



Te Reo
the Journal of the Linguistic
Society of New Zealand

Volume 62

Issue 1 (*Special Issue*): *Issue in Honour of Frantisek Lichtenberk*

Research Article

2019

pp. 130–154

September 2019

A corpus linguistic analysis of the Indirect Middle Construction in Romanian – just how peripheral is it?

Andreea S. Calude

University of Waikato, NZ

This paper is a peer-reviewed contribution from <https://www.nzlingsoc.org/journal/current-issue/>

©Te Reo – The Journal of the Linguistic Society of New Zealand

Guest Editors: Andreea S. Calude & Suzanne Kemmer

A corpus linguistic analysis of the Indirect Middle Construction in Romanian – just how peripheral is it?

Andreea S. Calude

Abstract

The indirect middle constitutes one component construction of the Middle Domain. As such, it has received comparatively little attention in the literature despite being closely related to the direct middle (and to body action events). This paper presents an empirical, quantitative analysis of the prevalence of the indirect middle voice construction in Romanian by examining a corpus of Romanian newspapers from the widely circulated newspaper *Evenimentul* (Mihalcea & Năstase, 2002). Findings from this analysis suggest that in Romanian, (1) the indirect middle is frequent and productive, (2) the indirect middle is closely associated with verbs depicting MENTAL events (rather than verbs involving a Recipient or Beneficiary), and (3) although various studies talk about “middle marked verbs”, in Romanian, the majority of verbs which occur with middle markers, can and do indeed occur without middle marking. This raises the question: to what extent are Romanian verbs which occur with middle markers strongly associated with this marking pattern? Moreover, if we assume that a strong association with middle marking is indicative of intrinsic middle semantics, can we identify connections between certain semantic verb classes and the indirect middle construction in Romanian?

Keywords

middle voice, indirect middle, Romanian, possessor raising/ascension, corpus linguistics

1 Rationale¹

Romanian is one the closest surviving daughters of Latin, and a Romance linguistic island geographically positioned among Slavic languages. In keeping with its Romance origin and like its genetic siblings, Romanian exhibits the following alternations:

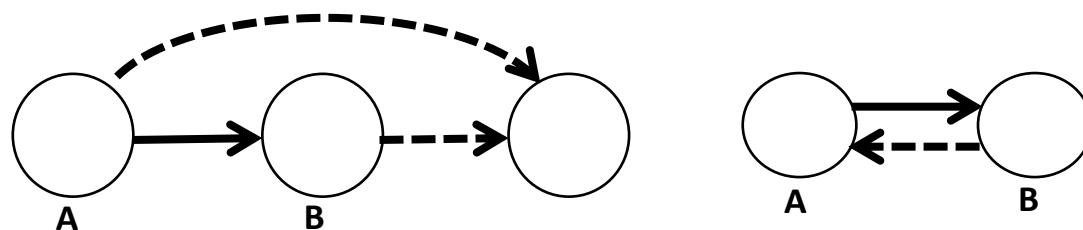
- | | | |
|--|---|-----------------------------|
| (1) | | |
| (a) <i>Clinton a cumpărat o casă pentru el însuși în New York.</i> | | INDIRECT REFLEXIVE |
| | ‘Clinton bought a house for himself in New York.’ | |
| (b) <i>Clinton și-a cumpărat o casă în New York.</i> | | INDIRECT MIDDLE |
| | ‘Clinton has bought himself a house in New York.’ | POSSESSOR RAISING/ASCENSION |
| (c) <i>Clinton a cumpărat casa lui în New York.</i> | | TRANSITIVE |
| | ‘Clinton bought his house in New York.’ | INDIRECT OBJECT LOWERING |

The sentences in (1a-c) show that in Romanian, expressing an idea like ‘Clinton has bought a house (for himself)’ can be achieved by various structural means. They each profile different component sub-parts of meaning: either the fact that the house was for himself rather than for someone else (1a), or the event of buying itself (1b), or the location of the house (1c). As the terminology given in small caps reflects, however, the constructions in (1b) and (1c) have been analysed in the previous literature under different labels, depending on the focus of the discussion, typically starting from two main perspectives, namely, either from a semantically-driven angle, or from a structurally motivated view. I discuss each perspective in turn in what follows.

Semantically oriented accounts focus on providing a unified account of the constructions which are cross-linguistically marked by middle markers, and include works by Lichtenberk (1985), Kemmer (1993), Manney (2000), Maldonado (1992), among many others. A chief goal concerning the Middle Domain in this line of research has been to pinpoint the precise motivation for common marking patterns among what seem to be rather disparate situation types (but see also a counter-position which argues explicitly for a lack of such a common motivation in Alexiadou et al., 2015). Tied to this endeavour is the issue of teasing out precise distinctions between reflexive constructions and middle constructions (be they indirect reflexive constructions and indirect middle constructions on the one hand, or direct reflexive constructions and body action/direct middles, on the other, see a discussion of this in English and Romanian in Calude, 2017). The literature in this area has focused primarily on body action constructions and direct middles specifically, while less attention has been placed on indirect middle constructions (with some few exceptions, e.g., Maldonado, 2000). The ample attention enjoyed by direct middles is most likely due to the fact that indirect middles are understood to be more loosely connected to the core of the middle voice (Kemmer, 1993, Chapter 6), rendering the indirect middle to the more peripheral parts of the Middle Domain.

Unlike direct or body action middles, the indirect middle construction involves three (not two) participants: Agent (subject), Patient (direct object) and Beneficiary or Recipient (indirect object), cf. Kemmer (1993, p. 81). The indirect middle is closely related to indirect reflexives, in that the Agent and Beneficiary/Recipient are co-referential. But it differs from these in regard to the conceptual separation between the two co-referential roles; while indirect reflexives involve more conceptual separation between the two co-referential roles, indirect middles involve comparatively less conceptual separation between them (*ibid.*).

Indirect reflexives and indirect middles can be represented schematically, as given in Figure 1 (from Kemmer, 1993, p. 77, inspired from event schemas proposed by Lichtenberk, 1985, p. 20). The figure shows, on the left-hand side, the prototypical indirect reflexive schema, in which, the three participants are represented by three circles, with participant A acting on participant B, and having an indirect effect on participant C, who is backgrounded, as indicated by the lack of a label for the third circle and a dashed line connecting it to circle A and to circle B. In the indirect middle schema (from Kemmer 1993, p. 81) given on the right-hand side, there are only two distinct participants, A and B, and these are connected to each other, A is acting on B and B on A; the dashed line shows a minimal conceptual separation between the participants involved. The very fact that a third participant is not even explicitly posited here shows the fusion of Agent and Endpoint (be it Recipient or Beneficiary).



Prototypical indirect reflexive schema

(Kemmer, 1993, p. 77)

Indirect middle schema

(Kemmer, 1993, p. 81)

Figure 1. Event schemas of the indirect reflexive and indirect middle constructions

In parallel to this work aiming to tie in semantic connections between differentially marked situation types, a whole other body of work had been set in motion, seeking to understand structures with a rather peculiar case-marking pattern: a possessor coded as an internal argument of a (non-possessive) verb and externally to the noun phrase containing the possessee (typically by a dative form), in other words, structures of the type SUBJECT + DATIVE PRONOUN + VERB + OBJECT. Testimony to the wealth of studies (too many to list them here) concerned with this phenomenon is the diversity of terms which this structure goes by in the literature: possessive raising, external possession, possessive ascension, possessive dative constructions, ethical datives, and sympathetic datives (though not all these terms are used completely synonymously). The “peculiar” marking pattern has now been recognized to be, in fact extremely widely used cross-linguistically and present in virtually all language families (Payne & Barshi, 1999). Analyses come from well-beyond Romance languages, and include, German, Hebrew, Russian, Dulong-Rawang and Bella Coola and many others:

(2)²

- (a) German (example from Lee-Schoenfeld, 2006, p. 137, ex. 54a)

Er massiert sich die Füße.

he massages self (DAT) the feet

‘He is massaging his feet.’

- (b) Hebrew (example from Berman, 1981, p. 47, ex. 18i)

Ima raxaca le Dan et ha panim

mom washed to Dan OM the face

‘Mom washed Dan’s face (for him).’ [lit. ‘Mom washed to Dan his face.’]

- (c) Russian (example from Haspelamath, 1999, p. 112, ex. 8)

Ja povetil ej nad korvat’ju kovrik.

I:NOM hung her.DAT over bed rug

‘I hung a rug over her bed.’

- (d) Dulong-Rawang (LaPolla & Jiangling, 2004, p. 7, ex. 8a)

àŋ mā.ɿ tɕiʔ-ɕu

3sg face wash-R/M

‘He is washing his face.’

- (e) Bella Coola (from Davis & Saunders, 1989, p. 135 cited in Beck, 2000, p. 236, ex. 31c)
kč=-ant-mi-cut-il
 wash=cloth-MD-REFL-1PL
 ‘We washed our own clothes.’

