Chapter 8

Derivation in the domain of multiword expressions

Verginica Barbu Mititelu

Romanian Academy Research Institute for Artificial Intelligence

Svetlozara Leseva

Institute for Bulgarian Language, Bulgarian Academy of Sciences

Multiword expressions and derivation have rarely been discussed together, even though analyzing the interaction between them is of great importance for the study of each topic and, in general, for the study of the language and for Natural Language Processing. Derivation is a means of enriching the lexicon with both words and multiword expressions. Various types of derivation (suffixation, prefixation or both, as well as other derivational devices) can act upon either words or multiword expressions. The focus of our work here is the formation of multiword expressions from other multiword expressions via derivation. We analyze the morphological, syntactic and semantic aspects of this process, providing examples from Romanian and Bulgarian, languages, which belong to different families but have been in contact throughout their history. The study can be further extended with data from other languages. The perspective adopted here is paradigmatic, but the syntagmatic approach, which can only be mentioned as further work, will add to the quality of the analysis of facts: corpus data will contextualize the phenomena discussed here and offer quantitative information about them.

1 Introduction

Widely accepted as a difficult task to deal with, the identification of multiword expressions (MWEs) in processing natural languages becomes even more difficult when the MWEs are new creations in the language or even ad-hoc creations in

the text as a result of the linguistic creativity of speakers, usually carrying an emotional load (1):[1]

(1)   *a băga de seamă – băgător de seamă*   (RO)

‘to pay attention to’ – ‘(the one) who only watches without playing any role (in the action)’

In example (1) the latter MWE is derived from the former and carries a negative connotation.

While the interest in the origin of MWEs has been manifested in all languages, specialists have normally investigated the social, economic, ethnographic, and other aspects motivating the process of turning certain word combinations into MWEs. When the origins cannot be found in the national background, MWEs are attributed to other languages, so they are borrowings or linguistic calques. Another (language internal) source of MWEs can be found in the inventory of already existing MWEs. In this paper we focus on one type of MWE formation: derivation from other MWEs, as shown in (1). We put together two topics that have rarely been discussed together in the same study.

On the one hand, MWEs have been classified and characterized according to syntactic and morphological variability (Nunberg et al. 1994; Sag et al. 2002; Baldwin et al. 2003; Baldwin & Kim 2010, among others) and/or semantic decomposability (Nunberg et al. 1994; Baldwin et al. 2003, among others), as well as according to types of idiomaticity (Baldwin 2004; 2006, among others). From a morphological perspective, only inflection and the reflexive form of verbs were discussed for each type of MWE (Sag et al. 2002; Savary 2008).

On the other hand, derivation is a process defined as involving words (Marouzeau 1933): it is the process of creating new words out of existing ones, by means of attaching or detaching affixes to or from a stem respectively, the latter type being better known as back-formation. An example of derivation is the word *survival*, created by attaching the suffix *−al* to the stem *survive*. An example of back-formation is the verb *to back-form*, obtained from *back-formation* by removing the suffix *−ation*. However, derivation can act both on words and MWEs.

---

[1]As convention of writing:

  (i) We adopt the use of the international two letter code of the country in which the language is spoken in front of each example to mark the language to which it belongs: RO for Romanian, BG for Bulgarian.

  (ii) We show the base MWEs on the left and the derived MWEs on the right.

In the former case, it always results in a new word; in the latter, it creates either a new word or a new MWE, as we will show below.

In the literature dedicated to either of the two topics (derivation, MWEs), one can identify two predominant trends: on the one hand, the discussion about derivation has always implied that words are the output and only rarely MWEs; on the other hand, the discussion about MWEs has implied, from time to time, reference to derivation: this interest has also been expressed, although sporadically, in studies on phraseology, particularly in analyzing the behavior of idioms with respect to their derivational morphology.

In this chapter, we describe the way derivation affects MWEs, providing examples from Romanian and Bulgarian, languages which belong to different language families (Romance and Slavic, respectively) but have had a long history of contact. We focus on MWEs derived from other MWEs, highlighting morphological, syntactic and semantic modifications triggered by these transformations.

In both Romanian and Bulgarian, derivation is much more productive than compounding or other internal means of enriching the vocabulary. Moreover, progressive derivation is more frequent than back–formation. In both languages suffixation is the prevalent derivational means. Prefixation in Romanian is much less productive. Bulgarian has a very developed deverbal verb formation as verbal prefixes express aktionsart and the language has a rich Aktionsart system. Cases of prefixation were not found in our data involving cross–part–of–speech derivation. Derivation affects all content word classes, simple or compound words.

2 Types of lexemes derived from MWEs

When subject to derivation, MWEs can serve as bases for the creation of either other MWEs or words. We discuss these types in the subsections below. In our discussion, we will use the term BASE MWE to denote the MWE that serves as the input to the derivation process.

2.1 MWEs derived from MWEs

We offer here some examples of MWEs derived from MWEs, in both Romanian and Bulgarian:

(2) *a mustra cugetul*   *(pe cineva)*   – *mustrare de cuget*   (RO)
     to chide the.conscience (on someone) – chiding by conscience

     'to have remorse' – 'having remorse'

(3) *съвестта   гризе (някого) – гризене на съвестта*   (BG)
    săvestta   grize (nyakogo) – grizene   na săvestta
    the.conscience gnaws (someone) – gnawing of the.conscience
    'to have remorse' – 'having remorse'

(4) *cronică literară – cronicar literar*   (RO)
    'literary review' – 'literary reviewer'

(5) *моден дизайн – моден дизайнер*   (BG)
    moden dizayn – moden dizayner
    'fashion design' – 'fashion designer'

These examples show that different types of MWEs can feed derivations: idioms in (2) and (3), terms in (4) and compounds in (5).

One content word (usually the syntactic head) of the base MWE is subject to affixation: e.g. in (2) above *mustrare* (noun) is derived from the verb *a mustra* with the suffix *–re*; in (4) *cronicar* is derived from *cronică* (the head of the base MWE noun phrase) with the agentive suffix *–ar*. Likewise, in (3) the process of derivation is carried out by means of suffixation of the verb *гриза* (griza) 'gnaw' with the suffix *–не*, thus obtaining the deverbal noun *гризене* (grizene) 'gnawing', and in (5) the head noun *дизайнер* (dizayner) 'designer' is obtained from the noun дизайн, dizayn, 'design' by means of the agentive suffix *–ер*.

2.2 Words derived from MWEs

When one content word of the MWE is subject to affixation and the derived word has the semantic content of the base MWE, we regard this as words being derived from MWEs; the other words of the base MWE simply do not occur in the result of the derivation:

(6) *a face   un lucru mușama – a mușamaliza*   (RO)
    to make a   thing oilcloth
    'to cover something up'

(7) *извадя (бизнес,   ...) на светло – изсветля*   (*бизнес,   ...*)   (BG)
    izvadya (biznes,   ...) na svetlo – izsvetlya   (biznes,   ...)
    bring.v (business, ...) to light   – make.brighter.v (business, ...)
    'to legalize (business, ...)'

(8)   *a face la rotisor – rotis**a***   (RO)
      'to cook in a rotisserie'

(9)   *instala**tor** de gaze – gaz**ist***   (RO)
      'gas installer'

(10)   *въздух под налягане   – въздух**ар***   (BG)
      văzduh pod nalyagane – văzduh**ar**
      'air under pressure' – 'an unreliable or incompetent person (especially one who pretends otherwise)'

This type of derivation involves semantic condensation as one of the content words of the MWE, the one that carries most of the semantic load, takes up the meaning of the whole. The word may be adapted morphologically to express the relevant part of speech, for example by means of suffixation with a verbal suffix, e.g. RO –*iza* (6), where the noun *mușama* 'oilcloth' yields the derivative verb *mușamaliza*, by means of back–formation, e.g. (8) where the verb *rotisa* is created from *rotisor*, or parasynthetically, e.g. BG *из–, –я* (7), where the nominalized adjective светло, svetlo, 'light' gives the verb *из–светл–я* (iz–svetl–ya) 'make brighter'. In addition, noun suffixes, such as the agentive suffixes RO –*ist* (9) and BG –*ap* (10), express the semantic role of the derived noun.

