

# **Noun-Verb Derivation in the Bulgarian and the Romanian WordNet – A Comparative Approach**

**Ekaterina Tarpomanova, Svetlozara Leseva, Maria Todorova,  
Tsvetana Dimitrova, Borislav Rizov**

Institute for Bulgarian Language – Bulgarian Academy of Sciences  
{katja, zarka, maria, cvetana, boby}@dcl.bas.bg

**Verginica Barbu Mititelu, Elena Irimia**

Research Institute for Artificial Intelligence – Romanian Academy  
{vergi, elena}@racai.ro

## **Abstract**

Romanian and Bulgarian are Balkan languages with rich derivational morphology that, if introduced into their respective wordnets, can aid broadening of the wordnet content and the possible NLP applications. In this paper we present a joint work on introducing derivation into the Bulgarian and the Romanian WordNets, BulNet and RoWordNet, respectively, by identifying and subsequently labelling the derivationally and semantically related noun-verb pairs. Our research aims at providing a framework for a comparative study on derivation in the two languages and offering training material for the automatic identification and assignment of derivational and morphosemantic relations needed in various applications.

## **1. Introduction**

Wordnet enrichment by linking synsets via semantically labelled derivational relations (called morphosemantic relations) has been reported for Turkish (Bilgin et al., 2004), Czech (Pala and Hlaváčková, 2007), Serbian (Koeva et al., 2008), Polish (Piasecki et al., 2009), Romanian (Barbu Mititelu, 2012) and Bulgarian (Dimitrova et al., 2014; Koeva, 2008; Stoyanova et al., 2013), among others. Depending on the derivational specificities of the language and / or the methodology adopted, different, possibly overlapping sets of morphosemantic relations have been identified and implemented in the different wordnets.

In this work we consider the morphosemantic relations encoded in the Princeton WordNet (PWN) (Fellbaum et al., 2009) as a stand-off file, which have been transferred automatically in BulNet and RoWordNet. These semantic links are established between literals “that are similar in meaning and where one word is derived from the other by means of a morphological affix” (Fellbaum et al., 2009). Although these relations are morphologically expressed on particular pairs of lexemes (literals) in English (and possibly in other languages), they also hold between the synsets to which these literals belong, given the semantic dimension of the relation.

The PWN morphosemantic links were automatically transferred to the Bulgarian and the Romanian WordNet provided that both synsets that were members of a relation were present. Afterwards, the teams working on the two wordnets performed automatic extraction of literal pairs and derivational models from the morphosemantically related synsets, followed by manual validation of the pair members.

The goal of this paper is to summarise the findings of our joint work with a view to proposing a framework for the automatic discovery of derivational relations and the automatic assignment of morphosemantic relations which makes use of the rich inventory of derivational patterns of the languages under study. A further objective is to implement these linguistic generalisations in applications that benefit from the existence of such wordnet relations.

## 2. Derivational Morphology of Bulgarian and Romanian

Bulgarian and Romanian show great similarities in derivational morphology due to the common Indo-European inheritance, and to the interaction of the two languages in the Balkan Sprachbund. In both languages suffixation is the most productive means of word formation, but also the most complicated: one or more suffixes may be added to a stem, or a suffix may be substituted for another; suffixation may or may not change the part of speech of a word, while prefixation (usually) does not change the part of speech. In Bulgarian prefixes have an important role for verb-to-verb derivation as they may involve change of verbal aspect. The two derivation processes, suffixation and prefixation, may occur simultaneously to form a new word (parasyntetic derivation) in both languages.

Conversion is a disputable notion in the traditional linguistic descriptions of both Romanian and Bulgarian. According to the Romanian tradition it is distinct from derivation and always implies homonymy. In the Bulgarian literature conversion is usually interpreted in a broader sense as a process of word formation in which the written forms of two words in a derivational pair differ only by their inflectional markers: (*rabotya* (“to work”) – *rabota* (“work”). Formation of deverbal nouns by removing the thematic vowel and the inflection of a verb without adding a suffix to the noun, such as in (*nanizha* (“to string”) – *naniz* (“string”), are called zero suffixation. In the Bulgarian data discussed below zero suffixation is subsumed under conversion and labelled accordingly. Cases of conversion in Romanian are not discussed in this paper, as it does not serve creating verbs (in their infinitive form) from nouns or, vice versa, nouns from infinitives.