Note that possessor raising (PR)³ constructions can, but need not necessarily involve coreferentiality of roles (and fusion of participants). This means that not all PR constructions are also instantiations of the middle voice. Examples (2a), (2d) and (2e) involve both Possessor Raising and middle semantics, whereas (2b) and (2c) involve Possessor Raising but not middle semantics. In Romanian too, a raised possessor can be coreferential with a subject (in which case we have a middle construction, as in 3a) but need not be (in which case we are not dealing with a middle construction, as in 3b).

- (3)
- (a) *Prietena mea și-a găsit post la școală nouă.*
 friend.the my MM-has solve job at school new
 ‘My friend found a job at the new school (for herself).’
- (b) *Prietena mea mi-a găsit post la școală nouă.*
 friend.the my MM-has solve job at school new
 ‘My friend found a job at the new school for me.’

In other words, there is only a partial overlap between the study of the middle voice (and specifically indirect middles) and that of Possessor Raising (and similar phenomena). Given the focus of the present paper on the middle voice, Possessor Raising is only relevant here in so far as it also involves coreferentiality of participant roles; in other words, the analysis of Romanian data concerns examples of the sort given in (3a) but not examples like (3b).

Structurally driven analyses of raised/ascended Possessors tend to tackle one of two issues. Some studies flesh out the nature of syntactic structures pursued by minimalist/formal accounts trying to ascertain how the thematic role of the dative Possessor might be assigned, or whether in fact something which looks like a PR structure really is one (e.g. Deal, 2012; Lee-Shoenfeld, 2006; Zimmerling, 2013). Other studies probe the nature of the relationship between Possessor and Possessee, typically, the alienability of the Possessee (e.g. Shibatani, 1994; O’Connor, 1996).

An observation to be made about Romanian is that it is not strictly speaking just Possessors which can be “raised”. In Romanian, Experiencers and (non-Possessor) Recipients can also be coded by means of a raised construction, as in (4) and (5), respectively. Examples (4) and (5) “borrow” the syntax of the middle construction⁴, in other words the middle marking pattern, because semantically, the construction carries middle semantics. The participants involved in each of the events below are conceptually difficult to disentangle from one another. In example (4), the subject – the person imagining themselves – and the indirect object – the person receiving that image – are not straight-forwardly separable, and similarly, in example (5), it is difficult to separate out the people proposing to solve problems from the people receiving the benefits of the solved problems.

- (4) *Niciodată nu și-a închipuit că va muri prematur.*
 never not MM⁵-has imagine that will die prematurely
 ‘He/She₁ never imagined that they₁ would die young.’
 (source: *Evenimentul* corpus, Mihalcea & Năstase, 2002)

- (5) *Bucureștenii și-au propus să-și rezolve azi problemele.*
 Bucharest-ians MM-have decide to-MM solve today problems
 ‘The inhabitants of Bucharest decided to resolve their problems today.’
 (source: *Evenimentul* corpus, Mihalcea & Năstase, 2002)

In cases where there is indeed a relationship of possession, this relationship can emerge as a result of the very event described, what Croft terms “possessor-to-be” (1985, p. 44), as in (6), or conversely, it can be dissolved as a result of it, as in (7), something like “possessor-no-longer”.

- (6) *Un țaran din Bihor și-a făcut casă din gunoaie.*
 a peasant from Bihor MM-has made house from rubbish
 ‘A peasant from Bihor made a house for himself out of rubbish.’
 (source: *Evenimentul* corpus, Mihalcea & Năstase, 2002)
- (7) *Claudio Mendez și-a pierdut orice speranță și credință.*
 Claudio Mendez MM-has lost any hope and faith
 ‘Claudio Mendez lost all [his] hope and faith.’
 (source: *Evenimentul* corpus, Mihalcea & Năstase, 2002)

The present paper aims to complement current research on the middle voice by bringing an empirical approach to the topic of indirect middles (and PR constructions, to the extent to which they involve coreferentiality of participant roles), from Romanian, as a case-study. The need for empirical study stems from a pertinent observation made by Vihman (2002, p. 140) who argues that in Estonian, body action middle verbs are *not* often coded by means of middle marking, despite the fact that cross-linguistically, the category constitutes the archetype of the Middle Domain. This observation begs the question: what might be the archetype of the indirect middle? Additionally, an extension of this question involves looking beyond *type* occurrences of middle-marked verbs, to also counting *token* occurrences of middle-marked verbs (of the kind given in Haspelmath, 2008). Thus, my aims are twofold:

- (1) to uncover the extent to which the indirect middle is pervasive in Romanian, (which verbs are involved, how do they cluster in terms of semantic categories, how frequently are they used), and
- (2) to capture an empirical snapshot of the “middle-ness” tendencies of a number of verbs which have been found to occur in indirect middle constructions, and thus indirectly probe at the peripherality of the indirect middle as a middle category (while case-studies of certain languages provide examples of verbs which can occur in middle constructions, it is not specified whether these uses are the most frequent ones or whether non-middle uses of those verbs prevail in frequency).

2 Indirect Middle Constructions and Possessor Raising Constructions in Romanian

As already mentioned, Romanian behaves like other Romance languages, such as French (Labelle, 2008) and Spanish (Maldonado, 1992, 2002, 2008), in that the middle voice is marked by various forms of the pronoun *se*. *Se* is typically labelled in grammars of Romanian as a

“reflexive pronoun” (e.g., Avram, 1986; Bărbuță et al., 2000; Daniliuc & Daniliuc, 2000; Dindelegan, 2013; Vasilescu, 2013). This *primary reflexive strategy* (Faltz, 1985) contrasts with a *secondary reflexive strategy* which in Romanian is marked by various forms of the emphatic pronoun *însuși* (or *unul altuia* ‘each other’ in reciprocal constructions). Examples (8) and (9) show the two strategies.

- (8) *Copiii se spală.*
 children.the MM wash
 ‘The children are washing.’

- (9) *Profesorul vorbea cu el însuși.*
 teacher.the talked with he EMPHATIC
 ‘The teacher was talking to himself.’

Romanian also allows a combination of both the emphatic pronoun and the middle marker to occur in the same sentence, for example, with verbs such as, *se certa* ‘scold oneself’, *se convinge* ‘convince oneself’, *se sprijini* ‘uphold oneself’, and *se spăla* ‘wash oneself/each other’ (see Calude, 2004, 2017 for details and examples).

Numerous types of middle constructions are found in the grammar of Romanian: direct middles, indirect middles, (direct and indirect) reciprocal middles, medio-passives, and impersonal middles (cf. Cornilescu, 1998; Dobrovie-Sorin, 1998; 2006; Manoliu-Manea, 1994). Given that this paper is directly concerned with the indirect middle construction, the discussion is focused solely on this construction henceforth.

As documented elsewhere (Calude, 2004, 2007), the indirect middle construction is marked in Romanian by dative forms of the pronoun *se*, namely *își* contracted to *-și* when occurring after the infinitive form *a-* and contracted to *și-* when preceding the perfect auxiliary *a-* (for phonological reasons). The dative form occurs in indirect middles (example 10) or naturally reciprocal middles (example 11), found with both inalienable possession (example 10) and alienable possession (example 11).

- (10) *Copiii își spală mâinile înainte de masă.*
 children.the MM wash hands.the before of meal
 ‘The children are washing their hands before the meal.’

- (11) *Ei își explică unul altuia problemele.*
 they MM explain one other problems
 ‘They explain their problems to each other.’

Structurally, the two constructions in (10) and (11) involve Possessor Raising, and as hinted in the introduction, the PR construction can alternate with a non-raised version, in what Croft (1985) terms ‘indirect object lowering’, see examples (12) and (13), respectively.

- (12) *Copiii spală mâinile lor/sale înainte de masă.*
 children.the wash hands.the POSS before of meal
 ‘The children are washing their hands before the meal.’

- (13) *Ei explică unul altuia problemele lor/sale.*
 they explain one other problems POSS
 ‘They explain their problems to each other.’

As a native speaker of Romanian, my intuition parallels the claim made by Conti (2011, p. 176) for Spanish, that non-raised constructions (e.g. examples 12 and 13) are more marked, stylistically archaic or discursively unusual, compared to their raised counterparts (e.g. examples 10 and 11). More research is needed to confirm this but the lack of availability of Romanian corpora, particularly of diverse genres, continues to hamper this line of inquiry.

The sentences in (12) and (13) differ not only in construction type: (12) being a reflexive, and (13) a reciprocal construction, but also in another respect: the indirect reflexive middle in (12) can alternatively be expressed as a direct (naturally reflexive) middle, together with the accusative preposition *pe* ‘on’ (termed ‘accusative-locative’ by Manoliu-Manea, 1994, 1996), whereas the indirect reciprocal in (13) does not allow such an alternation.

- (14) *Copiii se spală pe mâini înainte de masă.*
 children.the MM wash on hands before of meal
 ‘The children wash their hands before the meal.’

- (15) **Copiii se explică problemele pentru unul altuia.*
 children.the MM wash problems for one other
 ‘The children explain their problems.’

