor –, has been challenged by Barðdal et al. (2012), Barðdal & Smitherman (2013), Danesi, Johnson & Barðdal (2017), and Danesi & Barðdal (2018), according to whom the experiencer MIHI EST pattern traces back to the (Proto-)Indo-European dative experiencer construction.

Chapter five provides information about the corpus and the methodology used for this study. After explaining the choice for a web corpus for the present-day language, I describe the difficulties I encountered with finding an already existing corpus for the present-day language, and in collecting my own corpus for the pre-21st century period of Romanian. With the help of certain Romanian scholars, who provided me with a part of the documents in digital format, and by consulting open-access online libraries, I have been able to collect a significant number of texts for the pre-21st century period of Romanian, in digital format. I then processed these documents and uploaded them onto the Sketch Engine platform, which provided me with the necessary tools to make the newly created corpus searchable for research purposes. As for the organization of the obtained corpus in different periods, I relied on the periodization proposed by Gheție (1997), which I adapted slightly to the needs of my research.

This chapter also describes the methodology used in extracting relevant examples from the gathered corpora, more specifically the phases of the compilation of my dataset. During the first phase, I identified which nouns occur in the MIHI EST construction. In the second phase, I collected examples instantiating different experiential constructions, in which the nouns obtained in the first phase may enter, and I created the relevant dataset. After eliminating the noise, this dataset counts 8 458 examples to be analyzed, among which 4 828 examples (57 %) are with a dative experiencer, 3 412 (43 %) are with other kinds of experiencers (i.e. nominative, or accusative), and 2 784 examples are of the MIHI EST type only (33 %). The relevant examples from present-day Romanian are more numerous than the ones from the pre-21st century Romanian. More precisely, 74 % (6 248) of the examples represent present-day Romanian, and 26 % (2 209) the pre-21st century Romanian. Once manually annotated according to a selected number of variables, the dataset has been used for the computational analysis of the data and for the visualization of the results.

The methodology I applied in order to gather the list of nouns occurring in the MIHI EST construction is further developed in Chapter 6, where an inventory of the nouns found in the MIHI EST construction throughout the centuries is established. Based on the gathered

data, this inventory contains mostly state nouns and counts 29 different nouns for the pre-21st century period of Romanian, and 95 nouns for the present-day language.

Before having a closer look at the selected nouns, a few ambiguous cases are discussed such as adjectives, adverbs and interjections. These cases have been unintentionally extracted through my query in the corpus because they can sometimes be analyzed as nouns in Romanian. Due to their ambiguous status, I eliminated them from the final dataset. However, they represent extremely interesting cases, and are valuable material for further research.

The set of nouns turns out to be very dynamic throughout the centuries. New nouns enter into this construction with every historical period, while others disappear from it, although they are used in other types of constructions. Among the nouns that may fill the state noun slot of the MIHI EST construction there are new borrowings or newly formed words through morphological derivation, as well as nouns that have been existing in the language for some time. Although they come from different semantic fields, these nouns are always interpreted as physiological or psychological states when occurring in the MIHI EST construction, and hence are coerced into the meaning of the construction.

During the process of creation and annotation of the corpus, it has become clear that the nouns in my list are found also in the HABEO construction, which encodes the experiencer in the nominative but conveys the same meaning as the MIHI EST construction. The results of this study reveal that during the earliest documented period of Romanian (16th century), the HABEO construction is the first and the dominant construction for most of the state nouns. Starting with this period, the MIHI EST construction tends to attract mostly state nouns which already occur in the HABEO construction. Based on the evidence presented, three possible evolutions of the recruited nouns were observed with respect to their use in the two construction. The newly recruited noun may increase in frequency in the MIHI EST construction while it disappears from the HABEO construction; it may increase in frequency in both constructions, or it may increase in frequency in the HABEO construction while decreasing in frequency in the MIHI EST construction. In my dataset, most nouns follow the first path, increasing in frequency in the MIHI EST construction, while decreasing in the HABEO construction, and only a very limited number of nouns go either of the other two paths.

Chapter seven focuses on the analysis of the MIHI EST construction in Romanian, and more particularly on the behavior of the dative experiencer and of the state noun. The latter is traditionally seen as the subject of the structure because it has two coding subject properties, but it does not behave as a subject from a syntactic point of view. In contrast, the dative experiencer, which is not encoded in the nominative case and does not trigger verb agreement, shows several properties of a syntactic subject.

Indeed, the application of several subject criteria reveals that the dative experiencer occurring in the MIHI EST construction behaves like a subject with respect to word order, control, subject-to-subject raising, deletion of subjects in telegraphic style, bare

quantifiers in clause-initial position, and the ability to take secondary predicates. These criteria clearly support the hypothesis that the dative experiencer is a syntactic subject in the MIHI EST construction. However, other subject tests such as binding, deletion under coordination and in imperatives prove to be less reliable in Romanian. Following Barðdal & Eythórsson (2003, 2018), I have emphasized that the negative results of Romanian with respect to certain subject tests are due to a general property of clitics in Romanian, which are obligatory in experiencer constructions with specific reading. Therefore, tests that involve the omission of dative experiencers are non-conclusive but do not disconfirm the subject status of the dative experiencer. Subsequently, the state noun of the MIHI EST construction was shown to behave like a predicate with respect to modification, complementation, and word order, among others.

From a diachronic point of view, the data show that the state noun behaves like a predicate and the dative argument like a syntactic subject already in the earliest attested texts, tendency that becomes increasingly stronger throughout the centuries. Moreover, the dative experiencer tends more and more to occur clause-initially, whereas the state noun occurs postverbally, co-occurring with verbal complements and adverbial elements, a behavior that is specific for predicates.

In chapter eight, I aim to answer the question whether the MIHI EST construction is expanding or regressing in Romanian. This issue is investigated by means of two different studies, based on two different types of data: a survey among Romanian native speakers, presented in the first section, and a corpus study, described in the second section of the chapter. Through this survey, I intend to verify whether native speakers of Romanian accept in the MIHI EST construction a selection of nouns synonymous with nouns that instantiate the MIHI EST construction but which I have not found in this construction in my corpus. The acceptability of these sentences was judged on a 7-point Likert acceptability scale, with the acceptability level established at 4. The probability that all the test sentences containing the MIHI EST construction be rated unanimously with 1 – given that these sentences were artificially created by me through introspection and that the selected nouns were not attested in the MIHI EST construction – is contradicted by the data. This shows that there is a certain tendency toward innovation of the MIHI EST construction, since native speakers are open to original combinations within this construction. Moreover, the distribution of the scores for the nouns entering in the MIHI EST construction shows that every single test noun has been rated by at least one participant with 7, the maximum score. Likewise, the mean acceptability ratings for these nouns are situated between 1,5 for the least acceptable noun, *descurajare* 'despair, demoralization', and 3,0 for the most acceptable noun, *amărăciune* 'grief, sorrow'.

In order to understand what could influence the decision of speakers in their rating, a bimodal analysis of linguistic (Etymology, Derivation of the noun, Polarity of the event, and Stimulus type) and extralinguistic factors (Gender, Age, Region, Education level and Specialization) has been performed. With respect to the linguistic variables, this analysis

shows that the acceptability ratings tend to be higher with non-derived nouns and in structures with a positive polarity and with a PP stimulus, whereas etymology seems to have less impact on these average ratings. As for the extralinguistic factors, the investigation reveals that higher acceptability ratings correlate with participants younger than 45 years and with respondents originating from Western-Romania and from the Republic of Moldova. The participants who specialized in Romanian language (mostly university students), regardless of the level of their education, and the participants aged more than 45 years tend to be stricter in their ratings.

In the second study, selected productivity tests – estimating realized and potential productivity, together with global productivity – were applied to a dataset containing the MIHI EST construction. The quantitative analysis, which mostly dealt with the pre-21st-century period, has been complemented with a qualitative analysis of the 21st century. The aim of this investigation was, on the one hand, to observe possible changes in the degree of productivity of this construction throughout the centuries, and, on the other hand, to assess whether the usage of this construction is expanding or retracting in present-day Romanian. The diachronic study has revealed that, throughout the pre-21st-century period, innovation and entrenchment are intertwined, having as an outcome a dynamic, slowly increasing productivity degree of the MIHI EST construction. This tendency seems to continue in the 21st century as well, with another innovation wave revealed for this last century by the qualitative analysis. After three periods of stability, when only nouns expressing physiological or psychological states were admitted in the MIHI EST construction, the situation changes in present-day Romanian, as the construction allows original combinations with nouns coming from different semantic fields such as events, acts, time, meteorological phenomena, and elements of nature. In spite of that, the degree of schematicity of the MIHI EST construction does not seem to increase, since these nouns are used metaphorically and coerced into the interpretation of the construction, which continues to express a physiological or a psychological state. Nevertheless, this remains a very clear sign of expansion.

The results of the two studies corroborate each other in that, even in speaker's usage, there is a propensity toward allowing new nouns in the MIHI EST construction. Based on the average ratings, the tendency toward innovation of this construction in the present-day language seems to be conditioned by several factors. Nevertheless, younger speakers, mostly university students having, preferably, other specializations than Romanian language and originating from any region of Romania, except Central-Romania, clearly show a tendency toward innovation of the MIHI EST construction. In what follows, I highlight the most important findings of the present research, providing at the same time, an answer to the two central research questions.

9.2   Main findings

As mentioned *supra*, in this chapter, two key research questions have guided me in designing the present study:

(i)   To which extent is the dative experiencer in the MIHI EST construction a genuine syntactic subject?

(ii)   Is the MIHI EST construction expanding or regressing in Romanian?

The first question, i.e. whether the dative experiencer is a genuine syntactic subject in the MIHI EST construction, represents the core of this research. In order to generate an answer to this research question, I first had to address a series of other questions. Among these questions, the most important ones are: what characterizes a canonical subject in Romanian, or which subject tests are relevant in Romanian in distinguishing between subjects and objects, and, especially, which tests are applicable to complex predicate constructions, to which the MIHI EST construction belongs. Providing an answer to this plethora of research questions is a very delicate matter, since the most frequently used subject tests are highly language- and construction-specific.

In spite of the non-conclusiveness of certain tests, the subject status of the dative experiencer proves to be undeniable, given its behavior in the MIHI EST construction with respect to a series of subject criteria: word order, control, subject-to-subject raising, deletion of subjects in telegraphic style, bare quantifiers in clause-initial position, and the ability to take secondary predicates. Indeed, with respect to these tests, the dative experiencer behaves differently from dative goals and very much like canonical, nominative subjects. Moreover, the state noun, which is analyzed as the subject of the construction in traditional grammars, behaves like a predicate in terms of modification, complementation, and word order, among others. Therefore, I claim in this dissertation, that the dative experiencer is a genuine subject in the MIHI EST construction in Romanian.

The second research question, i.e. whether the MIHI EST construction is expanding or regressing in Romanian, is crucial for a better understanding of the tendency in Romanian with respect to the marking of core arguments. In order to investigate the evolution of the MIHI EST construction, I first observe which nouns can enter in this construction, then I turn my attention toward the competition between the MIHI EST and the HABEO constructions. The investigation of the nouns entering in the MIHI EST construction involves a diachronic study of a set of nouns occurring in the MIHI EST construction. My data show that the set of nouns clearly widens throughout the pre-21st century period in both absolute and relative figures, from 14 nouns in the earliest period, to 25 nouns in the 20th century, as absolute figures. As for the 21st century, due to the large size of the dataset, the high absolute type frequency (95 nouns) turns into a rather low relative type frequency for this period. It is important to mention that, in every

century, new nouns have been recruited (identified as unique occurrences, with the highest percentage in the 21st century: 50 % of the type nouns), whereas other nouns have been eliminated from the construction. As for their semantics, in the pre-21st century period, the construction recruited almost exclusively state nouns expressing physiological or psychological states. However, in the 21st century several nouns from different semantic fields are accepted, but are coerced into the meaning of the MIHI EST construction, that of a physiological or a psychological state.

Research into the competition between the MIHI EST construction and the HABEO construction is meant to provide further insight into the evolution of the MIHI EST construction, this time seen from the point of view of its interaction with the HABEO construction. As shown in chapter six of this dissertation, the data reveal that the MIHI EST construction tends to recruit mostly nouns that first occurred in the HABEO construction, and that this was more frequent between the 16th and 18th centuries. The MIHI EST construction then becomes, in most cases, their dominant construction in expressing psychological and physiological states.

Finally, the evolution of the MIHI EST construction in Romanian has been addressed also quantitatively, in chapter eight of this thesis, in which I carried out two separate studies, based on two types of data: a questionnaire survey and a corpus study. The questionnaire survey exposed the attitude of native speakers of Romanian toward innovation of the MIHI EST construction by recruitment of nouns synonymous with the ones already occurring in the construction. Based on the corpus study, I measured different aspects of the productivity of the construction under scrutiny, by means of a quantitative analysis, which was complemented by qualitative observations.

These data reveal that, during the course of the history, periods in which the lexical inventory of the construction was renewed alternate with periods of entrenchment, resulting in a slow increase in productivity of the MIHI EST construction. The new nouns entering in the MIHI EST construction in the innovation phases took their time in becoming fixed, or deep-rooted in the speaker's mind, during the entrenchment phases. Present-day Romanian is characterized by an unprecedented renewal of the lexical inventory of the construction, since more than 50 % of the types occurring in the MIHI EST construction in this period are unique newly recruited nouns (hapax legomena), some of them coming from different semantic fields. The MIHI EST construction coerces these new nouns into its interpretation, that of a physiological or a psychological state. This shows that the construction does not become more schematic, in that it does not develop new constructions carrying a distinct meaning. Nevertheless, based on the presented facts, the MIHI EST construction shows a clear tendency toward expansion.

To conclude, the results of this study are of great importance to the field of linguistics in general, and to the study of Romanian in the Romance, as well as in the Balkan context, in particular. On the one hand, the dative experiencer occurring in the MIHI EST construction shows several syntactic subject properties. This represents evidence in

favor of an analysis of the dative experiencer as a genuine syntactic subject, comparable to the ones in Icelandic. On the other hand, the study shows that non-canonical subject marking is expanding in Romanian, based on the evolution and the tendency toward expansion, documented for the MIHI EST construction.

Nevertheless, by the nature of its topic and by the special position of Romanian as a bridge between the Romance and the Balkan world, the present study turns out to raise more research questions than it aimed to answer. Indeed, the study of non-canonically marked subjects in Romanian opens a path for numerous other research projects, which go far beyond canonical vs. non-canonical marking of the subject.