These types of derivation seem to affect collocations (8), terms (9) and idioms (6), (7), (10) alike. In Romanian linguistics, the phenomenon has been described as very frequent and systematic (Groza 2011). However, no quantitative support has been offered for these claims and, as a consequence, we will not adhere to this estimation. In the Bulgarian literature, the specialists have remarked that while dephraseologization (and semantic condensation) is a productive process in the contemporary language, word–formation processes, including derivation, are relatively rare (Blagoeva 2011). We will not investigate this phenomenon here.

3 Data selection and processing

In order to study the behavior of MWEs with respect to derivation, we worked with an inventory of MWEs extracted from big Romanian and Bulgarian dictionaries containing MWEs.

For Romanian, this inventory was created starting from DELS (Dictionary of Expressions, Idioms and Collocations) (Mărănduc 2010). The dictionary was automatically parsed, MWEs were extracted and those marked as archaic were eliminated, along with expressions, as they are unproductive with respect to deriva-

tion; for the remaining 11,158 MWEs (collocations and idioms) we looked for derivationally related MWEs by searching the web and manually inspecting the results. Only for about 500 MWEs could we find derivationally related MWEs. So, a first remark is the relatively low impact that derivation has on MWEs, at least judging from the Romanian data. This may be a reason why the two phenomena have rarely been discussed together.

The Romanian MWEs were preprocessed and annotated morphosyntactically: they were automatically tokenized, lemmatized, tagged for part-of-speech (PoS) and chunked using the TTL web service (Ion 2007). Each word form in the MWE was identified, lemmatized and a PoS tag containing information about its part of speech and morphosyntactic characteristics (number, gender, case, etc., depending on the PoS) was attached to it. Syntactic groups were identified and marked as such, they are called chunks and are useful for the analysis in §6.

The Bulgarian data were excerpted from a large electronic dictionary of MWEs (Stoyanova & Todorova 2014). Named entities were removed since they are unproductive with respect to the phenomena explored in this work. The remaining MWEs were inspected and other unproductive types, such as proverbs, sayings and other expressions, were also filtered out automatically, using as a filter the code of the relevant type of MWE. Finally, obsolete and dialect entries were manually removed. The resulting dictionary of 4,039 entries consists predominantly of verb idioms and support verb constructions. The number of entries reflects two facts: (i) many Bulgarian verbs form aspectual pairs, whose members are distinct lexemes with their own inflectional and derivational morphology; therefore, unless there are semantic restrictions to the contrary, two MWE entries (one headed by a perfective aspect verb and one by an imperfective aspect verb) were encoded in the dictionary; (ii) (to a lesser degree) prefixation is a regular process which creates new verbs. Although the prefixed verbs meaning is modified to a lesser or to a greater extent, a variant of the MWE headed by a prefixed verb is often formed, thus the word family of a verb idiom may include a number of derived verb idioms. In the dictionary we kept only the more frequent MWEs, derived through prefixation, basically those bearing resultative meaning, e.g. (15). We found derivational MWEs for 2,612 entries in the dictionary, with a great prevalence of deverbal MWEs. The data were additionally supplemented with examples collected by the authors, adding up to 2,725 pairs.

The MWEs were automatically tokenized, lemmatized and PoS-tagged using the Bulgarian Language Processing Chain (LPC) (Koeva & Genov 2011), which is available as a web service, and subsequently chunked using a stand-alone tool which uses the LPC output (Stoyanova et al. 2015). As a result, all the words

in each MWE were marked with the relevant grammatical information, and the basic syntactic structure of the MWEs (head, dependent syntactic groups) was identified and marked explicitly.

4 Derivation types in the domain of MWEs

In this section we present the types of derivation detected in the domain of MWEs: progressive derivation by means of suffixes, prefixes or both, back-formation and zero-derivation.

4.1 Progressive derivation

The vast majority of derivation cases are **progressive** (i.e., MWEs are created by adding affixes to a word in a previously existent MWE). In Bulgarian and Romanian, these affixes can be suffixes, prefixes or both. Each subtype will be discussed in the subsections below.

4.1.1 Suffixation

In Romanian all 339 cases of progressive derivation are represented by **suffixation**. In Bulgarian almost all of the 2,704 instances of progressive derivation are accounted for by suffixation, with the exception of 10 cases of parasynthetic derivation. The productivity of the suffixes in the two languages is represented in Table 1.[2]

Other Romanian suffixes (*−a*, *−ime*, *−iza*) are much less productive in the set of pairs we dealt with, with only one or a maximum of two occurrences. In Bulgarian, other suffixes denoting events or results of events are instantiated with only a few examples in the data: *−еж* (three cases), *−ов* (one case), *−ица* (one case). The noun suffix *−ост*, which denotes properties, is found in three cases. Other agentive suffixes are *−ант/−ент* (two cases), *−ия* (two cases), *−ик* (one case). The suffixes *−ура*, *−ище*, *−ция* are found with institutions (one example per suffix).

There are cases when the same MWE serves as a derivational base for two different MWEs. There are two ways in which this can be achieved. The first one is through separate derivational paths. The derivative MWEs in (11) and (12) are formed through independent derivational processes:

---

[2]Abbreviations used in Table 1: Ag: Agent; Ev: Event; Instn: Institution; Instr: Instrument; L: Language; Re: Result; SVs: Semantic Values; St: State.

Table 1: Suffixes. Their productivity and semantics.

L	Suffix	P	SVs	Examples
RO	−*re*	305	Ev	*a−și băga mințile în cap*, to insert one's minds into head, 'to come to reason' – *băgare a minților în cap*, inserting one's minds into head, 'coming to reason'
RO	−*(ă)tor*	18	Ag	*a face rele* 'to do bad things' – *făcător de rele* 'wrongdoer'
RO	−*ie*	7	St	*sărac lipit* 'dog poor' – *sărăcie lipită* 'extreme poverty'
			Ev	*călători de plăcere* 'to travel for pleasure' – *călătorie de plăcere* 'travelling for pleasure'
			Instn	*judecător de pace* 'justice of the peace' – *judecătorie de pace* 'the court of a justice of the peace'
RO	−*ătură*	4	Ev	*a−ți arunca ochii* 'to cast a glance' – *aruncătură de ochi* 'glance'
BG	−*ne*	2,604	Ev, Re	*pisha istoriya*, to write history, 'to make history' – *pisane na istoriya* 'making of history'
BG	−*ba*	5	Ev, Re	*prodam na edro*, to sell in bulk, 'to wholesale' – *prodazhba na edro* 'a wholesale'
BG	−*ach*	42	Ag	*svalyam zvezdi*, to take down stars, 'to promise the moon' – *svalyach na zvezdi* 'one who promises the moon'
			Instr	*hvashtam brimki* 'to mend ladders/stitches (e.g. in stockings)' – *hvashtach na brimki* 'a tool for mending ladders'
BG	−*or/−er/−ir*	13	Ag	*komandvam parada*, to command the parade, 'to call the shots' – *komandir na parada* 'one who calls the shots'
BG	−*tel*	10	Ag	*stroya văzdushni kuli* 'to build castles in the air' – *stroitel na văzdushni kuli* 'one who builds castles in the air'
BG	−*ets*	4	Ag	*tărguvam na edro* 'to deal wholesale' – *tărgovets na edro* 'wholesaler'

(11) a. *a aduce laudă – aducere de laudă*   (RO)
          'to give praise' – 'giving praise'

      b. *a aduce laudă – aducător de laudă*   (RO)
          'to give praise' – 'the one who gives praise'

(12) a. *разбивам сърца – разбиване на сърца*   (BG)
          razbivam sărtsa – razbivane na sărtsa
          'to break hearts' – 'breaking of hearts'

      b. *разбивам сърца – разбивач на сърца*   (BG)
          razbivam sărtsa – razbivach na sărtsa
          'to break hearts' – 'heartbreaker'

The verb MWEs in (11) and (12) undergo suffixation and yield either an eventive noun (by means of the suffixes *–re* and *–не*, respectively) or an agentive one (by the suffixes *–tor* and *–ач*, respectively) in the derivationally related MWEs.