Manoliu-Manea (1994, p. 60 and 1996, p. 720) proposes that a key difference in interpretation between (10) and (14) has to do with separation: (10) is *Part*-centred (the hands being a *part* of the Agent’s body and conceived as separated or distinct from it in some sense), whereas (14) is *Whole*-centred (here, the hands are conceived as being an integral and inseparable part of the Agent’s body). The (body) part in the dative/indirect middle from (10) is topical and salient, according to Manoliu-Manea (ibid), which explains why it is encoded by a noun phrase with a definite article, in contrast with that in the accusative-locative (which is neither topical, nor salient and is expressed with an indefinite article).

What about the difference in interpretation between (12) and (14)? The exact semantic difference between these two sentences may be aspectual in nature and it may have to do with coverage: in (12), the hands were washed all over, whereas in (14), the washing can be construed as not being fully covering the entire area of the hands.

While that explanation fits in the case of (12) and (14), coverage is not the only possible source of differences in interpretation between such constructions. Another possibility involves intensity. Consider examples (16) and (17). The adversative meaning of the indirect middle, PR construction in (16) appears to be more vivid and intense than that expressed in the non-middle/ non-raised (17), which is more neutral and less emotionally loaded. Note also that (16) contrasts with earlier examples of indirect middles, in which the third participant is both Beneficiary and Possessor; here the third participant is neither Recipient nor Beneficiary, but *only* a Possessor (of sorts).

- (16) *Ion își bate nevasta.*
 Ion MM beats wife
 ‘Ion beats his wife.’

- (17) *Ion bate nevasta lui/sa/însuși.*
 Ion beats wife POSS
 ‘Ion beats his wife.’

In general, the precise semantic differences between indirect middles, PR constructions and non-middle, non-raised counterparts remain largely elusive, for Romanian and other

languages alike, and further research is required to clarify them, from individual languages, using (ideally) naturally occurring data or experimental data.

Interestingly, while the grammar of Romanian allows Possessor Raising with both alienable and inalienable possession relationships, the alternation between indirect reflexive middles and indirect reciprocal middles is possible only in cases involving inalienable possession, as illustrated by the contrast in (18) and (19). The verb *spăla* ‘wash’ can be used in both indirect reflexive middles and indirect reciprocal middles if the Possessee is a body part, like *mâini* ‘hands’, as given in the earlier examples (8) and (10), but it cannot be used in an indirect reflexive middle if the Possessee is alienable, such as, *mașina* ‘car’, cf. (18) and (19):

- (18) *Copiii își spală mașina unul altuia.*
 children.the MM wash car.the one other
 ‘The children are washing each other’s car.’
- (19) **Copiii se spală pe mașina unul altuia.*
 children.the MM wash on car.the one other
 ‘The children wash each other’s car.’

One final observation about PR in Romanian is the possible alternation in the position of the dative pronoun with respect to the possessed noun: it can come directly after the verb and before the direct object as in (20a), or it can come after the direct object as in (20b). The interpretation of (20b) is slightly ambiguous, leaving room for a referent receiving the consequences of the event (whether benefactive or adversative) which is distinct from the Possessor of the entity encoded by the direct object, as given in (20c) – although (20c) is perhaps borderline acceptable for some speakers of Romanian. In (20c), Monica is suffering the consequences of the breaking (perhaps she was in charge of looking after it), but the radio is still possessed by Dan (in some sense). Having said this, the default reading of (20b) is still the same as that of (20a), with Dan being a Possessor, as well as Receiver, of the consequences of the event.

- (20)
- (a) *Simona i-a spart lui Dan radioul.*
 Simona HIM-has broken to.him Dan radio.the
 ‘Simona has broken Dan’s radio [to him].’
- (b) *Simona i-a spart radioul lui Dan.*
 Simona HIM-has broken radio.the POSS Dan
 ‘Simona has broken Dan’s radio.’
- (c) ?*Simona i-a spart Monicăi radioul lui Dan.*
 Simona HIM-has broken Monica.to radio.the POSS Dan
 ‘Simona has broken for [lit. on] Monica Dan’s radio.’

Romanian data differs from German and (other) Romance languages, with respect to claims made regarding the importance of affected arguments. So while in German, Lee-Schoenfeld (2006, p. 103 and others cited by her), and in Romance, Guéron (1985) and Kempchinsky (1992) assume that “a PD [possessor dative] is not only a possessor but must also be an affectee argument of the verb” (Lee-Schoenfeld, 2006, p. 103), in contrast, in Romanian, example (20b) appears to encode a Possessor which, by virtue of being overtly distinct from the affectee, is not explicitly entailed to be themselves affected (though of course, logically speaking, it may

seem reasonable to assume that they may also be indirectly affected through their possessive relationship). The Romanian data seems to be more similar in this respect with the account given for Spanish by Conti (2011, p. 177ff.). These observations illuminate the importance of naturally occurring data from individual languages in illustrating subtle cross-linguistic differences.

While indirect reciprocal middles raise their own questions, they are left for future work due to the desire for providing a detailed analysis of indirect middle constructions. Hence, the remainder of the paper focuses exclusively on these constructions. I now move on to discuss the corpus data consulted for the current analysis.

3 Data

The data reported on in this paper comes from the Romanian newspaper corpus put together by Rada Mihalcea and Vivi Năstase, comprising 50 million words of newspaper language, from the major Romanian newspaper “*Evenimentul*” (Mihalcea & Năstase, 2002). The corpus is not tagged for word-class, so all analyses were done manually, with the help of AntConc (Anthony, 2018).

Due to the large corpus size and frequent use of middle markers in Romanian, I restricted my search to the Romanian third person (singular and plural) dative form of the middle marker, *își*. As mentioned in section 2, there are two phonologically (and orthographically) distinct forms for the third person dative middle marker in Romanian, the full *își* form and the contracted *și-* or *-și* form. As already explained, the contracted form is used when preceding or following an auxiliary verb containing vowel sounds adjacent to the middle marker. One difficulty with the *și-*/*-și* form is that *și* coincides orthographically with the conjunction “and” – which is, as one might expect, highly frequent. This coincidence needs disambiguating (manually since the corpus is not tagged for word-class). In theory, the use of the dash ought to be sufficient to distinguish the two forms from one another, the middle marker and the conjunction, but its use in the corpus is not always consistent. Added to that, the use of the past perfect form appears to be very frequent in the data and given the focus of my analysis on identifying semantic classes of verbs used with the indirect middle, combining two large datasets (the *își* occurrences and the *și-* occurrences), would have been extremely time consuming. In light of these difficulties, I decided to limit my search to the *își* form alone.

So the data extracted contains 28,209 indirect middle constructions involving the third person dative (singular and plural) form *își* (and ignoring the past perfect and infinitive constructions). Note that *își* can be used in both present tense constructions (see 21) and simple past constructions (see 22), so the examples analysed are not restricted to a single tense. Furthermore, the data also contain future tense examples (which are formed with *își* and the auxiliary verb *va*, i.e., *își va cumpăra* ‘will buy (for oneself)’).

- (21) Maria *își cumpără telefon mobil.*
 Maria MM **buy** telephone portable
 ‘Maria buys (herself) a mobile phone.’

- (22) Maria *își cumpăra telefon mobil.*
 Maria MM **buy.PAST** telephone portable
 ‘Maria bought (herself) a mobile phone.’

Finally, the analysis also includes examples which contain adverbs, such as, *mai* ‘even, still’ and the modal *putea* ‘can’, see example (23).

- (23) *Maria își mai putea cumpăra telefon mobil ieri.*
 Maria MM still could buy telephone portable yesterday
 ‘Maria could still buy (herself) a new mobile phone yesterday.’

The working hypothesis is that constraining the data set in this way will not affect the results presented, because there is no reason to believe that middle constructions have any special tendency to occur with verbs of different semantic classes when used with one or the other person (1st, 2nd or 3rd person), or when used with the past perfect tense or in the infinitive form. Given the total of 28,209 indirect middle constructions identified, the figure provides some reassurance that we are hopefully dealing with a representative view of the indirect middle construction in Romanian.

4 Results

This section provides the results of the corpus analysis described above. Before proceeding to the frequency of use results uncovered, a short discussion of the verb types found in Romanian is in order.

4.1 Romanian verbs: four types of verbs which occur in indirect middles

Verbs which allow multiple arguments in Romanian occur in the following construction types:

- (A) Ditransitive forms with three distinct participants, an Agent, a Patient and a coreferential Receiver/Beneficiary/Possessor,
- (B) Indirect middle constructions with the dative PR (*își* middle marker),
- (C) Indirect reflexive constructions with emphatic *însuși* reflexive pronoun form,
- (D) Indirect emphatic middle constructions with both *își* middle marker and *însuși* pronoun form.

Some verbs occur in all of the above construction types. For example, verbs like *pedepsi* ‘to punish’. The verb *pedepsi* can be used transitively with distinct Agent and Patient participants (as in 24a); in a middle construction where the Possessor and Receiver role are co-referential (as in 24b); in an emphatic construction where the Possessor role is marked by the emphatic pronoun *însuși* ‘himself’ (see 24c), or in a situation coded by both the middle marker and the emphatic pronoun (see 24d).