9.3   Limitations of this research and directions for future work

A first limitation of this study is that it has focused on experiencer constructions with nominal predicates, namely the MIHI EST construction, which is part of the larger topic of non-canonical marking of subjects in Romanian. In order to corroborate the findings of the present research, a larger study covering all types of predicates occurring with dative and accusative experiencers is necessary. Such a study could offer a different outlook on the matter and would allow for stronger conclusions. Some of these aspects are already being studied in work in progress, such as non-canonical marking in experiencer constructions with verbal predicates occurring with dative and accusative experiencers in Romanian (cf. Van Peteghem & Ilioaia In prep.)

Remaining in the same sphere of verbal predicates, I would also like to mention the special case of the verb *plăcea* 'please', which enters the dative experiencer structures of the type Dat-Nom and Nom-Dat, in Romanian. These two configurations of the verb *plăcea* 'please', Dat-Nom and Nom-Dat, in which the arguments exchange positions only, but keep their case and their semantic roles, both show neutral word order.

Moreover, *plăcea* 'please' seems to develop a new structure in which the core arguments are canonically marked in a transitive structure of the type Nom-Acc, in which the nominative subject expresses the experiencer. I am currently studying this verb in order to understand what drives the creation of this new transitive structure with a nominative experiencer, which is clearly an innovation. In my data, this construction is attested much later than the two dative experiencer structures and it remains very rare (35 occurrences found in my dataset for the 21st century). See further Ilioaia (In prep.) on the co-existence of the two opposite configurations of the verb *plăcea* 'please', Dat-Nom and Nom-Dat, with the newly developed transitive structure, Nom-Acc. There, I raise the question whether this innovation may be seen as a first sign toward a canonicalization of the marking of subjects in Romanian. I also point out that the emergence of a transitive

pattern with *plăcea* 'please' may be seen as a confirmation that the experiencer encoded in the dative case in the Dat-Nom structure is actually the syntactic subject of the structure. Nevertheless, the occurrence of this new (transitive) pattern, which, so far, remains an isolated phenomenon involving only the verb *plăcea* 'please', could disconfirm the hypothesis that non-canonical subject marking is progressing in Romanian.

In future work, it would be interesting to carry out comparative studies with other Romance languages, like Spanish, French or Italian, with which Romanian shares a number of properties, but from which Romanian also deviates due to its isolation in the Balkan area. A first look at the differences between Romanian and Spanish, for instance, shows that Romanian has oblique subjects in constructions where Spanish has canonical subject marking, as is the case in the MIHI EST constructions (e.g. *mi-e foame* me.DAT is hunger 'I'm hungry' vs. Sp. *tengo hambre* 'I have hunger', cf. Van Peteghem 2017). Furthermore, Romanian has accusative subject-like obliques in constructions where Spanish has dative ones (cf. Rom. *mă doare capul* me.ACC aches head-the, vs. Spanish *me duele la cabeza* me.DAT aches the head 'I have a headache', see Van Peteghem 2017), just like Latin or other Romance daughter languages.

A comparative study between Romanian and other Romance languages with regard to the CAPIO *inchoative* construction would also be of great interest. While in Romanian, the CAPIO *inchoative* construction is clearly of the *preoccupare*-type, with an accusative subject-like experiencer, in Italian this pattern develops a structure more like the *piacere*-type, with a dative object-like experiencer. As opposed to its sister languages, French has a structure that can be considered intermediate between the two types of constructions. In this case, the verb *prendre* 'take' freely alternates with the verb *venir* 'come', whereas the clitic form encoding the experiencer is ambiguous between the accusative and the dative, due to syncretism (cf. *Il me vient / prend l'envie de ...* lit. it me.DAT=ACC come/ take the mood to ... 'I begin feeling like ....').

An in-depth comparison of Romanian experiencer constructions with the Latin ones would allow reconstructing the development that took place between Latin and 16th-century Romanian, with respect to these constructions. Two main hypotheses can be advanced: (i) the existence of non-canonical experiencer constructions in Romanian could be a result of Latin conservatism, as has been argued for Spanish by Melis & Flores (2013), (ii) the larger proliferation of these constructions in Romanian, compared to other Romance languages, may be due to substratum influence, as speculated by Bossong (1998: 268). The first scenario is very appealing since Latin also has both accusative and dative subjects with various verb classes. A first thing to investigate would be to which extent the verb classes taking an oblique subject in Romanian are the same as in Latin and whether the case choice remains the same for predicates and verb classes.

For instance, in Romanian the verb *durea* 'ache' assigns the accusative to its experiencer argument (cf. *mă doare capul* me.ACC aches head-the 'I have a headache'), whereas its Latin etymon *dolere* assigns the dative (Lat. *mihi caput dolet* me.DAT head aches 'I have a headache'), as in Spanish (*me duele la cabeza* me.dat aches the head 'I have a headache'). This intriguing difference between Romanian, on the one hand, and Latin and Spanish, on the other hand, could be explained by the status of the dative experiencer in this construction, i.e. an external possessor in Latin and Spanish, and a verb argument in Romanian.

Furthermore, the complex predicate construction *mi-e foame* (me.DAT is hunger 'I'm hungry'), which is specific to Romanian in the Romance context, can be traced back to the MIHI EST experiencer construction in Latin (cf. *mihi est fames*, see Fedriani 2011, discussed also in Barðdal et al. 2012, Barðdal & Smitherman 2013, Danesi, Johnson & Barðdal 2017, and Danesi & Barðdal 2018). Nevertheless, the second explanation, that of a substratum influence, should a priori not be excluded, given a series of common features between Romanian and the Balkan languages.

With respect to the relation between Romanian and the surrounding languages, another fascinating direction for further research would be to compare Romanian with Balkan languages, such as Bulgarian, Albanian and Greek. Considered as part of the Balkan linguistic area, Romanian shares with these languages several typological features acquired by contact-induced influences. Some of these features are of great interest for the study of non-canonical subjects, namely the replacement of infinitives with subjunctives, and the non-finiteness of the embedded subjunctive in control and in raising configurations. This use of the subjunctive in Balkan languages – but also in certain Romance languages, such as Romanian and a few Southern Italian dialects – alongside the existence of inflected and personal infinitives, signaled in other Romance languages, has led Groothuis (2019) to claim that finiteness may be a scalar rather than a binary notion. I show in this dissertation that Romanian data provide clear instances of control subjunctives as well as cases of raising out of subjunctive clauses, which undoubtedly represents additional evidence in favor of this hypothesis.

In a nutshell, this dissertation has shown that the dative experiencer occurring in the MIHI EST construction in Romanian has several syntactic subject properties and that it patterns with canonical subjects rather than with objects or dative goals. Furthermore, the expansion of non-canonical marking of subjects in Romanian has been partially confirmed, based on the evolution of the MIHI EST construction. However, it is difficult to say whether the expansion of this specific construction and its innovative inclinations are representative enough in order to signal an increasing propensity in Romanian toward non-canonical marking of core arguments. If this were the case, the results of this study would contradict the hypothesis that the SAE languages tend to evolve toward canonical marking of core arguments (cf. Haspelmath 2001, Seržant 2013).

Further research including other types of predicates, such as adjectives, adverbs or verbs, as well as other types of experiencer constructions that exist in Romanian with non-canonically case-marked arguments is required in order to build enough foundation for such a claim.

Bibliography

Achard, Michel. 2015. *Impersonals and other agent defocusing constructions in French.* John Benjamins Publishing Company.

Alboiu, Gabriela. 2002. *The Features of Movement in Romanian.* Bucharest: The University of Bucharest Press.

Alboiu, Gabriela. 2005. *When CP domains are pro-Case.* Manuscript. University of Western Ontario.

Alboiu, Gabriela. 2007. Moving forward with Romanian Backward Control and Raising. In William D. Davies & Stanley Dubinsky (eds.), *New Horizons in the Analysis of Control and Raising.* Vol. 71. Dordrecht: Springer Netherlands.187–211.

Alexiadou, Artemis & Elena Anagnostopoulou. 2002. Raising without infinitives and the role of agreement. In Artemis Alexiadou, Elena Anagnostopoulou, Sjef Barbiers & Hans-Martin Gärtner (eds.), *Dimensions of Movement: From features to remnants.* John Benjamins Publishing. 17–30.

Alexiadou, Artemis & Gianina Iordăchioaia. 2014. The psych causative alternation. *Lingua.* Elsevier 148. 53–79.

Allen, Cynthia L. 1986. Reconsidering the history of "like". *Journal of Linguistics.* JSTOR 22(2). 375–409.

Allen, James. 1995. *Natural language understanding.* Lebanon, Indiana, U.S.A.: Benjamin/Cummings Pub. Co.

Anagnostopoulou, Elena. 1999. On experiencers. In *Studies in Greek syntax*, 67–93. Springer.

Anderson, John M. 1976. *The grammar of case: Towards a localistic theory.* Vol. 4. Cambridge University Press Archive.

Andrews, Avery D. 1976. The VP complement analysis in Modern Icelandic. In *NELS VI (Montreal Working Papers in Linguistics, 6), 1-22.*

Andrews, Avery D. 2001. Non-canonical A/S marking in Icelandic. *Typological studies in language* 46. 85–112.

Andrews, Avery D. 2007. The major functions of the NP. *Language typology and syntactic description.* 2nd Edition. Vol. 1 Clause structure. 132–223.

Anshen, Frank & Mark Aronoff. 1999. Using dictionaries to study the mental lexicon. *Brain and language.* Elsevier 68(1–2). 16–26.

Aronoff, Mark. 1976. Word formation in generative grammar. *Linguistic Inquiry Monographs* Cambridge, Mass. (1). 134 p.

Augustinova, Tania. 2006. A functional typology of copular "be": Towards an HPSG formalisation. In Stefan Müller (ed.), *Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar*, 27–38.

Auwera, Johan van der & Dirk Noël. 2011. Raising: Dutch between English and German. *Journal of Germanic Linguistics.* Cambridge University Press 23(1). 1–36.

Avram, Larisa. 1999. *Auxiliaries and the Structure of Language.* Vol. 4. Editura Universităţii din Bucureşti.

Baayen, Harald. 1993. On frequency, transparency and productivity. In *Yearbook of morphology 1992.* Springer. 181–208.

Baayen, Harald & Rochelle Lieber. 1991. Productivity and English derivation: a corpus-based study. *Linguistics* 29(5). 801–844.

Baayen, R. Harald. 2001. *Word frequency distributions.* Series: Text, Speech, and Language Technology. Vol. 18. Kluwer Academic.

Baayen, R. Harald. 2009. Corpus linguistics in morphology: morphological productivity. *Corpus linguistics: An international handbook.* 43. 900–919.

Bakker, Dik & Anna Siewierska. 2007. Another take on the notion Subject. *Structural-functional studies in English Grammar.* John Benjamins Amsterdam & Philadelphia 141–158.

Baldi, Philip & Andrea Nuti. 2010. Possession in Latin. In Philip Baldi & Pierluigi Cuzzolin (eds.), *New Perspectives on Historical Latin Syntax.* Vol. 3. Berlin: Mouton de Gruyter 239–387.

Barbu, Ana-Maria. 2015. *A fi* 'be' as a modal verb in Romanian. In Michail L. Kotin & Richard J. Whitt (eds.), *To be or not to be? The Verbum Substantivum from Synchronic, Diachronic and Typological Perspectives.* Cambridge Scholars Publishing. 316–334.

Barbu, Ana-Maria. 2017. Construcțiile modale ale verbului *a fi. Diacronia.*

Barðdal, Jóhanna. 1998. Argument structure, syntactic structure and morphological case of the impersonal construction in the history of Scandinavian. *Scripta Islandica* 49. 21–33.

Barðdal, Jóhanna. 2000. Oblique subjects in Old Scandinavian. *NOWELE. North-Western European Language Evolution* 37(1). 25–51.

Barðdal, Jóhanna. 2001. The perplexity of Dat-Nom verbs in Icelandic. *Nordic Journal of Linguistics.* Cambridge University Press 24(1). 47–70.

Barðdal, Jóhanna. 2002. "Oblique Subjects" in Icelandic and German. *Working papers in Scandinavian syntax.* Lund University 70. 61–99.

Barðdal, Jóhanna. 2004. The semantics of the impersonal construction in Icelandic, German and Faroese: Beyond thematic roles. *Focus on Germanic typology.* 105–137.

Barðdal, Jóhanna. 2006a. Construction-specific properties of syntactic subjects in Icelandic and German. *Cognitive Linguistics* 17(1). 39–106.

Barðdal, Jóhanna. 2006b. Predicting the productivity of argument structure constructions. In *Annual Meeting of the Berkeley Linguistics Society.* Vol. 32, 467–478.

Barðdal, Jóhanna. 2008. *Productivity: Evidence from case and argument structure in Icelandic.* Vol. 8. John Benjamins Publishing.

Barðdal, Jóhanna. 2011. The rise of Dative Substitution in the history of Icelandic: A diachronic construction grammar account. *Lingua* 121(1). 60–79.

Barðdal, Jóhanna, Valgerḥur Bjarnadóttir, Serena Danesi, Tonya Kim Dewey, Thórhallur Eythórsson, Chiara Fedriani & Thomas Smitherman. 2013. The story of 'woe'. *Journal of Indo-European Studies* 41(3). 321–377.

Barðdal, Jóhanna, Eleonora Cattafi, Laura Bruno & Serena Danesi. Forthcoming. Non-Nominative Subjects in Latin and Ancient Greek: Applying the Subject Tests on Early Indo-European Languages.

Barðdal, Jóhanna & Thórhallur Eythórsson. 2003. The change that never happened: the story of oblique subjects. *Journal of Linguistics* 39(3). 439–472.

Barðdal, Jóhanna & Thórhallur Eythórsson. 2006. Control infinitives and case in Germanic 'Performance error' or marginally acceptable constructions? *Case, Valency and Transitivity,* L. Kulikov, A. Malchukov and P de Swart (eds.) 147–77.

Barðdal, Jóhanna & Thórhallur Eythórsson. 2012. 'Hungering and Lusting for Women and Fleshly Delicacies': Reconstructing Grammatical Relations for Proto-Germanic. *Transactions of the Philological Society* 110(3). 363–393.