The second way to form two or more MWEs from the same source follows several steps along a single derivational path. We spotted six such instances in the Romanian data and three in Bulgarian (Table 2). Typologically, the examples are different: in Romanian, the noun–to–noun derivation yields antonyms. In Bulgarian, the derived nouns lexicalize different semantic roles in the eventuality denoted by the corresponding verb. Due to the small number of instances, no conclusions can be reached for either of the languages. More examples from

Table 2: Multiple derivations.

Language	Pattern	Productivity	Example
RO	V–N–N	3	*a ști carte*, to know book, 'to be educated' *știință de carte* 'education' *neștiință de carte* 'lack of education'
RO	V–A–A	3	*a ști carte*, to know book, 'to be educated' (*știutor de carte*, 'educated') *neștiutor de carte* 'uneducated'
BG	V–N$_{AGENT}$–N$_{LOCATION}$	3	*pera pari* 'to launder money' *perach na pari* 'money launderer' *perachnitsa na pari* 'a business involved in money laundering'

these languages (as well as from others) would help to better understand possible derivations.

Besides, as the verbs belonging to a given aspectual pair in Bulgarian are characterized by their own derivational morphology and derivational patterns, MWEs (just like single words) headed by different members of an aspectual pair may serve as a base for derived MWEs with similar semantics, e.g. the imperfective aspect verb gives rise to an eventive nominalization (13a) , while the perfective aspect counterpart yields a different deverbal MWE with an eventive (and possibly resultative) interpretation (13):

(13)   a. *побеждавам по точки – побеждаване по точки*   (BG)
          pobezhdavam po tochki – pobezhdavane po tochki

          'to outpoint, to outscore' – 'outpointing'

     b. *победя   по точки – победа по точки*   (BG)
          pobedya po tochki – pobeda po tochki

          'to outpoint, to outscore' – 'outpointing'

4.1.2 Parasynthetic derivation

Another derivational device detected only in the Bulgarian data is parasynthetic derivation, when both a suffix and a prefix are attached to an existing word. All ten cases we found in the data represent derivations of verbs from adjectives:

(14) *гладен като вълк – огладнея   като вълк*   (BG)
      gladen kato vălk – ogladneya kato vălk

      'as hungry as a wolf' – 'to become as hungry as a wolf'

4.1.3 Prefixation

Prefixation alone rarely serves as a means for deriving new MWEs in Romanian (see the examples of consecutive derivation in Table 2). In the Bulgarian data MWEs resulting from verb to verb derivation (15), where prefixation is a productive process, were included as separate entries in the dictionary and will not be discussed further below:

(15) *пера пари – из–пирам пари*   (BG)
      pera pari – iz–piram pari

      'to launder money' – 'to launder money up' (resultative meaning)

We made this decision because the derivationally related verb MWEs have different (although related) meanings and can themselves be subject to derivation, e.g. *пера пари* (pera pari) 'to launder money' – *пра–не на пари* (pra–ne na pari) 'money laundering', *из–пирам пари* (iz–piram pari) 'to launder money up' – *из–пира–не на пари* (iz–pira–ne na pari) 'money laundering', resultative meaning.

4.2 Back-formation

We found only one case of back-formation in Romanian, in which the verb *lucra* is derived from the noun *lucru* (16), and six cases in Bulgarian (17), all of which are neologisms:

(16) *lucru de mână – a lucra de mână*   (RO)
     'handiwork' – 'to work by hand'

(17) *промиване на мозъци – промивам мозъци*   (BG)
     promivane na mozătsi – promivam mozătsi
     'brainwash(ing))' – 'to brainwash' (example from Blagoeva 2008)

These data reflect a tendency noted in works on Bulgarian terminology and neology (Baltova 1986; Kolkovska 1993/1994; Kostova 2013, among others) concerning the creation of eventive nouns, in particular nouns ending in *–не* or ending in a verbal suffix followed by *–не* that do not have a verb counterpart. The corresponding verbs are often formed by back-formation (17) and the newly created verbs or verb MWEs can be subject to further derivations:

(18) *промивам мозъци – промивач на мозъци*   (BG)
     promivam mozătsi – promivach na mozătsi
     'to brainwash' – 'brainwasher'

4.3 Zero-derivation (conversion)

Fifteen cases in the Bulgarian data represent the process of conversion (also called zero-derivation) in which the derived MWE is formed without the attachment of a suffix and/or a prefix and usually involves detachment of a grammatical affix such as the inflection:

(19) *удара под кръста – удар под кръста*   (BG)
       udarya pod krăsta   – udar pod krăsta
       'to hit below the belt' – 'a hit below the belt'

With Romanian MWEs, conversion manifests itself in two ways: (i) the participle form functions as an adjective with more than 150 verb MWEs; (ii) the supine form of several verb MWEs functions as a noun. The participle and the supine are homonymous non-finite verb forms. However, the discussion below will exclude such cases (and therefore zero-derivation) and will focus only on affixal derivation.

5 The morphological classes of the MWE heads involved in MWE derivation

The formal description and analysis of the basic syntactic structure of MWEs and their representation in the lexicon are important for the encoding and prediction of some of the major morphological and syntactic properties of the MWEs, such as: the components that are likely to inflect; the possibilities for modification by optional elements (optional elements are placed in brackets), e.g. BG *пера (мръсни) пари* (pera (mrăsni) pari) 'launder (dirty) money'; the possibility for eliding modifiers with no change in meaning (placed in square brackets in this example), e.g. *вдигам летвата [високо]* (vdigam letvata [visoko]) 'raise the bar [high]'; paradigmatic restrictions on agreement, on singular/plural forms, and so forth. Among others, the syntactic analysis makes it possible to predict the potential of MWEs for derivation and the structural changes that may take place in this process (see §6).

The majority of the Romanian pairs extracted from the DELS involve verbs as bases for derivation. The most frequent type is represented by pairs of MWEs displaying verb nominalization, while derivative pairs involving other parts of speech are much rarer (see Table 3).

For Bulgarian 2,725 derivative pairs were found. The difference in the number of pairs as compared with the initial set of 4,039 entries is due largely to the fact that the perfective aspect verbs in the set are very unproductive with respect to the derivational processes discussed. Deverbal noun formation accounts for the majority of cases (2,663), with much smaller numbers for the opposite noun to verb pattern, verb to adjective, adjective to verb, noun to noun, adjective to noun (see Table 3).

Table 3: Morphological alternations occurring in MWE derivations.