- (24)
- (a) *El îi pedepsește pe copiii care nu ascultă.*
 he ACC punishes on children.the which not listen
 ‘He punishes the children which do not listen.’
 - (b) *El își pedepsește copiii.*
 he MM punishes children.the
 ‘He punishes his children.’

- (c) *El ar pedepsi copiii lui însuși pentru așa ceva.*
 he would **punish** children.the POSS **himself** for this something
 ‘He would punish his own children for something like this.’
- (d) *El își pedepsea copiii lui însuși pentru așa ceva.*
 he MM **punished** children.the POSS **himself** for this something
 ‘He punished his own children for something like this.’

However, not all verbs can be used in all of the above construction types. Such verbs fall into the following two types: (1) verbs which can occur in A, B, D but not C, as in (25), (2) verbs which can occur in A and B but not C or D, as in (26).

(25)

- (a) *El uită cheia de la casă regulat.*
 he **forgets** key.the of to house regularly
 ‘He regularly forgets the house key.’
- (b) *El își uită cheia de la casă regulat.*
 he MM **forgets** key.the of to house regularly
 ‘He regularly forgets his house key.’
- (c) **El își uită cheia de la casă lui însuși.*
 he MM **forgets** key.the of house him **himself**
- (d) *El uită cheia sa/ lui însuși în ziua interviului.*
 he **forgets** key.the his.POSS him **himself** in day.the interview.the
 ‘He forgets his key [on himself] on the day of the interview.’

(26)

- (a) *Adina petrece vacanța de vară la bunici.*
 Adina **spends** holiday.the of summer at grandparents
 ‘Adina spends her summer holiday at her grandparents.’
- (b) *Adina își petrece vacanța de vară la bunici.*
 Adina MM **spends** holiday.the of summer at grandparents
 ‘Adina spends her summer holiday at her grandparents.’
- (c) **Adina își petrece vacanța ei însăși la bunici.*
 Adina MM **spends** holiday.the POSS **herself** at grandparents
 ‘Adina spends her summer holiday at her grandparents.’
- (d) **Adina petrece vacanța ei însăși la bunici.*
 Adina **spends** holiday.the POSS **herself** at grandparents
 ‘Adina spends her summer holiday at her grandparents.’

Note that in Romanian, if a verb occurs in B, it will necessarily also be able to occur in A⁶. This contrasts with say a language like English, where some verbs are shown to only occur in a (necessarily) co-referential situation type, *She absented herself from work*, *He prides himself in his spaghetti* (examples 10a, b, from Siemund, 2000, p. 801) – though of course, here we are talking about reflexives and not indirect middles, but the point is that core-

referentiality of roles is built-in to the verb's valency patterns (**She absented her brother from work*, **He prides in his spaghetti*).

Finally, some verbs can occur in A and B but not C or D, but their meanings differ in A and B (in contrast to the verb in 26, whose meaning stays the same in A and B), see examples in (27). The meaning of the ditransitive construction is distinct from, but related to, that of the indirect middle construction (what Geniušienė terms “non-reversible reflexives of the semantic type” which can be both, of the figurative and non-figurative extension kind, cf. 1987, p. 145–152). Example (27a) gives a non-reversible indirect middle of the figurative nature involving a metaphorical extension, from *aminti* meaning “to remind (someone of something)” in the ditransitive sentence in (27b) to *își aminti* “to remember (something for oneself)” in the middle construction of (27a).

(27)

(a) *El își amintește cu plăcere de anii de școală.*
 he MM **remembers** with pleasure of years.the of school
 ‘He remembers fondly his school years.’

(b) *El îi amintește Mariei de anii de școală.*
 he DAT **reminds** Mary.DAT of years.the of school
 ‘He reminds Maria of (?their/her) school years.’

To sum up, in Romanian, verbs can be classified into the following classes, with regard to indirect middle and reflexive marking patterns:

Table 1. Verb types within ditransitive construction alternations in Romanian

Verb types	(A)	(B)	(C)	(D)
	ditransitive	indirect middle /PR raising	indirect emphatic reflexive	indirect reflexive
	clitic dative pronoun	<i>își</i>	<i>își</i> middle marker + reflexive pronoun <i>însuși</i>	reflexive pronoun <i>însuși</i>
	3 distinct roles	(Expected) coreferential Agent and Possessor/ Beneficiary	Coreferential Agent and Possessor/ Beneficiary	(Unexpected) Coreferential Agent and Possessor/ Beneficiary
(1) <i>pedepsi</i> ‘punish’	X	X	X	X
(2) <i>uita</i> ‘forget’	X	X	X	
(3) <i>petrece</i> ‘spend’	X	X		
(4) <i>aminti</i> ‘remind, remember’	X	X (but with different semantics in A and B)		

4.2 Romanian verbs: frequency of use

Table 2 gives the 100 most frequently occurring verbs with the middle *își* form (in this written newspaper corpus). The verbs appear in rank order, from most frequent to least frequent. For each verb, the table provides the Romanian form, the verb's English translation for each available constructional use (type A, B or C), its semantic category (following Levin, 1993) and total raw frequency of use in the middle/PR construction.

Assigning verbs to a particular semantic category was not always straight-forward and a few observations about how this was done are in order. First, the category “generic” refers to verbs which are semantically versatile, for example, *face* ‘do/make’ can occur with a variety of objects which pertain to different semantic domains; including personal benefit *își face parte* ‘to give oneself more’, motion *își face bagajele* ‘to leave’ (literally: to pack one’s bags), and emotion *își face curaj* ‘to be brave’ (literally: to make oneself courage) among others. Secondly, the semantic category was chosen on the basis of alignment with examples from Levin (1993) for English, but also by taking into account the Possessor objects that the various verbs occurred with in the Romanian corpus of newspapers (as detailed for instance above for the verb *face*).

Table 2. The 100 most frequently occurring verbs in the Romanian indirect middle

NO	VERB	as A (ditransitive)	as B (indirect middle)	as C (same meaning in both)	SEMANTIC CATEGORY	RAW FREQ.
1	<i>face</i>			do/make	GENERIC	1959
2	<i>aminti</i>	remind (X of Y)	remember X		MENTAL	1062
3	<i>da</i>			give	GIVE-VERBS	927
4	<i>dori</i>			want/desire	MENTAL	799
5	<i>lua</i>			take	GIVE-VERBS	612
6	<i>pune</i>			put	MOTION	537
7	<i>petrece</i>	cross OR party	spend		GENERIC	496
8	<i>propune</i>	propose	resolve to do something		COMMUNICATION	480
9	<i>pierde</i>			lose	MENTAL	477
10	<i>permite</i>	allow, permit	afford		EMOTION	433
11	<i>desfășura</i>			unfold	ASPECTUAL	428
12	<i>aduce</i>			bring	BENEFACTIVE	418
13	<i>asuma</i>	believes/ assumes	assume (the role of)		MENTAL	414
14	<i>găsi</i>			find	BENEFACTIVE	373
15	<i>continua</i>			continue	ASPECTUAL	359
16	<i>exprima</i>			express	COMMUNICATION	349
17	<i>vedea</i>	see	imagine		MENTAL	332
18	<i>avea</i>			have	GENERIC	324
19	<i>schimba</i>			change	GENERIC	304
20	<i>păstra</i>			keep	ASPECTUAL	301
21	<i>cumpăra</i>			buy	GET-VERBS	283
22	<i>vinde</i>			sell	DATIVE-VERBS	269

23	<i>căuta</i>			search	BENEFACTIVE	266
24	<i>bate</i>			beat/hit	CONFLICT	259
25	<i>plăti</i>			pay	GIVE-VERBS	239
26	<i>duce</i>			take	MOTION	233
27	<i>primi</i>			receive	DATIVE-VERBS	206
28	<i>începe</i>			start/begin	ASPECTUAL	199
29	<i>alege</i>			choose	GET-VERBS	193
30	<i>cere</i>		(idiomatic)	ask	COMMUNICATION	193
31	<i>menține</i>	also hold		keep up	ASPECTUAL	182
32	<i>achita</i>			acquit	GIVE-VERBS	177
33	<i>ține</i>			hold (on to)	MOTION	170
34	<i>spune</i>			say	COMMUNICATION	161
35	<i>deschide</i>			open	MOTION	158
36	<i>disputa</i>			contest	CONFLICT	152
37	<i>aștepta</i>			wait	MENTAL	151
38	<i>scoate</i>	take	uncover		MOTION	151
39	<i>prezenta</i>			introduce/present	COMMUNICATION	150
40	<i>anunța</i>			announce/report	COMMUNICATION	147
41	<i>trimite</i>			send	GIVE-VERBS	146
42	<i>lansa</i>			launch	MENTAL/MOTION	145
43	<i>recupera</i>			get back/salvage	GET-VERBS	140
44	<i>arăta</i>			show	MENTAL	138
45	<i>apăra</i>			defend	CONFLICT	136
46	<i>construi</i>			build	BENEFACTIVE	135
47	<i>imagina</i>			conceive/fancy/imagine/figure out	MENTAL	133
48	<i>pregăti</i>			prepare	BENEFACTIVE	130
49	<i>reveni</i>			come back / revive	MOTION(BODY)	128
50	<i>încheia</i>			close	MOTION	120
51	<i>lăsa</i>			let / allow	GENERIC	116
52	<i>retrage</i>			retract/recall/relinquish	CONTRIBUTE-VERBS	110
53	<i>ridica</i>			rise/ come up	MOTION(BODY)	102
54	<i>depune</i>			deposit / invest	CONTRIBUTE-VERBS	98
55	<i>câștiga</i>			win	GET-VERBS	96
56	<i>rezerva</i>			reserve	GET-VERBS	93
57	<i>împărți</i>			dispense/give out/split	GIVE-VERBS	91
58	<i>îndeplini</i>	carry out	honour/deliver		MENTAL/MOTION	90
59	<i>oferi</i>			offer/ volunteer	GIVE-VERBS	86
60	<i>asigura</i>			assure/secure	MENTAL	83
61	<i>iubi</i>			love	EMOTION	79