Barðdal, Jóhanna & Thórhallur Eythórsson. 2018. What is a subject: The nature and validity of subject tests. In Jóhanna Barðdal, Na'ama Pat-El & Stephen Mark Carey (eds.). *Non-Canonically Case-Marked Subjects: The Reykjavík-Eyjafjallajökull papers* 200. 257–274.

Barðdal, Jóhanna, Thórhallur Eythórsson & Tonya Kim Dewey. 2019. The alternating predicate puzzle: Dat-Nom vs. Nom-Dat in Icelandic and German. *Constructions and Frames.* John Benjamins 11(1). 107–170.

Barðdal, Jóhanna & Spike Gildea. 2015. Diachronic Construction Grammar: Epistemological context, basic assumptions and historical implications. *Diachronic construction grammar.* John Benjamins Amsterdam & Philadelphia 1–49.

Barðdal, Jóhanna & Thomas Smitherman. 2013. The quest for cognates: A reconstruction of oblique subject constructions in Proto-Indo-European. *Language Dynamics and Change* 3(1). 28–67.

Barðdal, Jóhanna, Thomas Smitherman, Valgerḥur Bjarnadóttir, Serena Danesi, Gard B. Jenset & Barbara McGillivray. 2012. Reconstructing constructional semantics: The dative subject construction in Old Norse-Icelandic, Latin, Ancient Greek, Old Russian and Old Lithuanian. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"* 36(3). 511–547.

Bauer, Brigitte LM. 1996. Residues of non-nominative Syntax in Latin: the MIHI EST construction. *Historische Sprachforschung/Historical Linguistics* 109(2). 241–256.

Bauer, Brigitte LM. 2000. *Archaic syntax in Indo-European: the spread of transitivity in Latin and French* (Trends in Linguistics 125). Berlin; New York: M. de Gruyter.

Bauer, Laurie. 2001. *Morphological productivity.* Vol. 95. Cambridge University Press.

Belletti, Adriana & Luigi Rizzi. 1988. Psych-verbs and Theta-Theory. *Natural Language & Linguistic Theory* 6(3). 291–352.

Benedetti, Marina. 2013a. Non-canonical subjects in clauses with noun predicates. *Argument Structure in Flux. The Naples-Capri Papers* 15–31.

Benedetti, Marina. 2013b. Experiencers and psychological noun predicates. In Serzant, Ilja A. & Leonid Kulikov (eds), *The Diachronic Typology of Non-Canonical Subjects* 140. John Benjamins Publishing. 121–138.

Benedicto, Elena. 1995. Mne ne citaetsja: (Relativized) modality, datives and reflexive suffixes. *Working Papers in Linguistics* 2(2). 1–16.

Bennett, Charles E. 1914. Notes on Horace. *The Classical Quarterly* 8(3). 145–150.

Bentley, Delia. 2006. *Split Intransitivity in Italian.* Walter de Gruyter.

Benveniste, Émile. 1966. *Problèmes de linguistique générale.* Gallimard. Paris.

Bernódusson, Helgi. 1982. Ópersónulegar setningar í íslensku [Impersonal Sentences in Icelandic]. *Master's thesis, University of Iceland, Reykjavík.*

Beys, Michael. 2009. Subject-to-Object Raising Constructions in Modern Greek. *Selected papers on theoretical and applied linguistics* 18. 105–113.

Bianchi, Valentina. 2003. On finiteness as logophoric anchoring. In J. Guéron & L. Tasmowsky (eds.), *Temps et point de vue/ Tense and point of view.* Paris: Nanterre. 213–246.

Bickel, Balthasar. 2004. The syntax of experiencers in the Himalayas. *Typological Studies in Language* 60. 77–112.

Blunden, Andy. 2005. The First Subject: Aristotle.

Bolkestein, A. Machtelt. 1983. Genitive and dative possessors in Latin. *Advances in functional grammar* 11. 55–91.

Bolkestein, A. Machtelt. 2001. Possessors and experiencers in Classical Latin. *Typological studies in language* 47. 269–284.

Borsley, Robert D. 1991. *Syntactic Theory: A unified approach.* New York: Chapman and Hall, Inc.

Borsley, Robert D. 1996. *Modern phrase structure grammar.* Blackwell.

Bossong, Georg. 1998. Le marquage de l'expérient dans les langues d'Europe. *Empirical Approaches to Language Typology.* Mouton de Gruyter 259–294.

Brâncuş, Grigore. 2007. O concordanţă gramaticală româno-albaneză: modul supin. *Studii de istorie a limbii române.* Bucureşti, Editura Academiei Române. 167–173.

Bresnan, Joan, Ash Asudeh, Ida Toivonen & Stephen Wechsler. 2015. *Lexical-functional syntax.* 2nd Edition. Oxford: Blackwell.

Burzio, Luigi. 1989. On the non-existence of disjoint reference principles. *Rivista di Grammatica Generativa.* Vol. 14. Unipress. 3-27.

Burzio, Luigi. 2010. Constraint violability and the Chain Condition. *Revista Virtual de Estudos da Linguagem.*

Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and cognitive processes* 10(5). 425–455.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language.* JSTOR 711–733.

Bybee, Joan & Sandra Thompson. 1997. Three frequency effects in syntax. In *Annual Meeting of the Berkeley Linguistics Society.* Vol. 23, 378–388.

Caluianu, Daniela. 2013. Four Romanian verbs of occurring: The effect of argument structure on verb meaning and use. *Argument structure in flux,* Ámsterdam, John Benjamins 231–253.

Campos, Héctor. 1999. Transitividad e intransitividad. *Gramática descriptiva de la lengua española.* Madrid, Espasa 2. 1519–1574.

Caragiu-Marioţeanu, Matilda. 1962. Moduri nepersonale. Editura Academiei.

Cardinaletti, Anna. 2004. Toward a Cartography of Subject Positions. In Luigi Rizzi (ed.), *The Structure of CP and IP: The Cartography of Syntactic Structures.* Vol. 2. Oxford University Press.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* M.I.T. Press.

Chomsky, Noam. 1980. On binding. *Linguistic inquiry.* JSTOR 11(1). 1–46.

Cinque, Guglielmo. 2007. A note on linguistic theory and typology. *Linguistic Typology* 11(1). 93–106.

Clausner, Timothy C. 2002. How conceptual metaphors are productive of spatial-graphical expressions. In *Proceedings of the Annual Meeting of the Cognitive Science Society.* Vol. 24.

Clausner, Timothy C. & William Croft. 1997. Productivity and schematicity in metaphors. *Cognitive science.* Wiley Online Library 21(3). 247–282.

Cole, Peter, Wayne Harbert, Gabriella Hermon & Shikaripur N. Sridhar. 1980. The acquisition of subjecthood. *Language.* JSTOR 719–743.

Comorovski, Ileana. 1985. Control and obviation in Romanian. In *Proceedings of the Eastern States Conference on Linguistics.* Ohio: Ohio State University. 47–56.

Comrie, Bernard. 1984. Subject and object control: syntax, semantics, pragmatics. In *Annual Meeting of the Berkeley Linguistics Society.* Vol. 10, 450–464.

Cornilescu, Alexandra. 1997. *The double subject construction in Romanian: Notes on the syntax of the subject.* Editura Academiei Române.

Cornilescu, Alexandra. 2000. The double subject construction in Romanian. In Virginia Motapanyane (ed.), *Comparative studies in Romanian syntax.* Dordrecht: Elsevier. 83–134.

Cornilescu, Alexandra. 2004. Direct object movement and the structure of the Romanian left periphery. *Antwerp Papers in Linguistics (APiL)* 141–166.

Cornilescu, Alexandra. 2009. Restructuring strategies of the Romanian verb *fi* 'be' and the analysis of existential sentences. *Null subjects, expletives and locatives in Romance.* Universität Konstanz. Konstanz. 199–230.

Cotfas, Maura Aurelia. 2016. Romanian implicative verbs revisited: "a încerca" 'try' versus "a reuşi" 'manage.' *Revue roumaine de linguistique* 61(1). Editura Academiei Române. 29–55.

Cristea, Dan, Nils Diewald, Gabriela Haja, Cătălina Mărănduc, Verginica Barbu Mititelu & Mihaela Onofrei. 2019. How to find a shining needle in the haystack: Querying CoRoLa: solutions and perspectives. *Revue roumaine de linguistique* 64(3). Editura Academiei Române. 279–292.

Cristea, Dan, Marius Răschip, Corina Forăscu, Gabriela Haja, Cristina Florescu, Bogdan Aldea & Elena Dănilă. 2007. The Digital Form of the Thesaurus Dictionary of the Romanian Language. In *Proceedings of the 4th International IEEE Conference SpeD,* 195–206.

Cristofaro, Sonia. 2009. Grammatical Categories and Relations: Universality vs. Language-Specificity and Construction-Specificity. *Language and Linguistics Compass.* Wiley Online Library 3(1). 441–479.

Croft, William. 1998. The structure of events and the structure of language. (Ed.) M Tomasello. *The New Psychology of Language: Cognitive and functional Approaches to Language Structure* 67–92.

Croft, William. 2001. *Radical construction grammar: Syntactic theory in typological perspective.* Oxford University Press on Demand.

Croft, William. 2002. *Typology and universals.* Cambridge University Press.

Croft, William. 2012. *Verbs: Aspect and causal structure.* Oxford: Oxford University Press.

Croft, William & D. Alan Cruse. 2004. *Cognitive linguistics.* Cambridge University Press.

Cuervo, María Cristina. 1999. Quirky but not eccentric: dative subjects in Spanish. *MIT Working papers in Linguistics* 34. 213–227.

Cuervo, María Cristina. 2010. Some dative subjects are born, some are made. In *Selected Proceedings of the 12th Hispanic Linguistics Symposium*, 26–37.

Culicover, Peter W. 1992. A note on quantifier binding. *Linguistic Inquiry.* JSTOR 23(4). 659–663.

Culicover, Peter W. 1997. *Principles and parameters: an introduction to syntactic theory* (Oxford Textbooks in Linguistics). Oxford: Oxford University Press.

Danesi, Serena & Jóhanna Barðdal. 2018. Case marking of predicative possession in Vedic. *Non-Canonically Case-Marked Subjects: The Reykjavík-Eyjafjallajökull papers.* John Benjamins Publishing Company. 181–212.

Danesi, Serena, Cynthia A. Johnson & Jóhanna Barðdal. 2017. Between the historical languages and the reconstructed language. *Indogermanische Forschungen.* De Gruyter 122(1). 143–188.

De Haan, Ferdinand. 2006. Typological approaches to modality. *The expression of modality.* Mouton de Gruyter 27. 40–69.

Den Dikken, Marcel. 1998. Predicate inversion in DP. *Possessors, predicates and movement in the determiner phrase.* John Benjamins Amsterdam 177–214.

Diaconescu, Constanţa Rodica & María Luisa Rivero. 2007. An applicative analysis of double object constructions in Romanian. *Probus.* De Gruyter Mouton 19(2). 209–233.

Diaconescu, Ion. 1971. Supinul în limba română din secolele al XVI-lea- al XVIII-lea. *Analele Universităţii din Bucureşti. Limba şi literature română* 1–2.

Dixon, Robert M. W. 1994. *Ergativity.* Cambridge University Press.

Dobrovie-Sorin, Carmen. 1987. A propos de la structure du groupe nominal en roumain. *Rivista di Grammatica Generativa.* Vol. 11. Unipress. 123-152.

Dobrovie-Sorin, Carmen. 1993. *The syntax of Romanian: Comparative studies in Romance.* Walter de Gruyter.

Dobrovie-Sorin, Carmen. 2001. Head-to-head merge in Balkan subjunctives and locality. *Comparative syntax of Balkan languages.* Oxford University Press on Demand 44–73.

Dobrovie-Sorin, Carmen. 2011. *The Syntax of Romanian: Comparative Studies in Romance.* Walter de Gruyter.

Dobrovie-Sorin, Carmen. 2013. Bare nouns. In *A Reference Grammar of Romanian.* John Benjamins. 49–96.

Dobrovie-Sorin, Carmen & Ion Giurgea. 2013. *A reference grammar of Romanian: Volume 1: The noun phrase.* Vol. 207. John Benjamins Publishing.

Donohue, Mark. 2011. Case and configurationality: Scrambling or mapping? *Morphology.* Springer 21(3–4). 499–513.

Dragomirescu, Adina. 2013. Complex predicates. In Gabriela Pană Dindelegan (ed.) *The Grammar of Romanian.* Oxford: Oxford University Press. 191–200.

Drinka, Bridget. 1999. Alignment in Early Proto-Indo-European. In Carol F. Justus and Edgar C. Polomé (eds.), *Language Change and Typological Variation: Papers in Honor of Winfred P. Lehmann on the Occasion of his 82nd Birthday. Vol. II: Grammatical Universals and Typology.* Washington, DC: Institute for the study of Man. 464–500.

Dryer, Matthew S. 1996. Focus, pragmatic presupposition, and activated propositions. *Journal of pragmatics* 26(4). 475–523.

Dryer, Matthew S. 1997. Are grammatical relations universal? *Essays on language function and language type* 115–143.

Dumitrescu, D. & Pascual J. Masullo. 1996. Romanian and the non-nominative subject paramter. *Aspects of Romance Linguistics.*

Elmer, Willy. 2011. *Diachronic grammar: The history of Old and Middle English subjectless constructions.* Vol. 97. Walter de Gruyter.

Elvira, Javier. 2011. Constructions of uncontrolled state or event: The increase in productivity of a new argument structure in Old Spanish. *Constructions and frames.* John Benjamins 3(2). 184–207.

Eythórsson, Thórhallur. 2000. Dative vs. nominative: Changes in quirky subjects in Icelandic. *Leeds Working Papers in Linguistics and Phonetics* 8. 27–44.

Eythórsson, Thórhallur. 2002. Changes in subject case marking in Icelandic. *Syntactic effects of morphological change.* Oxford University Press Oxford 196–212.

Eythórsson, Thórhallur & Jóhanna Barðdal. 2005. Oblique subjects: A common Germanic inheritance. *Language* 81(4). 824–881.

Falk, Cecilia. 1995. Lexikalt kasus i svenska [Lexical case in Swedish]. *Arkiv för nordisk filologi* 110. 199–226.

Fanselow, Gisbert. 2002. Quirky subjects and other specifiers. *More than words* 227–250.

Farkas, Donka F. 1985. Obligatory controlled subjects in Romanian. In *CLS. Papers from the General Session at the 21ˢᵗ Regional Meeting.* Vol. 21. 90–100.