Stem PoS-derived word PoS	#RO examples	#BG examples	Example from the RO data	Example from the BG data
V–N	349	2,663	*a depune jurământul* 'to take the oath' *depunerea jurământului* 'taking the oath'	*potrivam rătse* 'to rub (one's) hands' *potrivane na rătse* 'rubbing of (one's) hands'
V–A	2	16	*a sări în ochi*, to jump into eyes, 'to be straightforward' *săritor în ochi*, jumping into eyes, 'straightforward'	*mălcha kato păn*, to keep silent like a log, 'to be as mute as a maggot/fish' *mălchaliv kato păn*, silent like a log, '(as) mute as a poker'
N–V	4	8	*semnal luminos* 'light signal' *a semnaliza luminos* 'to signal with lights'	*igra na nervi* 'a battle of nerves' *igraya na nervi* 'to lead a battle of nerves'
N–N	5	18	*judecător de pace* 'justice of the peace' *judecătorie de pace* 'the court of a justice of the peace'	*voenen prokuror* 'military prosecutor' *voenna prokuratura* 'military prosecutor's office'
A–N	4	10	*sărac lipit* 'dog–poor' *sărăcie lipită* 'extreme poverty'	*nisht duhom* 'poor in spirit' *duhovna nishteta* 'spiritual poverty'
A–V	–	10	–	*byal kato platno* 'as white as a sheet' *pobeleya kato platno* 'to become as white as a sheet'

6 Syntactic reorganizations resulting from derivations

Dependency Grammar is used as a syntactic framework for our discussion. In this framework, verbs admit subjects, complements and adjuncts, nouns (even those derived from verbs) admit modifiers and adjectives admit complements. Syntactic functions are understood as in Quirk et al. (1985).

Out of the total number of 414 Romanian pairs, fifty do not undergo any internal reorganization in the process of derivation; in Bulgarian this holds true for 54 out of the 2,725 pairs:

(20)   *agent de publicitate – agenție de publicitate*   (RO)
       'advertising agent' – 'advertising agency'

(21)   *военен прокурор – военна прокуратура*   (BG)
       voenen prokuror – voenna prokuratura
       'military prosecutor' – 'military prosecutor's office'

In (20), *de publicitate* receives the same syntactic analysis in both MWEs: it is a modifier of the nouns *agent* and *agenție*, respectively. In (21) the adjectives *военен* (voenen) 'military' and *военна* (voenna) 'military' which modify the head noun *прокурор* (prokuror) 'prosecutor' and *прокуратура* (prokuratura) 'prosecutor's office', respectively, have the same analysis.

The cases without syntactic reorganization include the noun to noun, verb to adjective and adjective to verb patterns. In the following sections we will deal with the other two structural types of MWEs found in the data, that is: verb to noun and noun to verb MWEs.

The syntactic structure of the base MWE determines whether the syntactic expression of a dependent phrase is obligatory. For instance, a direct object NP_{DO} ($_{DO}$ stands for direct object) that is not a fixed part of a base MWE but is licensed by a transitive verb, as illustrated below, is not an obligatory dependent of the derived MWE, while an internal argument that is a fixed part of the base MWE is an obligatory component of the derived MWE. For example, in BG, *пъхвам (някого) зад решетките* (păhvam (nyakogo) zad reshetkite) 'put (someone) behind bars', the internal argument position (NP_{DO}) is not a fixed part of the idiom; rather, it is an open position that is filled by a suitable entity. In the nominalization *пъхване (на някого) зад решетките* (păhvane (na nyakogo) zad reshetkite) 'putting (of someone) behind bars' the position corresponding to the direct object may be left empty. On the contrary, if the NP_{DO} is a fixed part of the MWE, it cannot be omitted, e.g. *кърша ръце* (kărsha rătse) 'wring hands' – *кършене на*

ръце (kărshene na rătse) 'wringing of hands'. The syntactic structure of the base MWE also determines the word order of obligatory and non–obligatory components in the derived MWE (e.g. typically the object of the base MWE is closer to the deverbal noun than other base MWE components).

Next, we present the syntactic reorganizations observed in derived MWEs, as we found them in the available data for Romanian and Bulgarian. Their documentation facilitates text processing. Given the limited MWE dictionaries available, all knowledge facilitating the automatic morphosyntactic analysis of text is considered valuable. Below we offer rules that algorithms can use to process new MWEs which are derived from existing ones.

6.1 Verb PP or AdvP complement or adjunct – noun modifier

This pattern is observed when a verb MWE is related with a noun MWE via derivation (see (i1) below) or the other way round (i2). It accounts for 260 Romanian and 792 Bulgarian pairs:[3]

(i) (1) VP [V PP/AdvP] > NP [$N_{V–derived}$ PP/AdvP]
    (2) NP [N PP/AdvP] > VP [$V_{N–derived}$ PP/AdvP]

The verb admits a prepositional phrase (PP) or an adverbial phrase (AdvP) functioning as a complement or an adjunct in the MWE, but it can also admit other modifiers placed out of the MWE. Through derivation, the constituents, except for the head word, preserve their syntactic category and internal structure, as can be noticed in (i), where the form of the modifying phrase is the same. Although semantically the dependent functions similarly in the NP and the VP, its syntactic role is different according to our analysis: when the head is a verb, we analyse the particular dependent as a complement or an adjunct and, when the head is a noun, we analyse it as a modifier.

Below we indicate the syntactic category (PP, AdvP) and the status (complement, adjunct, modifier) of the dependent phrases. Complement or adjunct status is determined with respect to the argument structure of the verb that heads the respective MWE.

In the verb MWE in (22), *de credință* is the prepositional object (i.e., a complement) of the reflexive verb *a se lepăda*, whereas in the noun MWE the same PP is a modifier of the noun *lepădare* derived from *a se lepăda* with the suffix *–re*. Likewise, in the BG verb MWE in (23) the Goal PP *в джоба* (v dzhoba) 'in the pocket' is the prepositional object of the verb *бъркам* (bărkam) 'thrust

---

[3]Patterns are enumerated in the text with Roman numbers.

one's hand', whereas, in the noun MWE, it functions as a modifier of the noun *бъркане* (bărkane) 'thrusting one's hand', derived from the verb *бъркам* (bărkam) by means of the suffix *–не*.

(22) *a se lepăda de credință – lepădare de credință*   (RO)
      'to depart from the faith' – 'departing from the faith'

(23) *бъркам   в джоба   (на някого) – бъркане в джоба*   (BG)
      bărkam   v dzhoba   (na nyakogo) – bărkane v dzhoba
      thrust.one's.hand.v in the.pocket (of someone)
      'to incur expenses (on someone)' – 'incurring of expenses'

The PP *în peniță* (24) is a modifier in the first MWE, and an adjunct of the verb in the second MWE and the verb is derived from the noun in the former MWE. In (25), the PP *под кръста* (pod krăsta) 'below the belt' is an adjunct of the verb in the first MWE and a modifier of the noun *удар* (udar) 'hit' in the second MWE and the noun is derived from the verb *удара* (udarya) 'hit'.

(24) *desen în peniță – a desena în peniță*   (RO)
      'pen drawing' – 'to draw in pen'

(25) *удара   под кръста – удар под кръста*   (BG)
      udarya pod krăsta   – udar pod krăsta
      'to hit below the belt' – 'a hit below the belt'

In (26) the adverb *aminte* is a complement of the verb in the former MWE and a noun modifier in the latter. In (27) the adverb *отвисоко* (otvisoko) 'from above' is an adjunct of the verb *гледам* (gledam) 'look' in the former MWE and a noun modifier in the latter.

(26) *a lua aminte – luare aminte*   (RO)
      'to take into consideration' – 'taking into consideration'

(27) *гледам отвисоко (някого   / нещо) – гледане отвисоко*   (BG)
      gledam otvisoko   (nyakogo / neshto) – gledane otvisoko
      'to look down on (someone / something)' – 'looking down on (someone / something)'

In both languages derivation from a noun MWE to a verb MWE is much rarer:

(28) *игра   на нерви – играя на нерви*   (BG)
      igra   na nervi – igraya na nervi
      a.play of nerves – play.v   nerves

      'a battle of nerves' – 'to lead a battle of nerves'

6.2 Subject complement or object complement – noun modifier

In the Bulgarian data we detected a small number of verb MWEs that have a subject complement or an object complement (Quirk et al. 1985; Downing 2014) as part of their structure.[4] Syntactically, these complements are expressed as NPs, PPs or APs. We use the notation ($_{CS}$) for subject complements and ($_{CO}$) for object complements.