62	<i>manifesta</i>			manifest/demonstrate/display	MENTAL	79
63	<i>trage</i>			pull	MOTION(BODY)	78
64	<i>folosi</i>			use	GENERIC	77
65	<i>susține</i>			favour/support/uphold/bolster/bear up	MENTAL	76
66	<i>băga</i>	enter/insert/intrude	enter/insert/intrude but has a rather negative connotation		MOTION	76
67	<i>expune</i>			bare/exhibit/expose/flaunt	COMMUNICATION	75
68	<i>rezolva</i>			resolve	MENTAL	73
69	<i>relua</i>			continue/reinstate/resume	ASPECTUAL	72
70	<i>explica</i>			account for/explain/	COMMUNICATION	70
71	<i>muta</i>			move	MOTION	70
72	<i>respecta</i>			respect	MENTAL	69
73	<i>ascunde</i>			hide	MOTION	67
74	<i>sărbători</i>			celebrate	EMOTION	66
75	<i>juca</i>			play	GENERIC	66
76	<i>încerca</i>			try	MENTAL	63
77	<i>exercita</i>			carry out/practice/exercise	MENTAL/MOTION	63
78	<i>satisface</i>			satisfy	MENTAL/ BENEFACTIVE	62
79	<i>acuza</i>			charge /accuse	CONFLICT	61
80	<i>închipui</i>			depict/envision/fancy/make up	MENTAL	60
81	<i>urma</i>			follow	MOTION	60
82	<i>extinde</i>			extend	MENTAL	60
83	<i>recunoaște</i>			recognize (person) but also admit	MENTAL	60
84	<i>închide</i>			close/imprison/enclose	MOTION	59
85	<i>procura</i>			obtain/gain/procure	GET-VERBS	59
86	<i>termina</i>			finish /end	ASPECTUAL	56
87	<i>ispăși</i>	satisfy/atone	satisfy/atone, but with negative connotation		MENTAL/MOTION	54
88	<i>ajuta</i>			help	BENEFACTIVE	53
89	<i>trece</i>			pass	MOTION	53
90	<i>arunca</i>			throw	MOTION	52
91	<i>declara</i>			declare	COMMUNICATION	51
92	<i>revendica</i>			insist on/ make a claim for	MENTAL	51

93	<i>modifica</i>			modify	MENTAL/MOTION	51
94	<i>conduce</i>			drive/handle/ma nage	MOTION	51
95	<i>spăla</i>			wash	MOTION(BODY)	50
96	<i>părăsi</i>			leave/ditch	DATIVE-VERBS	50
97	<i>recăpăta</i>			recover / regain	GET-VERBS	49
98	<i>uni</i>			unite	MOTION/MENTAL	49
99	<i>merita</i>			merit	MENTAL	48
100	<i>crea</i>			create	MENTAL	47
TOTALS					20,104 (71% of all verbs found to occur in the indirect middle)	

Like other linguistic phenomena, the frequency distribution of the semantic classes of the verbs occurring in the indirect middle is skewed, with some relatively few semantic classes accounting for most of the data. Among the verbs found to occur in the indirect middle construction (and the indirect emphatic middle construction), the most commonly occurring ones are MENTAL verbs, such as *dori* ‘wish, desire’, *asuma* ‘assume’, *astepta* ‘wait’, *lansa* ‘launch’ and *recunoaște* ‘recognize’, with respect to both type and token counts, see Figures 2 and 3 below (summarised directly from the counts given in Table 1).

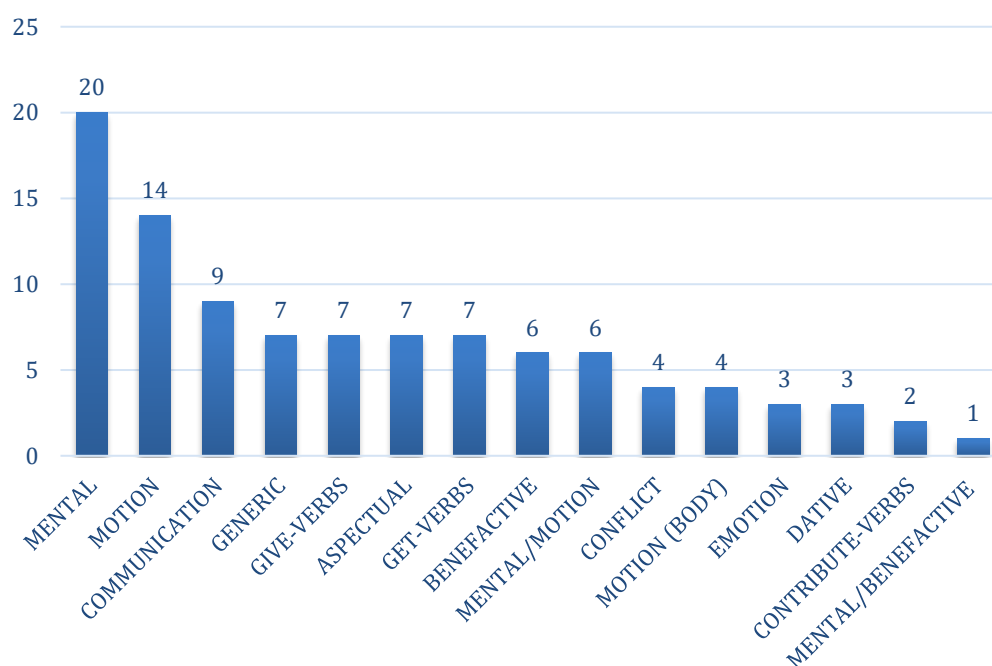


Figure 2. Type frequency of 100 most frequent verbs occurring in the Romanian indirect middle construction (in the newspaper corpus)

Assuming that previous findings are correct about the direct middle’s archetype being body action events, and assuming that the indirect middle is parallel to the direct middle, this finding seems surprising. In the Romanian corpus of newspapers consulted, body motion events figure low in terms of both types and tokens in the indirect middle construction. It remains to be seen whether this trend applies to Romanian middles more generally, that is, whether direct middles are also not commonly associated with body action events, as reported for Estonian (Vihman

2002, p. 140), or whether this is specific to Romanian indirect middles and not applicable to direct middles (or whether this trend is applicable to indirect middles cross-linguistically)⁷.

Similarly, GET-verbs, DATIVE verbs and BENEFACTIVE verbs are not especially frequent either. Instead, GIVE-verbs, MOTION verbs and COMMUNICATION verbs are the next most commonly found in the indirect middle construction, following the widespread use of MENTAL verbs.