Farkas, Donka F. 1988. On obligatory control. *Linguistics and philosophy.*

Farrell, Patrick. 2005. *Grammatical relations.* Vol. 1. Oxford University Press on Demand.

Fedriani, Chiara. 2009. The Behaviour-before-Coding Principle: Further Evidence from Latin. *Archivio glottologico italiano* 94(2). 156–184.

Fedriani, Chiara. 2011. Experiential metaphors in Latin: feelings were containers, movements and things possessed. *Transactions of the Philological Society* 109(3). 307–326.

Fedriani, Chiara. 2013. The ME PUDET construction in the history of Latin: why and how fast non-canonical subjects come and go.

Fernández-Soriano, Olga. 1999. Two types of impersonal sentences in Spanish: Locative and dative subjects. *Syntax* 2(2). 101–140.

Fillmore, Charles John & Paul Kay. 1993. *Construction Grammar.* University of California.

Fleisher, Nicholas. 2006. Russian dative subjects, case, and control. *Manuscript, University of California, Berkeley.*

Frâncu, Constantin. 1969. Cu privire la "Uniunea Lingvistică Balcanică": Înlocuirea infinitivului prin construcţii personale în limba română veche. *Anuar de lingvistică şi istorie literară* 20. 69–116.

Franks, Steven. 2007. Macedonian Pronominal Clitics as Object Agreement Markers. In Steven Franks, Vrinda Chidambaram & Brian Joseph (eds.), *A Linguistís Linguist: Studies in South Slavic Linguistics in Honor of E. Wayles Browne*, 189–221. Bloomington Indiana: Bloomington.

Geber, Dana. 2006. Processing subject pronouns in relation to non-canonical (quirky) constructions. *Ottawa Papers in Linguistics.*

Geber, Dana. 2011. *Romanian Dative Clitic Dependencies in Raising Constructions.* Université d'Ottawa/University of Ottawa PhD Thesis.

Gheţie, Ion. 1997. *Istoria limbii române literare: Epoca veche (1532-1780).* Bucureşti: Editura Academiei.

Gheorghe, Mihaela. 2013. Argument clauses. In Gabriela Pană Dindelegan, (ed.) *The Grammar of Romanian* (Oxford Linguistics). Oxford: Oxford University Press. 466–473.

Giurgea, Ion. 2017. Preverbal subjects and topic marking in Romanian. *Revue roumaine de linguistique* 63(3). Editura Academiei Române. 279–322.

Giurgea, Ion & E. Remberger. 2009. Postverbal subjects in Romance null-subject languages: Information-structural conditions and variation. *Going Romance.* Nice.

Giurgea, Ion & Eva-Maria Remberger. 2012. *Verum focus and polar questions.* Bibliothek der Universität Konstanz.

Givón, Talmy. 1976. Topic, Pronoun and Grammatical Agreement. In Charles N. Li (ed.), *Subject and Topic.* New York: Academic Press. 149–188.

GLR. 2005. *Gramatica Limbii Române. Vol. I Cuvântul (The Grammar of Romanian Language. Vol. I The Word)*. Valeria Guţu-Romalo (ed.). Bucureşti: Editura Academiei Române.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure.* University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language.* Oxford University Press on Demand.

Goldberg, Adele E. 2019. *Explain me this: Creativity, competition, and the partial productivity of constructions.* Princeton University Press.

Golumbia, David. 2004. The interpretation of nonconfigurationality. *Language & Communication* 24(1). 1–22.

González, Nora. 1988. A constraint on raising in Spanish. *On Spanish, Portuguese, and Catalan Linguistics.* Georgetown University Press 91.

Greenberg, Gerald R. & Steven Franks. 1991. A Parametric Approach to Dative Subjects and the Second Dative in Slavic. *The Slavic and East European Journal* 35(1). 71.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language* 2. 73–113.

Groothuis, Kim A. 2019. Postverbal subject positions in 'semi-finite'clauses in Southern Italo-Romance and Sardinian. In *Romance Languages and Linguistic Theory.* Vol. 15. Frankfurt: John Benjamins Publishing Company. 219–236.

Guéron, Jacqueline. 1986. Le verbe avoir. *Recherches linguistiques de Vincennes* 14(15). 155–187.

Gutiérrez-Bravo, Rodrigo. 2006. A reinterpretation of quirky subjects and related phenomena in Spanish. *Amsterdam Studies In The Theory And History Of Linguistic Science, Series* 4. Amsterdam; Philadelphia; J. Benjamins Pub. Co. 275. 127.

Haeberli, Eric. 1999. *Features, categories and the syntax of A-positions. Synchronic and diachronic variation in the Germanic languages.* University of Geneva PhD Thesis.

Haeberli, Eric. 2002. Syntactic Effects of Morphological Case. In *Features, Categories and the Syntax of A-Positions.* Springer. 101–195.

Haider, Hubert. 2005. How to turn German into Icelandic and derive the OV–VO contrast. *Journal of Comparative Germanic Syntax* 8. 1–53.

Haider, Hubert. 2010. *The syntax of German.* Cambridge University Press.

Hakulinen, Auli. 1983. *Subjektikategoria vai nominaalijäsenten subjektimaisuus.* [*Subject category and the subjecthood of nominal constituents*]. 238–250.

Hakulinen, Auli & Fred Karlsson. 1979. *Nykysuomen lauseoppia* [*Syntax of Modern Finnish*]. Society of Finnish Literature.

Halliday, M. 1985. *An Introduction to Functional Grammar.* Vol. 1. London: Edwa~ Amold.

Haspelmath, Martin. 1998. How young is standard average European. *Language sciences* 20(3). 271–287.

Haspelmath, Martin. 2001a. Non-canonical marking of core arguments in European languages. *Typological Studies in Language* 46. 53–84.

Haspelmath, Martin. 2001b. The European linguistic area: Standard Average European. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook.* Vol. 2. Walter de Gruyter.

Haspelmath, Martin & Sandro Caruana. 2000. Subject diffuseness in Maltese: On some subject properties of experiential verbs. *Folia linguistica* 34(3–4). 245–266.

Hermon, Gabriella. 2001. Non-canonically marked A/S in Imbabura Quechua. *Typological Studies In Language* 46. 149–176.

Hill, Virginia. 2002. Adhering focus. *Linguistic Inquiry.* MIT Press 33(1). 164–172.

Hill, Virginia. 2014. De-Infinitives as Complements to Romanian Nouns. In Enoch Oladé Aboh, Maria Teresa Guasti & Ian Roberts (eds.), *Locality.* Oxford University Press. 166–185.

Hill, Virginia & Alexandru Mardale. 2017. On the interaction of differential object marking and clitic doubling in Romanian. *Revue roumaine de linguistique* 62(4). Editura Academiei Române. 393–409.

Hill, Virginia & Olga Mišeska-Tomić. 2009. A typology of subjunctive complements in Balkan languages. *Bucharest Working Papers in Linguistics* (1). 197–209.

Holvoet, Axel. 2013. Obliqueness, quasi-subjects and transitivity in Baltic and Slavonic. In Serzant, Ilja A. & Leonid Kulikov (eds), *The Diachronic Typology of Non-Canonical Subjects.* John Benjamins Publishing. 257–282.

Holvoet, Axel. 2015. Non-canonical subjects in Latvian: An obliqueness-based approach. *Trends in Linguistics Studies and Monographs* 299.

Holvoet, Axel. 2016. *Argument marking in Baltic and Slavonic pain-verb constructions.* Amsterdam, Philadelphia: John Benjamins Publishing Company.

Hornstein, Norbert. 1999. Movement and control. *Linguistic inquiry.* MIT Press 30(1). 69–96.

Hornstein, Norbert. 2001. *Move! A minimalist theory of construal.* Vol. 10. Wiley-Blackwell.

Hornstein, Norbert. 2003. On control. *Minimalist syntax.* Wiley Online Library 6. 81.

Hyams, Nina & Sigridur Sigurjónsdóttir. 1990. The Development of" Long-Distance Anaphora": A Cross-Linguistic Comparison With Special References to Icelandic. *Language Acquisition.* Taylor & Francis 1(1). 57–93.

Iatridou, Sabine. 1988. Clitics, anaphors, and a problem of coindexation. *Linguistic Inquiry.* JSTOR 19(4). 698–703.

Iatridou, Sabine. 1993. On nominative case assignment and a few related things. *MIT Working Papers in Linguistics.*

Ilioaia, Mihaela. In prep. *Toward canonical marking of core arguments in Romanian? The case of "plăcea".*

Ilioaia, Mihaela. 2020. Productivity of the Romanian MIHI EST Pattern. *Revue roumaine de linguistique* 65(1). Editura Academiei Române. 49–67.

Ilioaia, Mihaela & Marleen Van Peteghem. In prep. Dative experiencers with nominal predicates in Romanian: a synchronic and diachronic study. *Submitted for publication.*

Ionescu, Emil. 2013. Construcţiile cu nume predicativ: o abordare lexicalistă [Constructions with Predicate Nominals: a Lexicalist Approach]. In *Limba română: Perspective actuale.* Polirom. 171–188.

Ionescu-Ruxăndoiu, Liliana. 2010. Tradiţie şi modernitate: Puncte de vedere asupra volumului I din noul Tratat de istorie a limbii române. *Studii de limba română: Omagiu profesorului Grigore Brâncuş.*

Iorga, Ana. 2014 [2013]. O tipologie a dativului românesc. Phd Thesis. University of Bucharest.

Janda, Laura A. 1993. *A Geography of Case Semantics: The Czech Dative and the Russian Instrumental.* Series: Cognitive linguistics Research. Berlin: Mouton de Gruyter.

Jónsson, Jóhannes Gísli. 2003. Not so quirky: On subject case in Icelandic. *New perspectives on case theory.* CSLI Publications Stanford, California 156. 127–163.

Jónsson, Jóhannes Gísli & Thórhallur Eythórsson. 2005. Variation in subject case marking in Insular Scandinavian. *Nordic Journal of Linguistics.* Cambridge University Press 28(2). 223.

Jordan, Maria. 2009. *Loss of infinitival complementation in Romanian diachronic syntax.* University of Florida.

Kachru, Yamuna, Braj B. Kachru & Tej K. Bhatia. 1976. The notion 'subject': a note on Hindi-Urdu, Kashmiri, and Panjabi. *The notion of subject in South-Asian languages.* South Asian Studies Publications Series, University of Wisconsin Madison. 79–109.

Karlsson, Fred. 1982. Kieliteorian relevanssi suomen kielen opetukselle. *Suomi vieraana kielenä.* WSOY Helsinki. 89–114.

Kayne, Richard S. 1993. Toward a modular theory of auxiliary selection. *Studia linguistica.* Wiley Online Library 47(1). 3–31.

Keenan, Edward L. 1976. Towards a universal definition of "subject." In Charles N. Lee (ed.), *Subject and topic.* New York: Academic Press. 303–334.

Keller, Frank, Maria Lapata & Olga Ourioupina. 2002. Using the web to overcome data sparseness. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002).* 230–237.

Keller, Frank & Mirella Lapata. 2003. Using the web to obtain frequencies for unseen bigrams. *Computational linguistics* 29(3). 459–484.

Kempchinsky, Paula Marie. 1986. *Romance subjunctive clauses and logical form.* UCLA.

Khan, Geoffrey A. 1984. Object markers and agreement pronouns in Semitic languages. *Bulletin of the School of Oriental and African Studies* 47(3). 468–500.

Kishimoto, Hideki. 2012. Subject honorification and the position of subjects in Japanese. *Journal of East Asian Linguistics.* Springer 21(1). 1–41.

Kiss, Katalin É. 2002. *The syntax of Hungarian.* Cambridge University Press.

Klaiman, Miriam H. 1980. Bengali dative subjects. *Lingua* 51(4). 275–295.

Kluge, Friedrich. 2002. *Etymologisches Wörterbuch der deutschen Sprache, 24th Editionn by E.* Seebold.

Kortlandt, Frederik HH. 2001. The Indo-Uralic verb. *Finno-Ugrians and Indo-Europeans: Linguistic and literary contacts (= Studia Fenno-Ugrica Groningana 2).* 217–227.

Koster, Jan & Eric Reuland. 1991. *Long distance anaphora.* Cambridge University Press.

Krapova, Iliyana. 1999. Subjunctive complements, null subjects and case checking in Bulgarian. *Amsterdam Studies In The Theory And History Of Linguistic Science.* Series 4. Amsterdam; Philadelphia; J. Benjamins Pub. Co. 239–262.

Krapova, Iliyana. 2001. Subjunctives in Bulgarian and Modern Greek. *Comparative syntax of Balkan languages.* Oxford University Press 105–126.

Laitinen, Lea. 1997. Norms made easy: case marking with modal verbs in Finnish. *Taming the vernacular: From dialect to written standard language* 110–124.

Lakoff, George. 1987. Women, fire, and dangerous things: What categories reveal about the mind. *Chicago: University of Chicago.*

Landau, Idan. 2001. *Elements of control: Structure and meaning in infinitival constructions.* Vol. 51. Springer Science & Business Media.

Landau, Idan. 2004. The scale of finiteness and the calculus of control. *Natural Language & Linguistic Theory.* Springer 22(4). 811–877.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar: Theoretical prerequisites.* Vol. 1. Stanford University Press.

Lauwers, Peter & Dominique Willems. 2011. Coercion: Definition and challenges, current approaches, and new trends. *Linguistics* 49(6). 1219–1235.

Lazard, Gilbert. 2009. Qu'est-ce qu'un sujet? *La linguistique* 45(1). 151–158.

Le Mair, Esther, Cynthia A. Johnson, Michael Frotscher, Thórhallur Eythórsson & Jóhanna Barðdal. 2017. Position as a behavioral property of subjects: The Case of Old Irish. Indogermanische Forschungen. Vol. 122(1), De Gruyter. 111–141.

Legendre, Géraldine. 1989. Inversion with certain French experiencer verbs. *Language.* JSTOR 752–782.

Legendre, Géraldine. 1990. French impersonal constructions. *Natural Language & Linguistic Theory.* Springer 8(1). 81–128.

Lombard, Alf. 1974. *La langue roumaine: Une présentation.* Paris: Éditions Klincksieck.

Madariaga, Nerea. 2011. Infinitive clauses and dative subjects in Russian. *Russian linguistics* 35(3). 301.