With this type of derivation, the verb MWE subject complement turns up as a modifier in the derived noun MWE (12 cases altogether). The derivation may be represented as in (ii).

(ii) VP [V NP$_{CS}$/PP$_{CS}$/AP$_{CS}$] > NP [N$_{V-derived}$ NP$_{CS}$/PP$_{CS}$/AP$_{CS}$]

The derivation involves a copular verb, such as *съм* (săm) 'be', *ставам* (stavam) 'become', *оставам* (ostavam) 'remain' or a verb that is not a typical copula (e.g., *отивам* (otivam) 'go') and admits a subject complement in the MWE. The deverbal noun (N$_{V-derived}$) derived from this verb heads the noun MWE, and the subject complement turns up as a post–modifier that preserves both its syntactic category and the type of syntactic linking to the head word. The examples below illustrate subject complements – PP$_{CS}$ (29), AP$_{CS}$ (30), N$_{CS}$ (31).

(29) *ставам   за смях   – ставане за смях*   (BG)
      stavam   za smyah   – stavane   za smyah
      to.become for ridicule – becoming for ridicule

      'to become a laughing stock' – 'becoming a laughing stock'

(30) *ставам   разноглед – ставане   разноглед*   (BG)
      stavam   raznogled – stavane   raznogled
      to.become cross–eyed – becoming cross–eyed

      'to become confused or overwhelmed (by something)'

---

[4] A subject or an object complement is a constituent that does not represent a new participant but completes the predicate by adding information about the subject or the object referent, respectively (Downing 2014), e.g. *a separate notion* in *The country became a separate notion*, *young* in *He died young* (subject complement); *a genius* in *People considered Picasso a genius* (object complement).

(31) *отивам войник – отиване войник*   (BG)
      otivam voynik – otivane voynik
      go.v   a.soldier – going   a.soldier
      'to go into the army' – 'going into the army'

Derivations involving an object complement are exemplified with 44 cases in the data. It typically applies on transitive verbs (but verbs admitting PP–object do occur, see (iv), (34) below). The direct object (NP_{DO}) is licensed by the verb that heads the MWE but it is not a fixed part of the MWE. In the formal representation this NP_{DO} is enclosed in curly brackets "{}". As we are particularly concerned with the way the structure of the MWE is reorganized, we do not consider the expression of the MWE–external NP_{DO} if it occurs, although it obeys the rules applying to any direct object:

(iii) VP [V {NP_{DO}} NP_{CO}/PP_{CO}/AP_{CO}] > NP [$N_{V-derived}$ {PP [P NP_{DO}]} NP_{CO}/ PP_{CO}/AP_{CO}]

Here are examples of an MWE headed by a transitive verb with different realizations of the object complement: an AP_{CO} (32) and a PP_{CO} (33):

(32) *дера (някого) жив – дране   жив (на някого)*   (BG)
      dera (nyakogo) zhiv – drane   zhiv (na nyakogo)
      skin.v (someone) alive – skinning alive (of someone)
      'to cause great trouble (to someone)'

(33) *правя (някого / нещо)   на решето – правене на решето*   (BG)
      pravya (nyakogo / neshto)   na resheto – pravene na resheto
      make.v (someone / something) a riddle – making a riddle
      (*на някого   / нещо*)
      (na nyakogo   / neshto)
      (of (someone / something))
      'to make a lot of holes in someone/something, to riddle some- one/something'

The lack of preposition insertion is a structural difference between the derivations involving subject/object complement NPs and direct object NPs, since the latter normally turn up as prepositional modifiers of the corresponding deverbal nouns. We leave aside the marginal cases of direct objects (not introduced by a preposition) that occasionally co–occur with the canonical form in formal administrative language.

The derivation involving an MWE headed by a verb that admits a prepositional object (PP$_O$) has the following representation (iv):

(iv) VP [V PP$_O$ AP$_{CO}$/NP$_{CO}$/PP$_{CO}$] > NP [N$_{V-derived}$ PP$_O$ AP$_{CO}$/NP$_{CO}$/PP$_{CO}$]

(34) exemplifies an MWE with a verb admitting a PP–object. The prepositional object retains its syntactic expression when it turns up as an NP modifier.

(34) *казвам на черното бяло   – казване на черното бяло*   (BG)
       kazvam na chernoto byalo – kazvane na chernoto byalo
       call.v   to the.black white – calling   to the.black white

       'to call black white' – 'an instance of calling black white'

6.3 Subject or direct object – noun modifier

This particular derivation pattern concerns fixed subject verb MWEs or fixed direct object verb MWEs. Unlike the cases discussed in §6.1 and §6.2, in this category nominalization triggers either insertion of a preposition (in both languages) that introduces the former subject or direct object as a prepositional noun modifier (see §6.3.1), or mapping of the former subject or direct object into a genitive modifier (only for Romanian) (see §6.3.2). We will reserve the term 'genitive modifier' for modifiers whose head noun is marked with the genitive case.

6.3.1 Subject or direct object – prepositional modifier

A subject (v) or a direct object (vi) in a verb MWE turns up as a prepositional modifier of the corresponding deverbal noun that heads the corresponding noun MWE. In (v1) and (vi1) a noun MWE is derived from a verb MWE, while in (v2) and (vi2) a verb MWE is derived from a noun MWE.

(v) (1) NP$_S$ V > NP [N$_{V-derived}$ PP[P NP$_S$]]
      (2) NP [N PP[P NP$_S$]] > NP$_S$ V$_{N-derived}$

(vi) (1) VP [V NP$_{DO}$] > NP [N$_{V-derived}$ PP [P NP$_{DO}$]]
       (2) NP [N PP [P NP$_{DO}$]] > VP [V$_{N-derived}$ NP$_{DO}$]

There are eight pairs in Romanian and seventy-five in Bulgarian involving the subject, and thirty-two pairs in Romanian and 1,732 in Bulgarian involving the direct object.

In Romanian the preposition *de* is always used and Bulgarian usually adds the preposition *на*. Both prepositions can be glossed in English with *of*. In Bulgarian, other prepositions may occur, generally when the noun is derived with a suffix other than *–не* or *–ние* (the prevalent suffixes for eventive and/or resultative deverbal nouns) or by other derivational means, e.g. zero-derivation, *обичам* (obicham) 'to love' – *обич* (obich) 'love'.

We repeat (2) and (3) as (35) and (36), respectively, in order to exemplify a case where the subject of a verb MWE corresponds to a prepositional modifier of the deverbal noun that heads the derived noun MWE. The subjects of the verb MWEs (*cugetul* and *съвестта* (săvestta)) correspond to the prepositional modifiers of the deverbal nouns *mustrare* and *гризене* (grizene) which are derived from the verbs *mustra* and *гриза* (griza) respectively.

(35) *a mustra cugetul   (pe   cineva)   – mustrare de*   (RO)
      to chide   consciousness (ObjMarker somebody) – chiding   by
      cuget
      consciousness
      'to have remorse' – 'having remorse'

(36) *съвестта   гризе (някого)   – гризене на съвестта*   (BG)
      săvestta   grize (nyakogo) – grizene   na săvestta
      the.consciousness gnaws (someone) – gnawing of the.consciousness
      'to have remorse' – 'having remorse'

(37) and (38) exemplify the case where the direct object of the verb MWE corresponds to a prepositional modifier of the deverbal noun that heads the derived noun MWE (derived from the verb) of the noun MWE. The direct objects *carte* and *локуми*, lokumi correspond to the prepositional modifiers of the (negative) adjective *(ne)știutor* and the noun *разтегач* (raztegach) respectively, which have been derived from the verbs *şti* and *разтягам* (raztyagam) respectively.