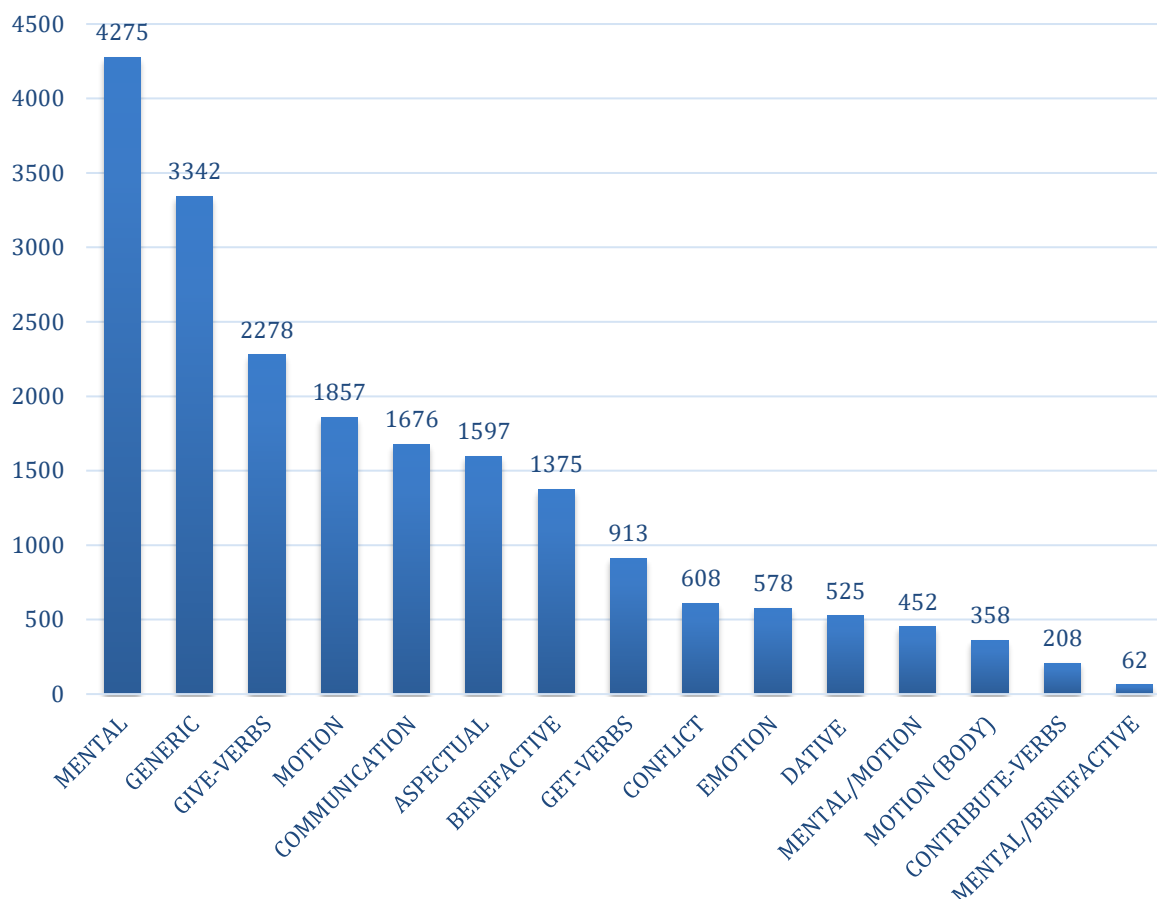


Figure 3. Token frequency of 100 most frequent verbs in the Romanian indirect middle construction (in the newspaper corpus)

The frequency patterns above suggest that the indirect middle in Romanian has a special affinity with the **cognition middle** (Kemmer 1993, p. 127–142). The cognition middle category has its basis in a semantic underpinning as it concerns events which pertain to cognition and which can straddle both indirect and direct middles. Many of the verbs classified here as belonging to the “mental domain” could be grouped within the broader range of the cognition middle, without stipulating a direct or indirect middle construction. This grouping may be further supported by the fact that many verbs which occur in the indirect middle construction can also occur in a direct middle construction, and in fact, in other middle construction types also, as illustrated in (28).

While it may be tempting to assign the cognition middle as the prototype of the indirect middle construction in Romanian, this would be misguided because, as discussed above, the cognition middle is not a subcategory of the indirect middle; it is encoded by a wider range of constructions than just indirect middles⁸. But the close association between the semantics of the cognition middle and the structural pattern stipulated by the indirect middle construction

creates a strong link between the two. Further to this point, the widespread use of the cognition middle in Romanian is by no means exceptional; it stands alongside similar claims made for Romanian indirect middle constructions on the basis of a smaller corpus of fairy tales and newspaper articles by Hartenstein (2004), and by similar claims made for the Greek middle voice more generally (cf. Manney, 2000, p. 42).

(28)

- | | | |
|-----|---|-----------------|
| (a) | <i>Băiatul acela și-a pierdut complet mințile.</i> | INDIRECT MIDDLE |
| | boy.the that MM lost completely minds.the | |
| | ‘That boy has completely lost his mind.’ | |
| (b) | <i>Ea s-a pierdut repede în orașul nou.</i> | DIRECT MIDDLE |
| | she MM lost completely in town.the new | |
| | ‘She got lost quickly in the new town.’ | |
| (c) | <i>Cheile acestea se pot pierde ușor.</i> | MEDIO-PASSIVE |
| | keys.the these MM can lose easily | |
| | ‘These keys can easily get lost.’ | |
| (d) | <i>Se pierde ușor ceea ce se câștigă repede.</i> | IMPERSONAL |
| | MM lose easily that what MM win quickly | |
| | ‘That which is earned easily is (then) lost quickly.’ | |

4.3 How ‘middle’ can a Romanian verb be?

Table 1. shows that some Romanian verbs are used more frequently in the indirect middle construction compared to other verbs. While this is meaningful in one sense, it is also ambiguous in another sense, namely, in that the verbs presented in the table do not occur exclusively as middles (and in some cases, not even exclusively as indirect middles, but can occur as other types of middles, as exemplified in 28). An immediate question follows: if a verb is used X number of times in an indirect middle construction, can that use be a mere effect of the frequency of occurrence of that particular verb in general? Put another way, is the low frequency of occurrence of Romanian *crea* ‘create’ in indirect middles simply an effect of overall low frequency-of-use of the verb in the corpus?

In order to shed light on the tendencies of the verbs which were most frequently found to occur in indirect middle constructions, ten verbs were investigated in further detail, by asking the question: for each verb, how does the frequency of its uses in a middle construction compare with its frequency in a non-middle construction? This line of inquiry follows a collostructional approach as outlined in Stefanowitsch & Gries (2003).

Due to the size of the corpus, it was not possible to manually code the verbs in question exhaustively (for a verb like *face* ‘do/make’, that would involve manually sifting through more than 20,000 uses alone). Instead, a random selection of 100 uses were extracted from the corpus (that is, any use of the verb, not just limited to the third person uses, nor to middles, obviously), and coded for the type of construction in which it occurred: non-middle or middle. Furthermore, for each middle construction, the type of middle construction was also recorded: direct middle, indirect middle, reciprocal middle, impersonal middle and medio-passive (following the broad categories outlined by Kemmer, 1993).

I chose to investigate verbs at both ends of the frequency spectrum; five verbs from the most frequently occurring verbs in the indirect middles construction, and five verbs from the least frequently occurring verbs in the indirect middle construction⁹ (the table only gives the 100 most frequently occurring verbs for each of the ten verbs, so there will be other verbs not included here that occurred fewer times still). The results of the manual coding are shown in Table 3 and in Figure 5. The table gives the highest middle category frequency of occurrence of each verb in bold font-face.

Table 3. Manually coding of 100 random uses of each verb below

VERB	non-middle	indirect middle	direct middle	reciprocal middle	medio-passive middle	impersonal middle
<i>face</i> ‘do/make’	72	9	5	-	1	13
<i>aminti</i> ‘remember /remind’	31	69	-	-	-	-
<i>dori</i> ‘wish/desire’	67	28	-	-	1	4
<i>lua</i> ‘take’	83	11	1	2	-	3
<i>pune</i> ‘put’	59	12	3	-	2	24
<i>spăla</i> ‘wash’	56	9	27	-	1	7
<i>părăsi</i> ‘leave / ditch’	92	8	-	-	-	-
<i>recăpăta</i> ‘recover/ regain’	13	87	-	-	-	-
<i>merita</i> ‘merit’	91	6	-	-	-	3
<i>crea</i> ‘create’	84	5	-	-	-	11