Malchukov, Andrej. 2018. Forty years in the search of a/the subject. *Non-Canonically Case-Marked Subjects: The Reykjavík-Eyjafjallajökull papers.* John Benjamins Publishing Company. 200–241.

Manoliu-Manea, Maria. 1987. From conversational to conventional implicature: the Romanian pronouns of identity and their substitutes. In *Papers from the 7th international conference on historical linguistics.* Vol. 48. John Benjamins Publishing. 419.

Manzini, M. Rita & Anna Roussou. 2000. A minimalist theory of A-movement and control. *Lingua.* Elsevier 110(6). 409–447.

Mardale, Alexandru. 2007. Bare plurals and differential object marking in Romanian and Spanish. In *International workshop on DOM in Romance.*

Mardale, Alexandru. 2008. Microvariation within differential object marking. *Revue roumaine de linguistique* 53(4). Editura Academiei Române. 448–467.

Martinet, André. 1960. *Éléments de linguistique générale*. Paris: Armand Colin.

Martinet, André. 1972. Should we drop the notion of subject? *La Revue canadienne de linguistique* 17(2–3). Editura Academiei Române. 175–179.

Martinet, André. 1985. *Syntaxe générale*. Paris: Armand Colin.

Martins, Ana Maria & Jairo Nunes. 2005. Raising issues in Brazilian and European Portuguese. *Journal of Portuguese Linguistics*. Ubiquity Press 4(2). 53–77.

Masica, C. P. 1976. *Defining a Linguistic Area: South Asia.* Chicago: Univ. Chicago Press.

Masullo, Pascual J. 1993. Two types of quirky subjects: Spanish versus Icelandic. In *Proceedings of NELS.* Vol. 23, 303–317.

Matasović, Ranko. 2012. *Clause alignment in Proto-Indo-European.*

Mathieu, Eric. 2006. Quirky subjects in Old French. *Studia Linguistica*. Wiley Online Library 60(3). 282–312.

McCloskey, Jim. 1997. Subjecthood and subject positions. In *Elements of grammar.* Springer. 197–235.

Melis, Chantal & Marcela Flores. 2013. On the historical expansion of non-canonically marked 'subjects' in Spanish. In Serzant, Ilja A. & Leonid Kulikov (eds), *The diachronic Typology of Non-Canonical Subjects.* John Benjamins Publishing. 163–184.

Michaelis, Laura A. & Knud Lambrecht. 1996. Toward a construction-based theory of language function: The case of nominal extraposition. *Language*. JSTOR 215–247.

Monachesi, Paola. 2000. Clitic placement in the Romanian verbal complex. In Birgit Gerlach & Janet Grijzenhout (eds.), *Clitics in Phonology, Morphology and Syntax.* John Benjamins Publishing. 256–294.

Montrul, Silvina & Makiko Hirakawa Chen. 1999. Psych verbs in second language acquisition. *The development of second language grammars: A generative approach*. John Benjamins Publishing 18. 171.

Moore, John & David M. Perlmutter. 2000. What does it take to be a dative subject. *Natural Language & Linguistic Theory* 18(2). 373–416.

Motapanyane, Virginia. 1994. On preverbal positions in Romanian. *Canadian Journal of Linguistics/ Revue canadienne de linguistique.* Cambridge University Press 39(1). 15–36.

Motapanyane, Virginia. 1995. NP-movement from finite clauses in Rumanian. In Guglielmo Cinque & Giuliana Giusti (eds.), *Advances in Roumanian Linguistics*, 161–170.

Nau, Nicole. 1998. Latvian (Languages of the World 217). *München-Newcastle: Lincom Europa.*

Nedelcu, Isabela. 2013. The Subject Of The Romanian Infinitive. *Revue roumaine de linguistique* 58(1). Editura Academiei Române. 41–54.

Nedelcu, Isabela. 2016. The infinitive and the infinitival construction. In Gabriela Pană Dindelegan (ed.), *The Syntax of Old Romanian.* Oxford: Oxford University Press. 232-249.

Nesset, Tore, Anna Endresen & Laura A. Janda. 2011. Two ways to get out: radial category profiling and the Russian prefixes vy-and iz. *Zeitschrift für Slawistik* 56(4). 377–402.

Nicolae, Alexandru. 2013. A note on the syntax of Romanian aspectual verbs: Aspectuals and voice. *„Cuvinte potrivite". Omagiu doamnei Maria Marin la aniversare.*

Noël, Dirk. 2001. The passive matrices of English infinitival complement clauses: Evidentials on the road to auxiliarihood? *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language."* John Benjamins 25(2). 255–296.

OED. 1989. Oxford English Dictionary. *Simpson, JA & Weiner, ESC.*

Onishi, Masayuki. 2001. Non-canonically marked subjects and objects: Parameters and properties. In Aikhenvald, Alexandra Y., Robert MW Dixon & Masayuki Onishi (eds.). *Non-canonical marking of subjects and objects.* Vol. 46, 1–53. Benjamins Amsterdam.

Palmer, Frank Robert. 2001. *Mood and modality.* Cambridge University Press.

Pană Dindelegan, Gabriela. 2013a. The subject. In Gabriela Pană Dindelegan (ed.), *The Grammar of Romanian.* Oxford: Oxford University Press, 100–125.

Pană Dindelegan, Gabriela. 2013b. Romanian – a brief presentation. In Gabriela Pană Dindelegan (ed.), *The Grammar of Romanian.* Oxford: Oxford University Press. 1–7.

Pană Dindelegan, Gabriela. 2013c. *The Grammar of Romanian* (Oxford Linguistics). Oxford: Oxford Univeristy Press.

Pană Dindelegan, Gabriela. 2016. *The Syntax of Old Romanian* (Oxford Studies in Diachronic and Historical Linguistics). Oxford: Oxford Univeristy Press.

Pană Dindelegan, Gabriela. 2016. The subject. In Gabriela Pană Dindelegan (ed.), *The Syntax of Old Romanian* (Oxford Studies in Diachronic and Historical Linguistics). Oxford: Oxford Univeristy Press.

Paykin, Katia & Marleen Van Peteghem. 2017. Nam ne pisalos'i mečtalos': alternating dative reflexive constructions revisited. *Russian Linguistics.* Springer 41(2). 151–175.

Perlmutter, David M. 1982. Syntactic representation, syntactic levels, and the notion of subject. In *The nature of syntactic representation.* Springer. 283–340.

Pesetsky, David. 1987. Binding problems with experiencer verbs. *Linguistic Inquiry.* JSTOR 18(1). 126–140.

Pesetsky, David & Esther Torrego. 2001. T-to-C movement: Causes and consequences. *Current Studies in Linguistics Series.* Mit Press 36. 355–426.

Philippaki-Warburton, Irene & Georgia Catsimali. 1999. On control in Greek. In *Studies in Greek syntax.* Springer. 153–168.

Pooth, Roland, Peter Alexander Kerkhof, Leonid Kulikov & Jóhanna Barðdal. 2019. The origin of non-canonical case marking of subjects in Proto-Indo-European: Accusative, ergative, or semantic alignment. *Indogermanische Forschungen.* De Gruyter 124(1). 245–264.

Postal, Paul. 1970. Cross-over phenomena. *Holt, Rhinehart and Winston.*

Reh, Mechthild. 1998a. *Experiens-Konstruktionen in afrikanischen Sprachen: Erkenntnisinteressen.* Ansätze, Ergebnisse. Reh (ed.), 1-22.

Reh, Mechthild. 1998. The role of metaphor in grammar: Experiencer constructions. *Experiens-Kodierung in afrikanische Sprachen typologisch gesehen: Formen und ihre Motivierung,* Hamburg, Universität Hamburg. 25–40.

Reis, Marga. 1982. *Zum Subjektbegriff im Deutschen [On the subject concept in German].* Sentences in German / Werner Abraham (Ed.).

Rivero, María Luisa. 2004. Spanish Quirky Subjects, Person Restrictions, and the Person-Case Constraint. *Linguistic Inquiry* 35(3). 494–502.

Rivero, María Luisa. 2009. Intensionality, high applicatives, and aspect: involuntary state constructions in Bulgarian and Slovenian. *Natural Language & Linguistic Theory.* Springer 27(1). 151–196.

Rivero, Maria Luisa & Dan Geber. 2004. Romanian Raising: Move and Agree. In *34th Linguistic Symposium on Romance Languages.*

Rivero, María Luisa & Dana Geber. 2003. Quirky subjects and person restrictions in Romance: Rumanian and Spanish. *Cahiers linguistiques d'Ottawa.*

Rivero, María Luisa & Ulyana Savchenko. 2005. Russian anticausatives with oblique subjects. In *Proceedings of FASL 13 The Columbia Meeting.* Michigan Slavic Publications. 267–288.

Rivero, María-Luisa. 2003. Reflexive clitic constructions with datives: syntax and semantics. In *Formal Approaches to Slavic Linguistics.* Vol. 11, 469–494.

Rizzi, Luigi. 1982. Negation, Wh-movement and the null subject parameter. In *An Annotated Syntax Reader.*

Rizzi, Luigi. 1986. Null Objects in Italian and the Theory of pro. *Linguistic Inquiry.* The MIT Press 17(3). 501–557.

Rizzi, Luigi. 2005. On some properties of subjects and topics. In Laura Brugè, Giuliana Giusti, Nicola Munaro, Walter Schweikert & Giuseppina Turano (eds.). *Contributions to the XXX Incontro di Grammatica Generativa* 203–224.

Rögnvaldsson, E. 1996. Word order variation in the VP in Old Icelandic. *Working papers in Scandinavian syntax* 58. 55–86.

Rott, Julian A. 2013. Syntactic prominence in Icelandic Experiencer arguments: quirky subjects vs. dative objects. *STUF-Language Typology and Universals.* De Gruyter 66(2). 91–111.

Roussou, Anna. 2001. Control and raising in and out of subjunctive complements. *Comparative syntax of Balkan languages.* Oxford University Press, USA 74–104.

Sands, Kristina. 2000. Complement clauses and grammatical relations in Finnish. The Australian National University.

Sands, Kristina & Lyle Campbell. 2001. Non-canonical subjects and objects in Finnish. *Typological Studies in Language* 46. 251–306.

Schulte, Kim. 2007. *Prepositional Infinitives in Romance: A Usage-based Approach to Syntactic Change.* Peter Lang.

Schultink, Henk. 1961. Produktiviteit als morfologisch fenomeen. In *Forum der letteren.* Vol. 2, 110–125.

Seefranz-Montag, Ariane von. 1983. *Syntaktische Funktion und Wortstellungsveränderung: Die Entwicklung" subjektloser" Konstruktionen in einigen Sprachen.* Wilhelm Fink.

Seefranz-Montag, Ariane von. 1984. "Subjectless" constructions and syntactic change. *Historical syntax* 521–553.

Seržant, Ilja. 2015. Dative experiencer constructions as a Circum-Baltic isogloss. *Contemporary Approaches to Baltic Linguistics. Berlin, Boston: De Gruyter Mouton* 325–348.

Seržant, Ilja A. 2013. Rise of canonical objecthood with the Lithuanian verbs of pain. *Baltic linguistics* 4.

Shibatani, Masayoshi. 1977. Grammatical relations and surface cases. *Language.* JSTOR 789–809.

Shibatani, Masayoshi. 2001. Non-canonical constructions in Japanese. *Typological studies in language* 46. 307–354.

Sigurðsson, Halldór Ármann. 1992 (1989). *Verbal syntax and case in Icelandic in a comparative GB approach.* PhD. Thesis. University of Lund

Sigurðsson, Halldór Ármann. 2002. To be an oblique subject: Russian vs. Icelandic. *Natural Language & Linguistic Theory* 20(4). 691–724.

Sigurðsson, Halldór Ármann. 2004. Icelandic non-nominative subjects: Facts and implications. In Peri Bhaskararao and KV Subbarao (eds.), *Non-nominative subjects.* Vol 2, John Benjamins Publishing. 137–159.

Sigurðsson, Halldór Ármann. 2006. The Nom/Acc alternation in Germanic. *Comparative studies in Germanic syntax* 13–50.

Smitherman, Thomas & Jóhanna Barðdal. 2009. Typological changes in the evolution of Indo-European syntax? *Diachronica* 26(2). 253–273.

Soare, Elena. 2009. Epistemic modality in Romanian: The role of BE. *Studies in Linguistics. Proceedings XXXV Incontro di Grammatica Generativa.* 3. 244–257.

Spătaru-Pralea, M. 2013. *Concurenţa infinitiv/conjunctiv în limba română: o călătorie în timp şi spaţiu.* Editura Universitară.

Stan, Camelia. 2007. Notă gramaticală: conjuncţia "ca (...) să". *Studii şi cercetări lingvistice.* Vol. LVIII (2). 451–458.

Stan, Camelia. 2013. *O sintaxă diacronică a limbii române vechi.* Bucureşti: Editura Universităţii din Bucureşti.

Ştefănescu, Ioana. 1997. *The Syntax of Agreement in Romanian.* Vol. 14. PhD Thesis.

Sulkala, Helena & Merja Karjalainen. 1992. *Finnish.* Psychology Press.

Terzi, Arhonto. 1992. PRO and Obviation in Modern Greek Subjunctives. In *Proceedings of the 10th West Coast Conference on Formal Linguistics.* Vol. 10. Arhonto Terzi CUNY Graduate Center. Center for the Study of Language (CSLI). 471.

Terzi, Arhonto. 1993. PRO in finite clauses: A study of the inflectional heads of the Balkan languages.

Terzi, Arhonto. 1997. PRO and null case in finite clauses. *The Linguistic Review.* De Gruyter Mouton 14(4). 335–360.

Thráinsson, Hoskuldur. 1979. *On Complementation in Icelandic*, Garland, New York.

Thráinsson, Höskuldur. 2007. *The syntax of Icelandic.* Cambridge University Press.

Turano, Giuseppina. 1994. *Syntactic dependencies in Albanian.* Florence: University of Florence.

Van Peteghem, Marleen. 2016. Verbs of pain and accusative subjects in Romanian. *Atypical predicate-argument relations* 33. 3.

Van Peteghem, Marleen. 2017. Les structures de la douleur: sur le marquage de l'expérienceur dans les langues romanes. In *De la passion du sens en linguistique: a Danièle Van de Velde.* Presses Universitaires de Valenciennes. 439–463.