(37) *a şti   carte – (ne)știutor   de carte*   (RO)
      to know book – (not)knowing of book
      'to be educated' – '(un)educated'

(38) a. *разтягам локуми*   (BG)
         raztyagam lokumi
         'to spin yarn, to tell tales'

b. *разтегач на локуми*   (BG)
raztegach na lokumi
spinner   of yarn
'yarnspinner'

As noted in §6.2, when the direct object is not a fixed part of the MWE, it may be left unexpressed. This is not the case with direct objects that are fixed parts of base MWEs: they are neither left out nor replaced with a possessive pronoun.

Apart from a direct object, the MWEs in this and in other categories may have other constituents, e.g. (an)other complement(s), such as a prepositional object or an adjunct. These constituents preserve their syntactic category and the syntactic link to the head word, but assume a different syntactic status (similarly to what was presented in §6.1.).

(39)   *a despica firul   în patru – despicarea firului   în patru*   (RO)
to split   the.hair in four – splitting   of.a.hair in four

'to make small and overly fine distinctions'

(40)   *цепя   стотинката на две – цепене   на стотинката на две*   (BG)
tsepya stotinkata   na dve – tsepene na stotinkata   na dve
split.v the.penny   in half – splitting of the.penny   in half

'to be very stingy'

In (39) and (40), *în patru* 'in four' and *на две* (na dve) 'in half' are PPs functioning as adjuncts in the VP and as modifiers in the derived NPs.

In the Bulgarian data we found MWEs headed by verbs that take a direct object and an object complement that are both part of the MWE (three pairs). The construction has the following form (vii):

(vii) VP [V NP_{DO} $AP_{CO}/NP_{CO}/PP_{CO}$] > NP [$N_{V-derived}$ PP [P NP_{DO}] $AP_{CO}/$ NP_{CO}/PP_{CO}]

The syntactic status of the MWE constituents in the derived structure is predictable: in the noun MWE, the NP_{DO} constituent of the base verb MWE corresponds to a modifier introduced by a preposition (*на*) and the object complement phrase of the base MWE turns up as a modifier of the derived MWE that is expressed in the same way: as an AP in (41) – *развързани* (razvărzani) 'untied', as an NP in (42) *бяло* (byalo) 'white' and as a PP in (43) – *с истинските им имена* (s istinskite im imena) 'by their proper names':

(41) *оставям ръцете* (*на някого*) *развързани* – *оставяне на* (BG)
ostavyam rătsete (na nyakogo) razvărzani – ostavyane na
leave.v the.hands (of someone) untied – leaving of

ръцете (*на някого*) *развързани*
rătsete (na nyakogo) razvărzani
the.hands (of someone) untied

'to untie someone's hands' – 'untying someone's hands'

(42) *наричам черното бяло* – *наричане на черното бяло* (BG)
naricham chernoto byalo – narichane na chernoto byalo
call.v the.black white – calling of the.black white

'to call black white' – 'an instance of calling black white'[5]

(43) *наричам нещата с истинските им имена* – *наричане на* (BG)
naricham neshtata s istinskite im imena – narichane na

нещата с истинските им имена
neshtata s istinskite im imena

'call things by their proper names' – 'calling things by their
proper names'

6.3.2 Subject or direct object – genitive modifier

In this case the subject or the direct object of the verb MWE corresponds to a
genitive modifier in the derived MWE. The reorganization may be represented
as (viii) and (ix) for the subject and the object, respectively:

(viii) NP_S V > NP [$N_{V–derived}$ $NP_{S–Genitive}$]

(ix) VP [V NP_{DO}] > NP [$N_{V–derived}$ $NP_{DO–Genitive}$]

We encountered 12 MWEs described by (viii) in our data. As shown in (44),
the subject *întunericul* corresponds to the genitive modifier of the noun *lăsarea*
derived from the verb *lăsa*:

---

[5]The expression in (42) is synonymous to the one in (34). The structural difference is due to the
different syntactic properties of the synonymous verbs казвам, kazvam, 'call' and наричам,
naricham, 'call': казвам takes a PP-object in the respective sense, while наричам takes an NP
object.

(44) *se   lasă   întunericul   – lăsarea   întunericului* (RO)
REFL lower the.darkness – lowering of.the.darkness
'it is getting dark' – 'the fact of getting dark'

Forty–six pairs display the type of derivation where the direct object of the verb MWE corresponds to a genitive modifier in the noun MWE (45):

(45) *a băga zâzanie   – băgarea zâzaniei* (RO)
to insert dissension – insertion of.dissension
'to sow dissent' – 'the sowing of dissent'

The direct object *zâzanie* corresponds to a genitive modifier of the noun *bă-garea*, derived from the verb *a băga*.

6.4 Adjunct – adjectival modifier

In this case, the adjunct (either a prepositional or an adverb phrase) modifying the verbal head of an MWE corresponds to an adjectival phrase in a noun MWE derived from a verb MWE (see (x1) below) or vice versa – a verb MWE is derived from a noun MWE and the adjective modifier in the noun MWE corresponds to an adjunct (either a prepositional or an adverb phrase) in the verb MWE (see (x2) below). This structure, represented in (x), was detected in six Romanian and sixteen Bulgarian pairs.

(x) (1) VP [V PP/Adv(P)] > NP [N$_{V-derived}$, A(P)]
     (2) NP [N$_{V-derived}$, A(P)] > VP [V PP/Adv(P)]

In Romanian the modifying adjective usually occurs after the modified noun. The normal position of the modifier in Bulgarian is to the left of the modified noun. The comma in (x) is used for signalling the possibility of having the modifier and the modified noun in either order with respect to each other.

Here are some examples of this type of syntactic reorganization:

(46) *arest preventiv – a aresta preventiv* (RO)
'preventive detention' – 'to subject to preventive detention'

(47) *честна игра – играя честно* (BG)
chestna igra – igraya chestno
'a fair play' – 'to play fairly'

In (46) *preventiv* is an adjective modifying the noun *arest* in the first MWE and an adverb modifying the verb *aresta*, derived from *arest*. Likewise, in the BG example (47), *честна* (chestna) 'fair' is the adjectival modifier of the noun *игра* (igra) 'play' and *честно* (chestno) 'fairly' is the adverb modifier of the verb *играя* (igraya) 'play'.

An example of a derivation of a noun MWE with an adjective modifier (*трезва* (trezva) 'straight') from a verb MWE with an adverb (*трезво* (trezvo) 'straight') is shown in (48):

(48)   мисля   трезво – трезва   мисъл   (BG)
       mislya trezvo   – trezva   misăl
       'to think straight' – 'straight thinking'

Examples (46), (47) and (48) involve an adverb modifier in the verb MWE. The other type of construction presented in (x) (involving a PP adjunct) is exemplified by (49):

(49)   търгувам   на едро – едра търговия   (BG)
       tărguvam   na edro – edra târgoviya
       'to deal wholesale' – 'wholesaling'

The PP modifier *на едро* (na edro) 'big/in bulk' of the verb MWE *търгувам на едро* (tărguvam na edro) 'to deal wholesale' corresponds to the adjective modifier *едра* (edra) 'big' of the noun MWE *едра търговия* (edra târgoviya) 'wholesaling'. Note that variants are possible where the PP adjunct of the verb MWE becomes a PP post–modifier in the noun MWE; these cases fall under §6.1.

Table 4 sums up the data presented in §6, along with their share (in percentage) in the overall number of cases that undergo syntactic reorganization (364 for Romanian and 2,671 for Bulgarian).