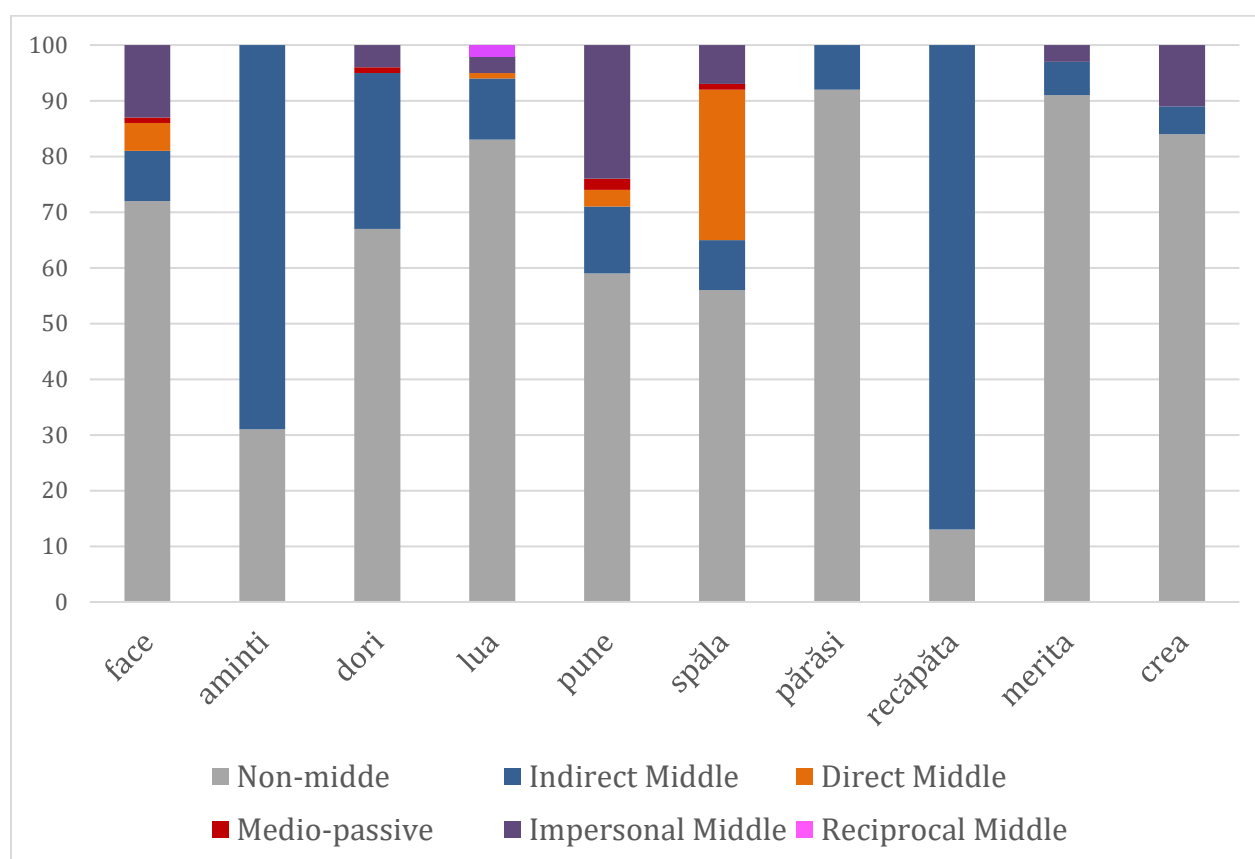


Figure 5. Graph of middle tendencies of 100 random verb-uses for ten Romanian verbs (*face*, *aminti*, *dori*, *lua*, *pune*, *spăla*, *părăsi*, *recăpăta*, *merita*, *crea*)

The results indicate that certain verbs which are among the most frequently found with indirect middle marking (*își* form) are generally more often found in non-middle constructions than in middle-marked constructions, for instance *face* ‘do/make’ or *lua* ‘take’. The converse is also true: at least one of the verbs which is least frequently found with indirect middle marking occurs more often in this type of construction than in a non-middle construction, namely, *recăpăta* ‘recover/regain’. Put another way, some infrequently occurring verbs can have strong associations with the Middle Domain (*își recăpăta*), while those that perhaps might spring to mind more often, such as, *își lua* ‘to take for oneself’ only do so because of their overall widespread use, and not necessarily because of their strong association with the middle. Only two verbs among the ten investigated show a distinct preference of occurrence with middle marking, namely *aminti* ‘remember, remind’ and *recăpăta* ‘recover/ regain’.

The table also shows that verbs tend to fall into two types: (1) they are either used in many different types of middle constructions (among 100 examples, these types of verbs will occur at least once in almost all middle construction types available, not just the indirect middle, but also direct middle, medio-passive or impersonal middle), or else (2) they ‘specialise’ in a particular middle construction type and have special affinity to that structure (in our case, this happens to be the indirect middle because this is the structure that was searched for in the first place). The least ‘versatile’ middle construction type appears to be the reciprocal middle, with only two uses of the verb *lua* ‘take’ occurring in a reciprocal middle (from the 100 examples checked).

Finally, the frequency counts suggest that many verbs can occur in various middle construction types, and specifically, that the impersonal middle is often one of these (only 3 from the 10 verbs occur in just the indirect middle construction type, possibly not coincidentally, 2 of these verbs have a special general association with the indirect middle, namely, *recăpăta* ‘recover/regain’ and *aminti* ‘remember/remind’).

The approach taken here is similar to Haspelmath (2008). Haspelmath compares the frequency of introverted verbs and extroverted verbs in corpora of German, Czech and English and suggests that there is a universal tendency for “verbs with higher frequency of reflexive use [to] show shorter reflexive-marking forms than verbs with lower frequency of reflexive use” (2008, p. 46–47). However, this universal is specifically formulated with regard to *pronoun* forms (reflexive versus non-reflexive phoric pronouns, ignoring full noun phrases). For example, he argues that the German verb *waschen* “to wash” occurs more frequently in the reflexive form “*sich waschen*” than as “*waschen* + pronoun other than the subject” (=disjoint pronoun), and similarly in English, “*wash* + reflexive *-self*” occurs more often than “*wash* + disjoint pronoun”. The reason for looking at pronoun use and not taking into account full noun phrases, according to Haspelmath is that the reflexive forms have as “direct competitors” those forms which are of the same type, that is, other pronouns (2008, p. 47).

While the analysis presented here is inspired from Haspelmath’s approach, it also differs from it: this analysis does not exclusively focus on pronoun use (though, for a direct comparison, I did check the objects of the verb *spăla* ‘wash’ and only seven cases from 100 were non-reflexive phoric pronouns, in agreement with Haspelmath’s findings for German and English). Because I was interested to compare the general occurrence of middle versus non-middle marking with the verbs in question (and direct and indirect middles are only a subset of the possible middle types found in Romanian), I took all uses of each verb into account, including intransitive and (periphrastic) passive uses.

5 Conclusion

This paper set out to investigate the Romanian indirect middle voice construction – a construction which has, to date, received comparatively less attention than other middle voice constructions, possibly due to its alleged peripheral status within the Middle Domain.

Being the second study to report actual frequency of use of such middle voice phenomena (following Hartenstein, 2004), this quantitative analysis puts forward four main findings:

- (1) The indirect middle voice construction is not as peripheral as first thought, at least, as far as Romanian is concerned; its use seems productive both in terms of verb types and verb tokens (exceeding 20,000 examples in an albeit rather large corpus).
- (2) The data has reveals an affinity between the structural marking patterns of the indirect middle and the semantic category of the COGNITION middle.
- (3) Clearly, not every Romanian verb occurs with middle marking, but for those that do, there will always be an available non-middle marked counterpart use available (even though in some cases, the meanings change slightly across the two marking patterns).
- (4) A collostructional analysis shows that controlling for overall frequencies of occurrence of verbs is important, and using this approach uncovers the fact that Romanian verbs fall into one of two possible categories: (a) verbs which occur in at least one middle construction type, and if that construction is an indirect middle (applicable to all the verbs investigated here), the next most commonly found possibility is that of an impersonal middle construction, and (b) verbs which occur predominantly in non-middle marked constructions (the most common type).

The paper also raises a number of questions, such as the need for further investigation of the relationship between the types of participants involved in indirect middle constructions, specifically not just the types of direct objects that can occur in indirect middle constructions, but also, the types of objects that *actually* occur regularly in these (are they of the alienable or inalienable kind, are they thought to be in a close relationship of some kind to the Agent subjects?).

A second question that requires further examination is the link between indirect middle constructions (and perhaps middle constructions in general) and aspectual properties of these. Some of the examples given in this paper from Romanian, namely (12) and (14) appear – at first glance at least – to differ in their aspectual properties, but this remains unclear for now.

In writing this paper, my aim is to complement existing studies of the Romanian middle voice, specifically by bringing large-scale, empirical data of middle voice constructions in this language. Secondly, the paper aims to contribute to wider work concerned with middle voice systems in general, by showing the benefits of looking deep inside patterns of use within individual languages in order to probe connections between the semantics of given verbs and their affinity to certain constructions types.

Acknowledgement

I am extremely grateful to the anonymous referee of this paper and to Suzanne Kemmer for her meticulous comments and for the time set aside to discuss this work over Skype. Any errors remaining are of course, my own. Finally, as explained in endnote 1, I am greatly indebted to Frank Lichtenberk for introducing me to the wonderful topic of the middle voice.

Notes

1. Frank Lichtenberk taught me from undergraduate study (in the typical LING101 course, in 1998), right through to my Honours year in Linguistics. At that time, I wanted to enrol in a generative/formal course in syntax, but as luck would have it, Frank was teaching graduate syntax that year, and it had to be cognitive/functional rather than generative/formal syntax. To top it all off, I was slightly scared of Frank (he once sent me away from his office for asking questions prematurely about an assignment). Seeing as I had hoped to write a thesis in an area of syntax for my Masters, things were not looking good for my future in linguistics. However, over the course of those twelve weeks, Frank changed my entire focus in syntax. He talked about such things as alienable and inalienable possession in the world's languages and about ergativity and unaccusatives. Those two topics reshaped and rewired my linguistics interests. What I had devoured during my three-year undergraduate study (Chomskyan trees and various Principles) was completely displaced in the space of about four hours of graduate lectures. To my complete surprise, I became hooked on cognitive/functional grammar. By the time I had mustered up the courage to ask Frank whether he would supervise me for an MA project on some aspect of Romanian syntax, he carefully took out from his bookcase, a thick-looking book with a bright orange cover on which three words stood out in bold black font "The Middle Voice". "Go home and read that and then come back to see me", he said matter-of-factly. Life has not been the same ever since. For that, I am eternally glad.

2. The bolding and italics in these examples are my own, abbreviations used in these example are from the original sources.