Van Peteghem, Marleen & Mihaela Ilioaia. In prep. *Expérienceurs datifs et accusatifs en roumain.*

Van Peteghem, Marleen & Mihaela Ilioaia. 2017. Nu mi-e frică de nimic: une structure MIHI EST en roumain? In A. Costăchescu & C. M. Popescu (eds.), *Hommages offerts à Maria Iliescu.* Editura Universitaria Craiova. 313–327.

Van Riemsdijk, H.C. & E. Williams. 1986. *Introduction to the Theory of Grammar.* MIT Press.

Van Valin, Robert D., Robert D. van Valin Jr, Robert D. van Valin Jr & Randy J. LaPolla. 1997. *Syntax: Structure, meaning, and function.* Cambridge University Press.

Varlokosta, Spyridoula & Norbert Hornstein. 1993. Control in Modern Greek. In *North East Linguistics Society.* Vol. 23, 17.

Vasilescu, Andra. 2013. Sentence types. In Gabriela Pană Dindelegan (ed.), The Grammar of Romanian. Oxford: Oxford University Press. 537–548.

Vasilescu, Andra. 2017. The descendants of Lat. *ipse* in Romanian. *Diacronia.* Editura Universităţii "Alexandru Ioan Cuza" din Iaşi (6). 1–24.

Vasiliu, E. & L. Ionescu-Ruxăndoiu. 1986. Limba română în secolele al XII-lea – al XV-lea (fonetică – fonologie – morfologie). Bucureşti.

Vázquez-Rosas, Victoria. 2006. *Gustar*-type verbs. In *Functional approaches to Spanish syntax.* Springer. 80–114.

Verhoeven, Elisabeth. 2007. *Experiential constructions in Yucatec Maya.* Amsterdam: Benjamins.

Verma, Mahendra K. & Karuvannur Puthanveettil Mohanan. 1990. *Experiencer subjects in South Asian languages.* Center for the Study of Language (CSLI).

Verma, Manindra K. 1976. *The notion of subject in South Asian languages.* University of Wisconsin–Madison.

Vilkuna, Maria. 1989. *Free word order in Finnish.* Suomalaisen Kirjallisuuden Seura. 280

Vivès, Robert. 1984. L'aspect dans les constructions nominales prédicatives: avoir, prendre, verbe support et extension aspectuelle. *Lingvisticae investigationes* 8(1). 161–185.

Willi, Andreas. 2018. *Origins of the Greek verb.* Cambridge University Press.

Yoon, James. 1996. Ambiguity of government and the Chain Condition. *Natural language & linguistic theory.* Springer 14(1). 105–162.

Yoon, James Hye-Suk. 2009. The distribution of subject properties in multiple subject constructions. *Japanese/Korean Linguistics.* CSLI Publications Stanford, CA 16. 64–83.

Zaenen, Annie, Joan Maling & Höskuldur Thráinsson. 1985. Case and grammatical functions: The Icelandic passive. *Natural Language & Linguistic Theory.* Springer 3(4). 441–483.

Zafiu, Rodica. 2012. Particula -şi: între intensificare şi indefinire. In Rodica Zafiu, Adina Dragomirescu, Alexandru Nicolae & Adriana Stefanescu (eds.), *Limba roână: direcţii actuale în cercetarea lingvistică*, 277–285. Bucureşti: Editura Universităţii din Bucureşti.

Zec, Draga. 1987. On obligatory control in clausal complements. *Working Papers in Grammatical Theory and Discourse Structure.* CSLI Publications 1. 139–168.

Zeldes, Amir. 2012. *Productivity in argument selection: From morphology to syntax.* Vol. 260. Walter de Gruyter.

Zribi-Hertz, Anne. 1996. *L'anaphore et les pronoms: une introduction à la syntaxe générative.* Presses Univ. Septentrion.

Appendices

Appendix 1.

Collection of Romanian texts (16ᵗʰ – 20ᵗʰ)

The corpus for old Romanian is established based on a subset of the texts employed in the Syntax of Old Romanian coordinated by Pană Dindelegan (2016). Each abbreviation is followed by a year that marks the date of the text. The localization of some of the texts is given, where possible, at the end of the entry in round brackets. The corpora for modern and contemporary Romanian have been totally collected by myself, from texts provided by several scholars.

Corpus of old Romanian (16ᵗʰ – 18ᵗʰ centuries)

A.1620	*Alexandria.* Ed. F. Zgraon. Bucharest: Fundaţia Naţională pentru Ştiinţă şi Artă. 2005 (*Cele mai vechi cărţi populare în literatura română.* 11). (South Transylvania, Braşov or Haţeg)
AA.1708	*Archirie şi Anadan.* Ed. M. Georgescu. Bucharest: Minerva. 1997 (*Cele mai vechi cărţi populare în literatura română.* 2). 157–168. (North Transylvania)
AD.1722–1725	Antim Ivireanul. 1972. *Didahii.* Ed.: Antim Ivireanul. *Opere.* ed. G. Ştrempel. Bucharest: Minerva. 238 p. (Wallachia, Bucharest)
AOD.1675–1676	*Alexie, omul lui Dumnezeu.* Ed. M. Stanciu-Istrate. Bucharest: Fundaţia Naţională pentru Ştiinţă şi Artă. 2001 (*Cele mai vechi cărţi populare în literatura română.* 5). 89–101. (North Moldova)
BB.1688	*Biblia.* Ed.: *Biblia adecă Dumnezeiasca Scriptură a Vechiului şi Noului Testament.* tipărită întâia oară la 1688 în timpul lui Şerban Vodă Cantacuzino. Domnul Ţării Româneşti. Bucharest: Editura Institutului Biblic. 1977. (Wallachia, Bucharest)
Bert.1774	*Bertoldo.* Ed. Magdalena Georgescu. Bucharest: Minerva. 1999 (*Cele mai vechi cărţi populare în literatura română.* 3). 157–239. (Moldova)
CazV.1643	Varlaam. *Cazania.* ed. J. Byck. Bucharest: Editura Academiei. [s.a.]. 1– 506. (Moldova)
CC¹.1567	Coresi. *Tâlcul Evangheliilor.* Ed.: Coresi. *Tâlcul evangheliilor şi molitvenic românesc.* ed. V. Drimba. Bucharest: Editura Academiei Române. 1998. 31–187 (Transylvania, Wallachian subdialect; Gheţie and Mareş 2001: 115)
CC².1581	Coresi, *Evanghelie cu învăţătură.* Ed. S. Puşcariu. Al. Procopovici: Diaconul Coresi. *Carte cu învăţătură (1581).* vol. I. *Textul.* Bucharest: Socec. 1914. (Braşov)
CCat.1560	Coresi. *Catehism.* Ed. Al. Roman-Moraru. in I. Gheţie (coord.). *Texte româneşti din secolul al XVI-lea.* I. *Catehismul lui Coresi;* II. *Pravila lui Coresi;* III. *Fragmentul Todorescu;* IV. *Glosele Bogdan;* V. *Prefeţe şi Epiloguri.* Bucharest: Editura Acadmiei Române. 1982. 101–105. (Braşov)
CD.1698	Dimitrie Cantemir. *Divanul.* Ed.: D. Cantemir. *Opere complete.* I. *Divanul.* ed. V. Cândea. Bucharest: Editura Academiei. 1974. 103–405. (Moldova, Iaşi)

CH.1717–1723	Dimitrie Cantemir. *Hronicul vechimei a romano moldo-vlahilor*. ed. S. Toma. Bucharest: Minerva. 1999-2000. 1–274 (vol. I). 5–223 (vol. II). (Moldova)
CII.~1705	Dimitrie Cantemir. *Istoria ieroglifică*. Ed.: D. Cantemir. *Opere complete*. IV. *Istoria ieroglifică*. ed. S. Toma. Bucharest: Editura Academiei. 1974. 51–289. (Moldova)
CIst.1700–1750	Constantin Cantacuzino. *Istoria Ţării Româneşti*. Ed.: *Istoria Ţărâi Rumâneşti atribuită stolnicului Constantin Cantacuzino*. Ed. O. Dragomir. Bucharest: Editura Academiei Române. 2006. 145–202. (Bucharest)
CÎ.1678	*Cheaia înţelesului*. Ed. Ioannykij Haleatovskyi. *Cheia înţelesului*. ed. R. Popescu. Bucharest: Libra. 2000. 13–194. (Bucharest)
CLM.1700–1750	Miron Costin. *Letopiseţul Ţărâi Moldovei*. Ed.: M. Costin. *Opere*. ed. P. P. Panaitescu. Bucharest: Editura de Stat pentru Literatură şi Artă. 1958. 41–201. (Moldova)
CNM.1700–1750	Miron Costin. *De neamul moldovenilor. din ce ţară au ieşit strămoşii lor.* Ed.: M. Costin. *Opere.* ed. P. P. Panaitescu. Bucharest: Editura de Stat pentru Literatură şi Artă. 1958. 241–274. (Moldova and Braşov)
CP[1].1577	Coresi. *Psaltire slavo-română.* Ed.: Coresi. *Psaltirea slavo-română (1577) în comparaţie cu psaltirile coresiene din 1570 şi din 1589.* ed. S Toma. Bucharest: Editura Academiei RSR. 1976. 35–662. (Braşov, Wallachian subdialect)
CPrav.1560–1562	Coresi. Pravila. Ed. Gh. Chivu. in I. Gheţie (coord.). *Texte româneşti din secolul al XVI-lea.* 218–231. (Braşov)
CT.1560–1561	Coresi. *Tetraevanghel.* Ed.: *Tetraevanghelul tipărit de Coresi. Braşov 1560 – 1561.* comparat cu *Evangheliarul lui Radu de la Măniceşti. 1574.* ed. F. Dimitrescu. Bucharest: Editura Academiei. 1963. (Wallachian subdialect. Braşov)
CTd.1600–1640	*Codicele Todorescu.* Ed.: N. Drăganu. *Două manuscripte vechi. Codicele Todorescu şi Codicele Marţian.* 191–229. (North Transylvania. ILRL: 83).
DDL.1679	Dosoftei. *Dumnezăiasca liturghie.* Ed. N. A. Ursu. Iaşi: Mitropolia Moldovei şi Sucevei. 1980. 3–313. (Moldova, Iaşi)
DIR.A	*Documente privind istoria României.* veacul XVII. A. Moldova. Bucharest: Editura Academiei. (1612); 1956: IV (1616–1620); 1957: V (1621–1625). (Moldova)
DIR.B	*Documente privind istoria României.* veacul XVII. B. Ţara Românească. Bucharest: Editura Academiei. (1606); (1607); (1614); 1954: IV (1621–1625). (Wallachia)
DÎ	*Documente şi însemnări româneşti din secolul al XVI-lea.* text stabilit şi indice de Gh. Chivu. M. Georgescu. M. Ioniţă. Al. Mareş. Al. Roman- Moraru. Bucharest: Editura Academiei Române. 1979.
DPV.1673	Dosoftei. *Psaltirea în versuri.* Ed.: Dosoftei. *Opere.* 1. Versuri. ed. N. A. Ursu. Iaşi: Mitropolia Moldovei şi a Sucevei. 1974. 3–1065. (Ukraine, Uniev)
DRH.A	*Documenta Romaniae Historica. A. Moldova.* Bucharest: Editura Academiei Române. 1996: vol. XXIII (1635–1636); 2006: XXVIII (1645 – 1646). (Moldova)
DRH.B	*Documenta Romaniae Historica. B. Ţara Românească.* Bucharest: Editura Academiei Române. 1965: vol. XXI (1626); 1969: vol. XXIII (1630–2); 1974: vol. XXIV (1633–4); 1998: vol. XXX (1645); 2003: vol. XXXI (1646); 2002: vol. XXXIV (1649); vol. XXXV (1650); 2006: vol. XXXVII (1652); 2009: vol. XXXVIII (1653). (Wallachia)
DVS.1682–1686	Dosoftei. *Viaţa şi petreacerea svinţilor.* Iaşi. (Moldova)
FN.1693–1704	*Foletul Novel. Calendarul lui Constantin Vodă Brâncoveanu (1693- 1704).* ed. E. Vârtosu. Bucharest: Monitorul Oficial şi Imprimeriile Statului. 1942. 4–194. (Wallachia)

FT.1571–1575	*Fragmentul Todorescu (Carte de cântece).* Ed. I. Gheţie. in I. Gheţie (coord.). *Texte româneşti din secolul al XVI-lea.* 336–343. (Transylvania, Cluj; Gheţie and Mareş 2001: 95)
GB.XVI-XVII	*Glosele Bogdan.* Ed. M. Georgescu in I. Gheţie (coord.). *Texte româneşti din secolul al XVI-lea.* 422–438. (North Moldova)
GIst.~1750	Radu Greceanu. *Începătura istoriii vieţii luminatului şi preacreştinului Domnului Ţării Rumâneşti.* Ed. *Cronicari munteni.* vol. II. ed. M. Gregorian. Bucharest: Editura pentru Literatură. 1961. 5–272. (Wallachia)
LD	*Legenda duminicii.* Ed. E. Timotin. Bucharest: Fundaţia Naţională pentru Ştiinţă şi Artă. 2005 (*Cele mai vechi cărţi populare în literatura română.* 10).
	LDI.1601–post1619 – ms. 447 = *Codex Sturdzanus.* 313–319
	LDIII.1630–1650 – ms. 4746 BCU Cluj. 329–340. (Sălaj; Mareş 2003: 151) LDIV.1659–1681 – ms. 26 Şcheii Braşovului. 341–352. (Braşov)
	LDVa.1678 – ms. 5910. 353–363. (Crişana, Bihor)
	LDVb.1680–1692 –ms. 4182. 364–369. (Crişana, Bihor)
	LDVI.1732 – ms. 701. 370–377. (Crişana, Bihor)
	LDVIII.1725–1750 – ms. 1317. 389–391. (Oltenia, Râmnic)
Mărg.1691	*Mărgăritare.* Ioan Gură de Aur. *Mărgăritare.* ed. R. Popescu. Bucharest: Libra. 2001. 11–493. (Wallachia, Bucharest)
MC.1620	M. Moxa. *Cronograf.* Ed.: Mihail Moxa. *Cronica universală.* ed. G. Mihăilă. Bucharest: Minerva. 1989. 95–223. (Wallachia, Cozia Monastery)
MI.~1630	*Manuscrisul de la Ieud.* Ed. M. Teodorescu. I. Gheţie. Bucharest: Editura Academiei. 1977. 153–170. (North Transylvania, Maramureş)
NÎnv.~1700	*Învăţăturile lui Neagoe Basarab către fiul său Teodosie.* Ed. F. Moisil. D. Zamfirescu. Bucharest: Minerva. 1971. 125–352. (Wallachia, Bucharest)
NL.~1750–1766	Ion Neculce. *Letopiseţul.* Ed.: Ion Neculce. *Letopiseţul Ţării Moldovei şi O samă de cuvinte.* ed. I. Iordan. Bucharest: Editura de Stat pentru Literatură şi Artă. ed. a II-a. 1959. 31–388. (Moldova and Wallachia)
NT.1648	*Noul Testament.* Ed. Alba Iulia: Reîntregirea. 1998. (Alba Iulia)
PE	Prefeţe şi epiloguri din secolul al XVI-lea. Ed. E. Buză. F. Zgraon. in I. Gheţie (coord.). *Texte româneşti din secolul al XVI-lea.* 555–571.
PH.1500–1510	*Psaltirea Hurmuzaki.* ed. I. Gheţie and M. Teodorescu. Bucharest: Editura Academiei Române. 2005. (Moldova)
PIst.~1780	Radu Popescu. *Istoriile domnilor Ţării Româneşti.* Ed. *Cronicari munteni.* I. ed. M. Gregorian. 225–577. (Wallachia)
PO.1582	*Palia de la Orăştie.* Ed. V. Pamfil. Bucharest: Editura Academiei. 1968. (Banat-Hunedoara; Gheţie and Mareş 2001: 99)
Prav.1581	*Pravila ritorului Lucaci.* Ed. I. Rizescu. Bucharest: Editura Academiei. 1971. 161–183. (Moldova. Putna Monastery)
Prav.1780	*Pravilniceasca condică.* 1780. ed. Colectivul pentru vechiul drept românesc condus de acad. A. Rădulescu. Bucharest: Editura Academiei. 1957 (*Adunarea izvoarelor vechiului drept românesc scris.* 2). 36–156. (Wallachia, Bucharest)