Several conclusions can be drawn. The MWEs in which the fixed base MWE subject corresponds to a fixed PP modifier in the derived MWE or vice versa (§6.3.1) have a very similar share in the two languages and, as the numbers show, the construction is relatively rare. The same holds for verb MWE adjuncts that turn up as adjectival modifiers or vice versa (§6.4). The cases involving subject complements or object complements (§6.2) are found only in Bulgarian. Still, this pattern is potentially productive as the head verbs involved in it are very common (e.g. *make/do*). In Romanian, the correspondence between a subject or an object and a genitive modifier (§6.3.2) is more common than the correspondence to a PP modifier (§6.3.1).

Table 4: Distribution of Romanian and Bulgarian MWEs across types.

Type	No. of RO examples	No. of examples	RO data %	BG data %
§6.1 PP/AdvP	260	792	71.2%	29.65%
§6.2 Subject/object complement	0	56	0%	2.1%
§6.3.1 Subject	8	75	2.5%	2.81%
§6.3.1 Object	32	1,732	8.8.%	64.84%
§6.3.2 Subject	12	0	3.3%	0%
§6.3.2 Object	46	0	12.63%	0%
§6.4 Adjunct	6	16	1.65%	0.6%

There is a striking difference with respect to the prevalent base MWE structure in each of the languages. In Romanian the most frequent construction is verb–prepositional object/adjunct (§6.1), while the verb–direct object construction is quite uncommon (§6.3.1). In Bulgarian the most frequent type is verb–direct object (§6.3.1); verb–prepositional object/adjunct (§6.1) is also typical, although there are twice as many verb–direct object constructions. This points to a significant difference in the syntactic expression of complements (as reflected in the structure of MWEs); PP objects are by far the preferred choice in Romanian, while in Bulgarian both direct objects and PP objects are common, with a marked preference for the former.

7 Semantics of the derivational patterns

In this section we present the semantic aspects of the MWEs that involve derivation. Although we refer to the semantics of the base MWE, we are more interested in the semantics of the derived MWEs. Tables 5 and 6 offer an overview of the derived MWE semantics.[6]

In Romanian, the great majority of the base MWEs (349) designate events. This remark correlates with the data in Table 4, where most base MWEs are verbs. Furthermore, the derived nominalizations (mostly with the suffix –*re*) also denote events (322 cases): this correlates with the number of V–N pairs in Table 3.

---

[6]For verb–noun pairs we used the inventory of morpho–semantic relations from PWN (Fellbaum et al. 2009), but we added to it some roles whenever they proved necessary.

Table 5: Semantics of the base and the derived MWEs. Frequencies and examples.

Language	Base MWE	Derived MWE	Occurrences	Examples
RO	Event	Event	322	*a cădea în păcat* 'to fall into sin' *cădere în păcat* 'falling into sin'
BG	Event	Event	2,590	*promivam mozătsi* 'to brainwash' *promivane na mozătsi* 'brainwash'
RO	Event	Agent	18	*a vâna zestre* 'to hunt dowry' *vânător de zestre*, hunter of dowry, 'fortune hunter'
BG	Event	Agent	53	*promivam mozătsi* 'to brainwash' *promivach na mozătsi* 'one who does brainwashing'
RO	State	State	11	*a mustra cugetul (pe cineva)* 'to have remorse' *mustrare de cuget* 'having remorse'
BG	State	State	12	*zhiveya po tsarski* 'to live regally' *zhivot po tsarski* 'a regal life'
RO	Event	Instrument	3	*arunca flăcări* 'to throw flames' *aruncător de flăcări*, thrower of flames, 'flamethrower'
BG	Event	Instrument	3	*razbărkvam karti* 'to shuffle cards' *razbărkvach na karti* 'card shuffler'
RO	State	Experiencer	3	*a voi binele* 'to wish well' *voitor de bine*, wisher of well, 'well–wisher'
RO	Event	Distance	3	*arunca cu băţul* 'to throw with a stick' *aruncătură de băţ* 'as far as the stick can be thrown'
BG	Event	Institution	5	*kova zakoni*, forge laws, 'to create and promulgate laws' *kovachnitsa na zakoni*, smithy of laws, 'the parliament'

Table 6: Semantics of the base and the derived MWEs. Frequencies and examples.

Language	Base MWE	Derived MWE	Occurrences	Examples
RO	Job	Institution	8	*judecător de pace* 'justice of the peace' *judecătorie de pace* 'the court of a justice of the peace'
BG	Job	Institution	3	*voenen prokuror* 'military prosecutor' *voenna prokuratura* 'military prosecutor's office'
RO	Event	Vehicle	1	*a vâna submarine* 'to hunt for submarines' *vânător de submarine*, hunter of submarines, 'a vessel for locating and attacking submarines'
RO	Result	Action	1	*lucru de mână* 'handiwork' *a lucra de mână* 'to work by hand'
RO	Artefact	Event	2	*desen în peniță* 'pen drawing' *a desena în peniță* 'to draw in pen'
RO	Event	Characteristic	2	*a sări în ochi*, to jump into eyes, 'to be straightforward' *săritor în ochi*, jumping into eyes, 'straightforward'
BG	Event	Charac–teristic	8	*rabotya kato vol*, work like an ox/horse, 'to work hard' *rabotliv kato vol*, as hard–working as an ox, 'very hard–working'
BG	State	Characteristic	8	*mălcha kato păn*, keep silent like a log, 'to be as mute as a maggot/fish' *mălchaliv kato păn*, silent like a log, '(as) mute as a maggot/fish'
BG	Inchoative state	Characteristic	10	*gladen kato vălk* 'as hungry as a wolf/bear' *ogladneya kato vălk* 'to become as hungry as a wolf/bear'
BG	Job	Agent	18	*softuerno inzhenerstvo* 'software engineering' *softueren inzhener* 'software engineer'

The Bulgarian data also support the productivity and regularity of the derivation of eventive nominalizations (2,590) predominantly with the suffix −*не*. Another interesting tendency (though represented by few examples), especially with respect to neologisms, is the back−formation of verbs from nouns. The other semantic types encountered with derived MWEs constitute a small number of the overall data. The derivatives such as agents, experiencers, instruments and locations are derived primarily from VP [V NP$_{DO}$] MWEs, and less frequently from VP [V PP/AdvP]. No examples were found for such nouns derived from MWEs with the following syntactic structure: NP$_S$ V or VP [V NP$_{CS}$/PP$_{CS}$/AP$_{CS}$].

The productivity of event nominalization is not unexpected, because in the process of MWE-to-MWE derivation the majority of cases account for idiomatic (partial) predicate−argument structures. As the structure of eventive nominalizations may reflect the argument structure of the base verb (Grimshaw 1990), it readily renders these idiomatic structures. The frequency of use of eventive nominalizations, whether single words or MWEs, is substantiated by the fact that they make it possible to refer to an action/event regardless of its doer and the time of occurrence (as expressed by verbal categories) (Pometkova 2006), and hence they may be used interchangeably with the verb-headed construction, or even may be preferred contextually in certain cases or in certain registers, such as scientific discourse.

In Romanian, the number of nominalizations increases greatly (by almost 200 cases in our data) when taking into account supine forms of verbs which, via conversion, become nouns; and the rest of the MWE behave, in these cases, similarly to the cases displaying affixal derivation, i.e. almost the same types of syntactic changes occur. Moreover, supine nouns are (for more than 150 cases in our data) alternatives to derived nominalizations, with a semantic difference: Cornilescu (2001) maintains that −*re* nominalizations tend to express results, while supines express events.

As the data show, other types of verb−noun derivational patterns, such as the ones resulting in agents, experiencers, instruments, locations and so forth, are significantly fewer in number. In our opinion, the semantic grounds for this phenomenon is that the situations described by the respective verb MWEs frequently do not conceptualize a particular type of agent, experiencer or instrument and so forth that needs to be lexicalized. Moreover, in terms of their semantic and syntactic properties, these types of nouns do not as readily inherit and express the base verb arguments and/or adjuncts. This is supported by the fact that when the need arises for expressing the relevant agentive or instrumental, etc. meaning, participle−headed constructions are preferred, at least in Bulgarian (50).