3. I am using this term as a cover for all phenomena given in examples (2).

4. I am grateful to Suzanne Kemmer for making me aware of this connection.

5. The abbreviations used in this paper were deliberately kept to a minimum in order to not distract the reader, they include: MM – middle marker, POSS – possessive marker, EMPHATIC – emphatic pronoun, DAT – dative marker.

6. The only exception (from the verbs in Table 1.) is the verb *își închipui* 'imagine, dream up' which – to the best of my knowledge – can (mostly) only occur with the middle marker.

7. One anonymous reviewer suggests that this effect could also be due to the genre of the corpus (newspaper language), and this is, of course, a possibility. I am not entirely sure why newspaper language is more likely to concern mental processes than other events, but this plausibility cannot be ruled out until further data is gathered.

8. I am grateful to Suzanne Kemmer for pointing this out and for the stimulating discussion surrounding this issue.

9. Two verbs were not coded, namely *uni* 'unite' and *da* 'give' because of their high polysemy with other Romanian lexical forms.

References

- Alexiadou, Artemis, Anagnostopoulou, Elena, & Schafer, Florian. (2015). *External Arguments in Transitivity Alternations*. Oxford: Oxford University Press.
- Anthony, Laurence. (2018). AntConc (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University. (<http://www.laurenceanthony.net/software>).
- Avram, Mioara. (1986). *Grammatica Pentru Toți*. București: Editura Academiei Republicii Socialiste România.
- Bărbuță, Ion, Cicală, Armenia, Constantinovici, Elena, & Cotelnic, Teodoro. (Eds.) (2000). *Grammatica Uzuală a Limbii Române*. Chișinău: Grupul Editorial Litera.
- Beck, David. (2000). Unitariness of Participant and Event in the Bella Coola (Nuxalk) Middle Voice. *International Journal of American Linguistics*, 66(2), 218–256.
- Berman, Ruth. (1981). Dative Marking of the Affectee Role: Data from Modern Hebrew. *MIT Working Papers in Linguistics*, 3, 35–59.
- Calude, Andreea S. (2004). Reflexive–middle and reciprocal–middle continua in Romanian. *Proceedings of the 2004 Conference of the Australian Linguistic Society*. <http://ses.library.usyd.edu.au/handle/2123/93> (accessed 3 April 2017).
- Calude, Andreea S. (2007). Light and heavy reflexive marking: The middle domain in Romanian. *Annual Review of Cognitive Linguistics*, 5, 239–269.
- Calude, Andreea S. (2017). Testing the boundaries of the middle voice: Observations from English and Romanian. *Cognitive Linguistics*, 28(4), 599–630.
- Cornilescu, Alexandra. (1998). Remarks on the syntax and the interpretation of Romanian middle passive *se* constructions. *Revue Roumaine de Linguistique*, 43, 317–342.
- Conti, Carmen. (2011). Possessive Dative revisited: Another view of external possession in Spanish. *Studia Linguistica*, 65(2), 170–197.
- Croft, Bill. (1985). Indirect Object “Lowering”. *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society*, 11, 39–51.
- Daniliuc, Laura, & Daniliuc, Radu. (2000). *Descriptive Romanian Grammar*. LINCOM Studies in Romance Linguistics 14. Munich: LINCOM Europa.
- Davis, Phillip W., & Saunders, Ross. (1973). Lexical suffix copying in Bella Coola. *Glossa*, 7, 231–251.
- Deal, Amy Rose. (2013). Possessor Raising. *Linguistic Inquiry*, 44(3), 391–432.
- Dindelegan, Pană Gabriela. (Ed.) (2013). *The Grammar of Romanian*. Oxford: Oxford University Press.
- Dobrovie-Sorin, C. (1998). Impersonal *se* constructions in Romance and the passivization of unergatives. *Linguistic Inquiry*, 29, 399–437.
- Dobrovie-Sorin, C. (2006). The SE-anaphor and its role in argument realisation. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell Companion to Syntax* (pp. 118–179). Malden: Blackwell.
- Faltz, Leonard. (1985). *Reflexivization: A Study in Universal Syntax*. New York: Garland.
- Geniušienė, Emma. (1987). *The Typology of Reflexives*. Berlin: Mouton de Gruyter.
- Guéron, Jacqueline. (1985). Inalienable possession, PRO-inclusion and lexical chains. In Hans-Georg Obenauer, Jean-Yves Pollock, & Jacqueline Guéron (Eds.), *Grammatical representation* (pp. 43–86). Dordrecht: Foris.
- Hartenstein, Anne-Marie. (2004). The Middle Voice in Romanian. A corpus based analysis, *Kobe Papers in Linguistics*, 4, 20–40.
- Haspelmath, Martin. (2008). A frequentist explanation of some universals of reflexive marking. *Linguistic Discovery*, 6(1), 40–63.

- Haspelmath, Martin. (1999). External Possession in a European Areal Perspective. In Doris L. Payne and Immanuel Barshi (Eds.), *External Possession* (pp. 109–135). Amsterdam / Philadelphia: John Benjamins.
- Kempchinsky, Paula. (1992). The Spanish possessive dative construction: θ -role assignments and proper government. In P. Hirschbuhler (Ed.), *Linguistic Symposium on Romance Languages*, Vol 20 (pp. 135–149).
- Kemmer, Suzanne. (1993). *The middle voice*. Amsterdam: John Benjamins.
- Labelle, Marie. (2008). The French Reflexive and Reciprocal *se*. *Natural Language & Linguistic Theory*, 26(4), 833–876.
- LaPolla, Randy, & Jiangling, Yang. (2004). Reflexive and middle marking in Dulong-Rawang. *Himalayan Linguistics*, 2, 1–14.
- Lee-Schoenfeld, Vera. (2006). German Possessor Datives: Raised and Affected. *The Journal of Comparative Germanic Linguistics*, 9(2), 101–142.
- Levin, Beth. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago press.
- Lichtenberk, Frantisek. (1985). Multiple Uses of Reciprocal Constructions. *Australian Linguistics Journal*, 5, 19–41.
- Maldonado, Ricardo. (1992). *The case of Spanish se*. PhD Dissertation. San Diego: University of California.
- Maldonado, Ricardo. (2000). Conceptual distance and transitivity increase in Spanish Reflexives. In Zygmunt Frajzyngier, & Traci Walker (Eds.), *Reflexives: Forms and functions* (pp. 153–186). Philadelphia/Amsterdam: John Benjamins.
- Maldonado, Ricardo. (2008). Spanish Middle Syntax. A Usage Based Proposal for Grammar Teaching. In Sabine de Knop (Ed.), *Cognitive Approaches to Pedagogical Grammar* (pp. 155–196). Berlin: Mouton de Gruyter.
- Manny, Linda. (2000). *Middle Voice in Modern Greek: Meaning and Function of an Inflectional Category*. Amsterdam/Philadelphia: John Benjamins.
- Manoliu-Manea, Maria. (1994). *Discourse and Pragmatic Constraints on Grammatical Choices*. North Holland Linguistics Series 57. Amsterdam: Elsevier.
- Manoliu-Manea, Maria. (1996). Inalienability and topicality in Romanian: Pragma-semantics of syntax. In Hilary Chappell & William McGregor (Eds.), *The Grammar of Inalienability* (pp. 711–743). Berlin: Mouton de Gruyter.
- Mihalcea, Rada, & Vivi Năstase. (2002). Letter Level Learning for Language Independent Diacritics Restoration *Proceedings of the 6th Conference on Natural Language Learning* (CoNLL 2002), Taiwan.
- O'Connor, Mary Catherine. (1996). The Situated Interpretation of Possessor Raising. In Masayoshi Shibatani & Sandra Thompson (Eds.), *Grammatical constructions: their form and meaning* (pp. 125–156). Oxford: Clarendon Press.
- Payne, Doris L., & Barshi, Immanuel. (Eds.) (1999). *External Possession* Amsterdam / Philadelphia: John Benjamins.
- Shibatani, Masayoshi. (1994). An Integrational Approach to Possessor Raising, Ethical Datives, and Adversatives Passives. *Berkeley Linguistics Society*, 20(1), 461–486.
- Siemund, Peter. (2010). Grammaticalization, lexicalization and intensification: English itself as a marker of middle situation types. *Linguistics*, 48, 797–835.
- Stefanowitsch, Anatol, & Gries, Stefan. T. (2003). Collocations: Investigating the interaction of words and constructions. *International journal of corpus linguistics*, 8(2), 209–243.
- Vasilescu, Andra. (2013). The Semantics of Personal and Reflexive Pronouns. In Dobrovie-Sorin, Carmen & Ion Giurgea (Eds.), *A Reference Grammar of Romanian. Volume I: The noun phrase* (pp. 268–282). Amsterdam/Philadelphia: John Benjamins.

- Vihman, Virve-Anneli. (2002). Middle Voice in Estonian. *Transactions of the Philological Society*, 100(1), 131-160.
- Zimmerling, Anton. (2013). Possessor Raising and Slavic Clitics. In Irina Kor Chahine (Ed.), *Current Studies in Slavic Linguistics* (pp. 43-60). Amsterdam: John Benjamins.