TD	E. Timotin. *Decântecele manuscrise româneşti (secolele al XVII-lea – al XIX-lea).* Bucharest: Editura Academiei Române. 2010

TDI.1601–1618 – ms. 447 (= CS). 237. (Alba)

TDII.ante1700 – ms. 2267. 238. (Sălaj)

TDIII.ante1700 – ms. 5910. 239. (Crişana, Bihor)

TDIV.1720 – ms. 3137. 240. (North Dacoromania)

TDV.~1754 – ms. 34. 241. (Crişana, Bihor)

TDVI.1757 – ms. 5911. 242–243. (Arad)

TDVII.~1750 – ms. 4254. 244. (Transylvania)

TDVIII.1733–1783 – ms. 10. 245–246. (Transylvania)

TDXI.post1600 – ms. 418. 249. (North Dacoromania)

TDXII.1750–1800 – ms. 1507. 250. (Wallachia)

TDXIII.1750–1800 – ms. 1320. 251. (Wallachia)

TDXIV.1777 – ms. 2183. 252. (Oltenia)

TDXXIV.1601–1618 – ms. 447 (= *Codex Sturdzanus*). 262. (Alba)

TDXXV.1750–1800 – ms. 1507. 263. (Wallachia)

TDXXXV.1676 – ms. 5318. 276. (North Dacoromania)

TDXXXVI.1750–1800 – ms. 1320. 277. (Wallachia)

TDXXXVIII.1779 – ms. 1739. 279. (Wallachia)

TDXXXIX.1777–1784 – ms. 4104. 280. (Republic of Moldova, Tighina) TDXLVIII.1777 – ms. 2183. 289. (Oltenia)

TDL.1750–1800 – ms. 1320. 291. (Wallachia)

TDLIII.~1754 – ms. 34. 294–295. (Crişana, Bihor)

TDLIV.1779 – ms. 1739. 296. (Wallachia)

TDLXI.1779 – ms. 1739. 303. (Wallachia)

TDLXIII.~1754 – ms. 34. 305. (Crişana, Bihor)

TDLXXV.1779 – ms. 1739. 317. (Wallachia)

TDLXXVII.1779 – ms. 1739. 319. (Wallachia)

TDLXXXI.1777 – ms. 2183. 323. (Oltenia)

TDLXXXIV.1777–1784 – ms. 4104. 326. (Republic of Moldova, Tighina)

TDXCVIII.1750–1800 – ms. 1507. 345. (Wallachia)

ULM.~1725	Grigore Ureche. *Letopiseţul Ţării Moldovei.* Ed. P. P. Panaitescu. Bucharest: Editura de Stat pentru Literatură şi Artă. 1955. 57–210. (Wallachia, original from Moldova)
VE.1703	*Viaţa lui Esop.* Ed. Violeta Barbu. Bucharest: Minerva. 1999. 175–208. (Braşov)
VM	*Cele douăsprezece vise în tâlcuirea lui Mamer.* Ed. Al. Mareş. Bucharest: Fundaţia Naţională pentru Ştiinţă şi Artă. 2003 (*Cele mai vechi cărţi populare în literatura română.* 8).

VMI.~1750 – ms. 5054. 193–199. (Wallachia)

VMII.1643–1654 – ms. 6. 200–208. (Maramureş)

VMIII.1625–1700 – ms. sl. 75. 209–213. (Banat or South-West Hunedoara)

VRC.1645	Varlaam. *Răspunsul împotriva catihismusului calvinesc.* Ed.: Varlaam. *Opere. Răspunsul împotriva catihismusului calvinesc.* ed. M. Teodorescu. Bucharest: Minerva. 1984. 143–230. (Wallachia, Târgovişte, Dealu Monastery)

Corpus of modern Romanian (19th century)

BD.1760–1820	Budai-Deleanu, Ion. *Ţiganiada* / Ion Budai-Deleanu; Prolog: Leonachi Dianeu. Bucureşti: Editura Litera. 1997. – 368 p.
CO.1890	Creangă, Ion. 1993. *Opere*. Editura Fundaţiei Culturale Române. Colecţia Fundamente
DE.1800–1830	Limona, D. and Trandafirescu, N. 1983. *Documente economice din arhiva Casei Comerciale Ioan Stamu (1714-1876)*. Vol. 1. Bucureşti: Direcţia Generală a Arhivelor Statului.
DE.1830–1859	Limona, D. and Trandafirescu, N. 1983. *Documente economice din arhiva Casei Comerciale Ioan Stamu (1714-1876)*. Vol. 2. Bucureşti: Direcţia Generală a Arhivelor Statului.
DP.1878–1883	Delavrancea, Barbu Ştefănescu. 2011. *Basme şi povestiri*. Bucureşti : Editura TEDIT FZH. 191 p.
DS.1805	Donici, Alecu. 1997. *Scrieri*. Litera Chişinău.
EB.1872–1902	Eminescu, Mihai. 1987. *Basme*. Bucureşti: Editura Ion Creangă. 128 p.
El.1786	*Eliodor*. Ed. M. Georgescu. Bucharest: Minerva. 1997 (*Cele mai vechi cărţi populare în literatura română. 2*)
EP.1872–1902	Eminescu, Mihai. 1996. *Proză literară*. Litera. Chişinău. 316 p.
Er.1785	*Erotocrit*. Ed. M. Georgescu. Bucharest: Minerva. 1997 (*Cele mai vechi cărţi populare în literatura română. 2*)
F.1863	Filimon, Nicolae. 1997. *Ciocoii vechi şi noi*. Litera. Chişinău.
F.ante1837	*Filerot*. Ed. M. Georgescu. Bucharest: Minerva. 1997 (*Cele mai vechi cărţi populare în literatura română. 2*)
I.1789	*Imberie*. Ed. M. Georgescu. Bucharest: Minerva. 1997 (*Cele mai vechi cărţi populare în literatura română. 2*)
IP.1700–1799	*Imparatia poamelor*. Ed. M. Georgescu. Bucharest: Minerva. 1997 (*Cele mai vechi cărţi populare în literatura română. 2*)
MS.I.1783	Maior, Petru. *Scrieri*. vol. I. 1976 [1783]. ediţie critică alcătuită de Florea Fugariu. Prefaţă şi tabel cronologic de Maria Protase. Bucureşti. Editura Minerva.
MS.II.1783	Maior, Petru. *Scrieri*. vol. II. 1976 [1783]. ediţie critică alcătuită de Florea Fugariu. Prefaţă şi tabel cronologic de Maria Protase. Bucureşti. Editura Minerva.
PC.1824	*Patru corăbieri*. Ed. M. Georgescu. Bucharest: Minerva. 1997 (*Cele mai vechi cărţi populare în literatura română. 2*)
SM.1894	Slavici, Ioan. 1996. *Mara*. Litera. Chişinău
SN.bef1900	*Nuvele*, Slavici. Ioan. 1996. *Moara cu noroc: Nuvele şi poveşti*. Litera. Chişinău
VA1793	*Viata lui Anastasie*. In *Cele mai Vechi Cărţi Populare în Literatura Română*. Volumul VI - Scrieri Eshatologice Postbizantine, Vedenia Sofianei , Vedenia Lui Chir Daniil. 2002. Editată De Academia Română. Fundaţia Naţională Pentru Ştiinţă şi Artă. 176 p.
VD.1793	*Vedenia lui Daniil*. In *Cele mai Vechi Cărţi Populare în Literatura Română*. Volumul VI - Scrieri Eshatologice Postbizantine, Vedenia Sofianei, Vedenia Lui Chir Daniil. 2002. Editată De Academia Română. Fundaţia Naţională Pentru Ştiinţă şi Artă. 176 p.
VI.1791	*Varlaam si Ioasaf*. Ed. M. Georgescu. Bucharest: Minerva. 1997 (*Cele mai vechi cărţi populare în literatura română. 2*)

Corpus of contemporary Romanian (20th century)

CC.1929	Caragiale, Mateiu. 2009. *Craii de Curtea-Veche*. Litera. Chişinău.
CE.1938	Călinescu, George. 2010. *Enigma Otiliei*. Litera. Chişinău.
CV.1932	Călinescu, George. 1997. *Viaţa lui Eminescu*. Litera. Chişinău.
DL.1966–1994	Dragomir, Constantin. 2002. *Legende populare româneşti*. Litera. Chişinău.
DT.1878–1913	Delavrancea, Barbu Ştefănescu. 1996. *Teatru*. Litera. Chişinău.
EC.1913	Eftimiu, Victor. 1997. *Cocoşul negru*. Litera. Chişinău.
IA.1933	Ibrăileanu, Garabet. 2009. *Adela*. Litera. Chişinău.
P-BC.1927	Papadat-Bengescu, Hortensia. 2010. *Concert din muzică de Bach*. Litera. Chişinău.
PU.1930	Petrescu, Camil. 2012. *Ultima noapte de dragoste. Întâia noapte de război*. Agora.
RP.1922	Rebreanu, Liviu. 2010. *Pădurea spânzuraţilor*. Litera Chişinău.
RR.1932	Rebreanu, Liviu. 1998. *Răscoala*. Litera Chişinău.
SA.1909	Sadoveanu, Mihail. 2004. *Alexandria. Esopia*. Prut. 216p.
SP.1908	*Poveşti*, Slavici. Ioan. 1996. *Moara cu noroc: Nuvele şi poveşti*. Litera. Chişinău
ST.1926	Sadoveanu, Mihail. 2015. *Ţara de dincolo de negură*. Ed. Mihail Sadoveanu.

Appendix 2. Sentences Survey presented in Chapter 8, Section 8.1 – translation sentences. The "***" indicates sentences with control nouns.

Nr.	Q_Nr	Sentence in Romanian	Sentence in English	Average Rating	Verb
1	Q28***	*Mi-era foarte **frică** înainte de a intra în sala de examen.*	I was very afraid before entering the exam room.	6,74	*fi*
2	Q78***	*Îi era şi mai **frică** văzând că sala era plină.*	He was even more afraid, seing that the room was full.	6,63	*fi*
3	Q55***	*Îi e **frică** de ea însăşi.*	She is afraid of herself.	6,57	*fi*
4	Q19***	*Îi era **jenă** văzând că toţi o privesc.*	She felt emarrassed seing that everybody stares at her.	6,55	*fi*
5	Q33***	*Mi-era **jenă** înainte de a-mi mărturisi greşeala.*	I was feeling embarrassed even before admitting my mistake.	6,24	*fi*
6	Q23***	*Am mare **oroare** de insecte.*	I am horrified of insects.	6,17	*avea*
7	Q64	*Am un mare **chin** în suflet.*	There's great anguish in my heart.	5,79	*avea*
8	Q52	*A avut mare **îndurare** de noi!*	He had a lot of pity for us!	5,77	*avea*
9	Q71	*Am o **amărăciune** în suflet, de neînţeles!*	There's so much sorrow in my soul, nobody can understand it!	5,69	*avea*
10	Q21***	*Avea **necaz** pe fratele său.*	He felt grudge/ rancor/ envy/ resentment toward his brother.	5,38	*avea*
11	Q51***	*Mi-e **oroare** de accidente.*	I'm horrified by accidents.	5,33	*fi*
12	Q1***	*Îi era **jenă** de ea însăşi.*	She was embarrassed of herself.	5,22	*fi*
13	Q27***	*Îmi era şi mai **necaz** văzând că el nu venise.*	I was even more upset to see that he had not come.	5,13	*fi*
14	Q62***	*Am mare **frică** să nu se simtă cumva deranjat.*	I'm very afraid he'll be bothered.	5,05	*avea*
15	Q66	*Am mare **zbucium** în suflet.*	There's so much turmoil in my soul.	4,83	*avea*
16	Q46	*Am **ranchiună** în suflet când îi privesc.*	I have resentment in my soul when I look at them.	4,71	*avea*
17	Q69	*Am avut mare **mânie** în suflet la aflarea veştii.*	I was very angry when I heard the news.	4,67	*avea*
18	Q34	*Când intru în casa ta, am mare **tihnă** în suflet.*	When I enter your house, I have great peace in my soul.	4,65	*avea*
19	Q9***	*Îi era **târşă** de aşa oameni.*	She felt disgust for such people.	4,64	*fi*
20	Q37	*Am o mare **răceală** în suflet.*	I feel very indifferent.	4,61	*avea*
21	Q43	*Are mare **pizmă** pe fratele ei!*	She is very jealous of her brother!	4,56	*avea*
22	Q60***	*Îi era **oroare** înainte de a intra la examenul oral.*	He was horrified before taking the oral exam.	4,32	*fi*
23	Q47***	*Mi-era **târşă** auzindu-l cum striga.*	I was ashamed hearing him shout.	4,24	*fi*
24	Q20***	*Mi-era **deja** târşă înainte de a-l vedea.*	I was already horrified before I saw him.	4,10	*fi*
25	Q80	*Mare **râvnă** are de a urca munţii!*	He has great zeal to climb the mountains!	3,97	*avea*
26	Q14***	*Îi era **necaz** pe sine însuşi.*	He was upset with himself.	3,95	*fi*
27	Q25***	*Am mare **târşă** de oameni ca el.*	I have a great horror of people like him.	3,87	*avea*
28	Q76	*Are mare **sfială** de băiatul vecinilor.*	She's very shy about the neighbors' boy.	3,72	*avea*