(50)   *вземам решение – вземащ   решение*   (BG)
       vzemam reshenie – vzemasht reshenie

       'to make a decision' – '(the one) making a decision'

These participial constructions may be either contextually used or may undergo nominalization and lexicalization.

Besides, another word-formation device that is also frequent is compounding, in which case the arguments/adjuncts are incorporated in the word structure. Here are some examples of one-word compounds that have MWE counterparts in the data e.g. *сърцеразбивач* (sărtserazbivach) 'heartbreaker', *миторазбивач* (mitorazbivach) 'mythbuster', *кодоразбивач* (kodorazbivach) 'codebreaker', *монетосекач* (monetosekach) 'coiner/minter', etc.

8 Conclusions

Putting MWEs and derivation together, we notice that derivation affects MWEs, creating either words or other MWEs. The productivity of this phenomenon seems to depend on language characteristics: Bulgarian, a language with aspect, allows for more cases of derivation than Romanian, which lacks aspect. Another factor influencing productivity is the data set: Romanian DELS lacks terms, which do occur in the Bulgarian dictionary and are productive in terms of derivation, serving the need for expressing different actors, instruments, objects, places, etc. within a domain of activity.

We have presented data from Bulgarian and Romanian. However, derivation has been reported to act upon MWEs in other languages: Piela (2007) discusses examples of words created from idioms and argues that this process is productive in Polish; in Russian, the process of creating MWEs from MWEs seems to be the most productive internal means of MWE formation (Ermakova et al. 2015). We can conclude that MWEs are subject to derivation in more languages and comparing and contrasting them from such a perspective can be of linguistic interest.

9 Acknowledgements

Abbreviations

AG	Agent	INSTR	Instrument
BG	Bulgarian	L	Language
LPC	Bulgarian Language Processing Chain	POS	part of speech
		RE	Result
DELS	Dictionary of Expressions, Idioms and Collocations	RO	Romanian
		SVS	Semantic Values
EV	Event	ST	State
INSTN	Institution	V	Verb (in the glosses)

References

Baldwin, Timothy. 2004. *Multiword expressions. Advanced course.* Australasian Language Technology Summer School (ALTSS 2004). Sydney, Australia.

Baldwin, Timothy. 2006. *Compositionality and multiword expressions: Six of one, half a dozen of the other.* Invited talk given at the COLING/ACL'06 Workshop on Multiword Expressions: Identifying & Exploiting Underlying Properties.

Baldwin, Timothy, Colin Bannard, Takaaki Tanaka & Dominic Widdows. 2003. An empirical model of multiword expression decomposability. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, vol. 18, 89–96. Sapporo.

Baldwin, Timothy & Su Nam Kim. 2010. Multiword expressions. In Nitin Indurkhya & Fred J. Damerau (eds.), *Handbook of Natural Language Processing*, 2nd edn., 267–292. Boca Raton: CRC Press.

Baltova, Yulia. 1986. Za nyakoi yavleniya i tendentsii v izgrazhdaneto na leksikalnata sistema na savremenniya balgaski knizhoven ezik. In *Vaprosi na savremennata balgarska leksikografiya i leksikologiya*, 74–80. Sofia: BAS Publishing House.

Blagoeva, Diana. 2008. Novi frazeologichni kalki v balgarskiya ezik (v sapostavka s drugi slavyanski ezitsi). In *Izsledvaniya po frazeologiya, leksikologiya I leksikografiya*, 149–153. Sofia: Prof. M. Drinov Publishing House.

Blagoeva, Diana. 2011. Defrazeologizatsiyata kato iztochnik na leksikalni i semantichni inovatsii v savremenniya balgarski ezik. In *Ezikovedski izsledvaniya v chest na Prof. Siyka Spasova-Mihaylova*, 139–151. Sofia: Prof. M. Drinov Publishing House.

Cornilescu, Alexandra. 2001. Romanian nominalizations: Case and aspectual structure. *Journal of Linguistics* 37(3). 467–501.

Downing, Angela. 2014. *English grammar: A university course.* London & New York: Routledge.

Ermakova, Elena Nicolayevna, Natalia Nikolaevna Zolnikova, Guzel Chakhvarovna Faizullina, Milyausha Sakhretdinovna Khasanova & Tatiana Nikolaevna Khlyzova. 2015. Derivation and the derivational space in Phraseology as a problem of contemporary language development. *Mediterranean Journal of Social Sciences* 6. 335–340.

Fellbaum, Christiane, Anne Osherson & Peter E. Clark. 2009. Putting semantics into WordNet's "morphosemantic"links. In Zygmunt Vetulani & Hans Uszkoreit (eds.), *Human Language Technology. Challenges of the Information Society: Third Language and Technology Conference, LTC 2007. Revised selected papers*, 350–358. Poznan, Polland.

Grimshaw, Jane. 1990. *Argument structure.* Cambridge, MA: MIT Press.

Groza, Liviu. 2011. *Probleme de frazeologie: studii, articole, note.* Editura Universității din București.

Ion, Radu. 2007. *Word Sense Disambiguation Methods Applied to English and Romanian.* Romanian Academy, Bucharest dissertation.

Koeva, Svetla & Angel Genov. 2011. Bulgarian language processing chain. In *Proceedings of Integration of Multilingual Resources and Tools in Web Applications. Workshop in conjunction with GSCL*, vol. 26.

Kolkovska, Siya. 1993/1994. Slovoobrazuvane na imena za deystviya ot otimenni glagoli (s ogled na terminologiyata). *Balgarski ezik* 4. 478–480.

Kostova, Nadya. 2013. Novite imena za deystviya v balgarskiya ezik i tyahnoto leksikografsko predstavyane. In Diana Blagoeva, Sia Borisova Kolkovska &

Margarita Lishkova (eds.), *Problemi na neologiyata v slavyanskite ezitsi*, 77–108. Sofia: Prof. M. Drinov Publishing House.

Mărănduc, Cătălina. 2010. *Dicționar de expresii, locuțiuni și sintagme ale limbii române*. București: Corint.

Marouzeau, Jules. 1933. *Lexique delà terminologie linguistique*. Paris: Librairie orientaliste Paul Geuthener.

Nunberg, Geoffrey, Ivan A. Sag & Thomas Wasow. 1994. Idioms. *Language* 70(3). 491–538.

Piela, Agnieszka. 2007. Od Frazeologizmu do derywatu. *LingVaria* II 1(3). 41–48.

Pometkova, Yana. 2006. Spetsifikatsiya na konstruktsiite s otglagolno sashtestvitelno ime vav funktsiyata na podlog v nauchnata rech. In *Sbornik Yordanka Marinova. Izsledvaniya po sluchay neyniya sedemdesetgodishen yubiley*, 276–284. V. Tarnovo: Sv. sv. Kiril i Metodiy University Press.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. Longman.

Sag, Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In *Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002)*, 1–15.

Savary, Agata. 2008. Computational Inflection of MultiWord Units. A contrastive study of lexical approaches. *Linguistic Issues in Language Technology* 1(2). 1–53.

Stoyanova, Ivelina, Svetlozara Leseva & Svetla Koeva. 2015. *Partial syntactic analysis of Bulgarian*. Paisievi chetenia, Plovdiv, 30–31 October 2015.

Stoyanova, Ivelina & Maria Todorova. 2014. Razrabotvane na rechnitsi ot sastavni edinitsi za balgarski. In Svetla Koeva & Diana Blagoeva (eds.), *Ezikovi resursi i tehnologii za balgarski ezik*, 185–201. Sofia: Prof. M. Drinov Publishing House.