Nr.	Q_Nr	Sentence in Romanian	Sentence in English	Average Rating	Verb
29	Q72***	*Ne fusese **oroare** văzând atâtea nedreptăţi.*	We had been horrified to see so many injustices.	3,70	*fi*
30	Q7	*Am mare **temere** de a nu-l fi supărat.*	I'm afraid I upset him.	3,66	*avea*
31	Q73	*Îi era **pizmă** pe ea ştiind că va câştiga concursul.*	He was jealous of her knowing she would win the contest.	3,35	*fi*
32	Q61	*Mi-e **amărăciune** în suflet.*	My soul is full of grief.	3,28	*fi*
33	Q56	*Mi-e **amărăciune** în suflet fără a avea un motiv întemeiat.*	There is griefs in my heart without any good reason.	3,23	*fi*
34	Q48	*Am **descurajare** in suflet când mă gândesc la relaţia noastră.*	I feel discouraged when I think about this relationship.	3,23	*avea*
35	Q24	*Le este **ranchiună** fără a cunoaşte încă tot adevărul.*	They hold a grudge without knowing the whole truth.	3,14	*fi*
36	Q68	*Îi era **chin** văzând atâta suferinţă.*	Seeing so much sufference was an ordeal to him.	3,02	*fi*
37	Q3***	*Am mare **jenă** de vecina mea.*	I am ashamed of my neighbour.	2,95	*avea*
38	Q2***	*Mi-era **necaz** înainte să intru în sala de examen.*	I was troubled before getting into the exam room.	2,93	*fi*
39	Q38	*Nu ţi-e **îndurare** de acel om sărac?*	Don't you take pity on that poor man?	2,91	*fi*
40	Q77	*Mi-e **chin** în suflet.*	I have a troubled soul.	2,91	*fi*
41	Q18	*Mi-e **tihnă** lângă tine.*	I feel at ease when I am with you.	2,87	*fi*
42	Q35	*Mi-era **râvnă** deja, fără să-i fi văzut formele perfecte.*	I already wanted her without having seen her perfect curves.	2,76	*fi*
43	Q13	*Mi-a fost mare **chin** înainte de a-l revedea.*	I was really troubled before seeing him again.	2,67	*fi*
44	Q5	*Mi-era aşa **pizmă** pe el!*	I envied him so much!	2,62	*fi*
45	Q63	*Ne era **sfială** ştiind căt de buni erau ceilalti concurenţi.*	We were a bit embarassed knowing how good the other participants were.	2,61	*fi*
46	Q67	*Mi-era **amărăciune** auzind vorbele lui.*	I felt miserable hearing what he said.	2,60	*fi*
47	Q30	*Îi era **sfială** înainte de a-l zări coborând din maşină.*	He was a bit coy before seeing him getting out of the car.	2,59	*fi*
48	Q65	*De obicei mi-e **tihnă** înainte de a intra în sala de examen.*	I am usually at ease before going into the exam room.	2,59	*fi*
49	Q16	*Îi era **pizmă** pe el, înainte de a afla că e aşa bogat.*	She envied him before knowing how rich he was.	2,58	*fi*
50	Q8	*Le era **râvnă** văzând atâtea bunătăţi pe masă.*	Seeing all those goodies on the table, whetted their appetite.	2,53	*fi*
51	Q44	*Îi fu **îndurare** văzând atâţia copii nenorociţi.*	He felt pity seeing so many distressed children.	2,51	*fi*
52	Q29	*Mi-era o **mânie** grozavă înainte să văd raportul de vânzări.*	I was really angry before seeing the sales report.	2,47	*fi*
53	Q39	*Mi-e **ranchiună** auzind cum îl laudă părinţii.*	Mi-e ranchiună auzind cum îl laudă părinţii.	2,47	*fi*
54	Q53	*Mi-e **mânie** pe conducerea ţării.*	Mi-e mânie pe conducerea ţării.	2,47	*fi*
55	Q32	*Mi-e **ranchiună** când îi văd împreună.*	I hate the sight of them together.	2,44	*fi*
56	Q26	*Mi-e **răceală** în suflet.*	I am cold-hearted.	2,44	*fi*

Nr.	Q_Nr	Sentence in Romanian	Sentence in English	Average Rating	Verb
57	Q57	*Are mare **tângă** de când i-a decedat părintele.*	He feels powerless since his dad died.	2,42	*avea*
58	Q45	*Mi-e **sfială** de tine.*	I embarrassed of you.	2,35	*fi*
59	Q15	*Mi-e **tângă** fără să ştiu de ce.*	I feel powerless without knowing why.	2,32	*fi*
60	Q22	*Mi-era **răceală** în suflet, auzind o istorie ca asta.*	Hearing a story like this made me cold - hearted.	2,27	*fi*
61	Q54	*Mi-e **temere** de moarte.*	I am afraid of death.	2,26	*fi*
62	Q10	*Mi-era **temere** văzând că sala era plină.*	I got nervous seeing a room full of people.	2,23	*fi*
63	Q79	*Am o **indolenţă** in oase, de nu-mi vine să fac nimic!*	I am so lazy, I don't want to do anything.	2,23	*avea*
64	Q59	*Îi era **temere** fără să aibă vreun motiv.*	He was scared for no reason.	2,22	*fi*
65	Q36	*Le fu **mânie** văzând câţi bani au pierdut.*	They got angry seeing how much money they lost.	2,19	*fi*
66	Q49	*Mi-e **zbucium** deja de-o lună.*	I have been struggling for a month now.	2,19	*fi*
67	Q41	*Îi era **tihnă** văzând că se liniştise bebe.*	She was at ease now that tha baby had cooled down.	2,17	*fi*
68	Q12	*Le era **tângă** de soarta strămoşilor lor.*	He felt sorry for his ancestors fate.	2,10	*fi*
69	Q58	*Mi-e **râvnă** de tine.*	I want you!	2,05	*fi*
70	Q6	*Ne fu **tângă** văzând atâta sărăcie în jur.*	We felt sorry seeing so much poverty around us.	2,01	*fi*
71	Q40	*Îi era **zbucium** fără a avea un motiv întemeiat.*	He was troubled without any good reason.	2,00	*fi*
72	Q31	*I-a fost **îndurare** de el când i-a născocit pedeapsa.*	He took pity on him when he made up his punishment.	1,82	*fi*
73	Q75	*Ne e **zbucium** văzând cum se comportă cu fiica noastră.*	It is unsettling seeing how he treats our daughter.	1,82	*fi*
74	Q70	*Mi-e **indolenţă** şi nu-mi vine să fac nimic.*	I am lazy and I don't want to do anything.	1,80	*fi*
75	Q74	*Îi era **indolenţă** înainte să vină prietenii lui.*	He was lazy before his friends came.	1,77	*fi*
76	Q17	*Mi-e **descurajare** când mă uit la rezultatele noastre.*	I feel disheartened when looking at our results.	1,75	*fi*
77	Q42	*Îi era **răceală** fără a înţelege de ce.*	He was cold without knowing why.	1,71	*fi*
78	Q50	*Îmi fu **indolenţă** auzind acordurile acestei melodii.*	I got lazy listening to that song.	1,63	*fi*
79	Q11	*Îi era **descurajare** fără a avea vreun motiv.*	He was disheartened with no reason.	1,42	*fi*
80	Q4	*Mi-e **descurajare** văzând cum se comportă în ultima vreme.*	I am disheartened seeing how he has been acting lately.	1,40	*fi*

Abstract

This dissertation deals with the MIHI EST construction in Romanian, illustrated in (1), in which the verb *fi* 'be' combines with a dative experiencer and a state noun. This construction represents in Romanian the most natural way of expressing psychological or physiological states. It traces back to Latin, but it disappeared from all other Romance languages, which use a HABEO structure to express this kind of states. Hence, within the Romance context the MIHI EST construction is a unique phenomenon in Romanian.

(1) *Mi-   e   foame /   sete /   frică*
    me.DAT=   is   hunger /   thirst /   fear
    'I am hungry/ thirsty/ afraid'

The present study is a part of a larger project that aims to measure Romanian's tendency to non-canonical subject marking claimed in the literature. If confirmed, this tendency contradicts the hypothesis that European languages replace non-canonical structures with canonical structures. Within this comprehensive project, my dissertation contributes with an in-depth analysis of the MIHI EST construction.

By means of a synchronic and diachronic corpus-based study, I investigate (i) the status of the core arguments of the MIHI EST structure, i.e. the dative experiencer and the nominative state noun, traditionally analyzed as the subject, and (ii) the evolution of the MIHI EST construction from the first texts in Romanian dating from the 16th century until today.

My investigation reveals that, with respect to a series of largely accepted syntactic subject criteria, the dative experiencer behaves like nominative subjects. These criteria are the following: word order, non-realization of the subject in subordinate clauses when coreferential with the subject of the main clause, movement of the subject of the subordinate clause to the position of subject of the main clause, deletion of subjects in telegraphic style, bare quantifiers in clause-initial position, and the ability to take secondary predicates. In contrast, a thorough examination of the state noun shows that, although it is nominative-marked and triggers verb agreement, it does not behave like a syntactic subject, but shows predicate behavior.

As for the evolution of the MIHI EST structure, the analysis of the data reveals that, throughout the centuries, periods of modernization alternate with periods of stabilization. With other words, periods in which new nouns are accepted in the MIHI EST structure alternate with periods in which the construction gains in stability by a more frequent usage of the same existing combinations.

Based on the presented facts, I claim that the MIHI EST construction shows a certain tendency toward expansion, since in present-day Romanian it can coerce nouns coming from other semantic fields into the construction's psychological or physiological interpretation. The question arises whether the expansion of the MIHI EST construction constitutes sufficient evidence for a propensity in Romanian toward non-canonical marking of core arguments, which would go against the tendency of the European languages toward canonical marking. Further research covering other types of predicates, such as adjectives, adverbs or verbs that occur with non-canonical subjects is required in order to validate this claim.

Samenvatting

In dit proefschrift wordt de MIHI EST-constructie in het Roemeens onderzocht. In die constructie, geïllustreerd in (1), wordt het werkwoord *fi* 'zijn' gecombineerd met een *dative experiencer* en een toestandsnomen. In het Roemeens is deze constructie de meest natuurlijke manier om een psychologische of fysiologische toestand uit te drukken. Hoewel ze overgeërfd is uit het Latijn, is de constructie verdwenen in alle andere Romaanse talen, waar een HABEO-structuur gebruikt wordt om dit soort toestanden uit te drukken. Binnen de Romaanse context is de Roemeense MIHI EST-constructie dus een uniek fenomeen.

(1) *Mi-*   *e*   *foame /*   *sete /*   *frică*
    mij.DAT=   is   honger /   dorst /   angst
    'Ik heb honger / dorst / ben bang'.

Deze studie maakt deel uit van een groter project met als doel na te gaan of er in het Roemeens een tendens is tot niet-canonieke onderwerpsmarkering, zoals beweerd in de literatuur. Indien dat het geval blijkt te zijn, is die tendens in tegenspraak met de hypothese dat Europese talen de neiging hebben niet-canonieke structuren te vervangen door canonieke structuren. Binnen dit veelomvattende project draagt mijn dissertatie bij met een diepgaande analyse van de MIHI EST-constructie.

Door middel van een synchrone en diachrone corpusgebaseerde studie onderzoek ik (i) de status van de kernargumenten in de MIHI EST-structuur, d.w.z. van de *dative experiencer* en van het nominatieve toestandsnomen, dat traditioneel geanalyseerd wordt als het subject, en (ii) de evolutie van de MIHI EST-constructie vanaf de eerste teksten in het Roemeens (daterend uit de zestiende eeuw) tot vandaag.

Uit mijn onderzoek blijkt dat de *dative experiencer* zich met betrekking tot een reeks algemeen aanvaarde syntactische criteria voor subjectstatus gedraagt zoals een nominatief subject. Die criteria zijn de volgende: woordvolgorde, niet-realisatie van het onderwerp in ondergeschikte bijzinnen indien coreferentieel met het onderwerp van de hoofdzin, verplaatsing van het onderwerp van de ondergeschikte bijzin naar de positie van onderwerp van de hoofdzin, weglating van onderwerpen in telegramstijl, kale kwantoren in zinsinitiële positie, en de mogelijkheid om secundaire predicaten te nemen. Een grondige studie van het toestandsnomen laat dan weer zien dat, hoewel het in de

nominatief staat en congrueert met het vervoegde werkwoord, het zich niet gedraagt als een syntactisch subject, maar predicaatgedrag vertoont.

Wat de evolutie van de MIHI EST-structuur betreft, blijkt uit de analyse van de gegevens dat periodes van vernieuwing worden afgewisseld met periodes van stabilisering. Met andere woorden, periodes waarin nieuwe zelfstandige naamwoorden worden geaccepteerd in de MIHI EST-structuur worden afgewisseld met periodes waarin de constructie aan stabiliteit wint doordat reeds bestaande combinaties frequenter worden gebruikt.

Op basis van de gepresenteerde data claim ik dat de MIHI EST-constructie een zekere neiging tot expansie vertoont, omdat zij in het hedendaags Roemeens de mogelijkheid biedt zelfstandige naamwoorden uit andere semantische domeinen te dwingen in de psychologische of fysiologische interpretatie van de constructie. De vraag rijst of de uitbreiding van de MIHI EST-constructie voldoende bewijs vormt dat het Roemeens een neiging vertoont tot niet-canonieke markering van kernargumenten, die zou ingaan tegen de tendens van de Europese talen tot canonieke markering. Verder onderzoek naar andere soorten predicaten, zoals bijvoeglijke naamwoorden, bijwoorden of werkwoorden die voorkomen met niet-canonieke onderwerpen, is nodig om die bewering te valideren.