Word formation in Bulgarian and Romanian often involves vowel or consonant alternations, or both. Most of the alternations are phonetically motivated (metaphony, palatalisation), others mark grammatical forms (apophony in Bulgarian). Because of their irregular behaviour, while phonetic alternations often impede the automatic detection of derivational pairs.

## 3. The Nature of Morphosemantic Relations

The morphosemantic relations encoded in the Princeton WordNet and transferred to BulNet and RoWordNet, usually denote a relation between a predicate and a participant in its semantic representation. In consequence, most of the relations correspond to thematic roles in the representation of the respective derivationally related predicates: Agent, Event, State, Result, Undergoer, Property, Vehicle, Destination, Material, Body-part, Cause, Instrument, Location, By-means-of, (cf. Fellbaum et al. (2009). The only exception is the relation Event, which links verbs to deverbal nouns denoting the same event.

Given the semantic dimension of morphosemantic relations, the semantic label associated with such a relation holds between synsets and is transferable across languages, even though the morphological relation (between literals) needs not be expressed (Koeva, 2008). Besides, even in a language in which a morphosemantic relation or an instance of such a relation has morphological expression, its specific semantics may be derivationally expressed only by certain literal pairs in the respective synsets. For example, consider the synsets *write*, *compose*, *pen*, *indite* – *writer*, *author*, where only *write* and *writer* are derivationally related, as opposed to *cry*, *weep* – *weeper*, *crier*, where both *cry* and *crier* and *weep* and *weeper* are morphologically related.

The morphosemantic relations may be expressed through direct or non-direct derivation (obtained through different derivation paths). Consider the Bulgarian verb *analiziram* (“to analyse”) and the agentive noun *analizator* (“analyser”), each derived independently from the noun *analiz* (“analysis”). In this case, two pairs are linked via morphosemantic relations – *analizator* (“analyser”) – *analiziram* (“to analyse”) (Agent) and *analiz* (“analysis”) – *analiziram* (“to analyse”) (Event). Also, the derivation path may involve more than one operation, such as the two-step derivation of the nouns from the corresponding verbs in: Bulgarian *kova* (“to forge”) > *kovach* (“blacksmith”) > *kovachnitsa* (“a forge”), and Romanian *topi* (“to melt”) > *topitor* (“melter”) > *topitorie* (“foundry”).

## 4. Methods of Assigning Morphosemantic Relations

In the framework of our research we assigned the morphosemantic relations from the PWN stand-off file to the lexicalised synsets in BulNet and RoWordNet and checked the literals in the relevant pairs of

synsets to establish the derivationally related pairs of literals. We used different approaches for identification of the derivational subsets in BulNet and Ro-WordNet as each team had started working independently, with different resources at their disposal and with different aims. The particularities of the methods used for each language are described in Dimitrova et al. (2014) and Barbu Mititelu (2012).

The derivationally related pairs of literals were verified manually. In Romanian 2,767 pairs were found altogether, with the following distribution: 2,429 cases of suffixation, 318 cases of verbal suffixation equivalent to the Bulgarian conversion and 20 cases of parasynthetic derivation. In Bulgarian 6,135 pairs were found, as follows: 4,590 cases of suffixation, 930 cases of substitution of a noun suffix for a verb suffix or vice versa, 433 cases of conversion, 139 cases of prefixation, 12 cases of parasynthetic derivation, 31 cases of non-transparent derivation.

## 5. Expression of Morphosemantic Relations through Derivational Patterns

Morphosemantic relations are expressed both at the form level and at the meaning level, which makes their cross-lingual analyses informative and useful in different NLP tasks. Table 1 shows the derivational patterns associated with each morphosemantic relation in Bulgarian (BG) and Romanian (RO), with the number of occurrences found in the respective database in brackets. The verbal suffixes are written in parentheses and preceded by a dash: for example, – (-a) among the Romanian suffixes. “Total” refers to the total number of affixes/derivational patterns with the respective semantic label.

Semantic label	BG Affixes (number of occurrences)	RO Affixes (number of occurrences)
Agent	-tel (169), -ach/-yach (128), -(n)ik (87), -sht (83), -or/-yor (44), -tor (-tor/-tyor/-ator/-itor/-ityor) (42), -ets (33), – (-iram/-iziram) (22), -ar/-yar (19), – (-stvam) (19), -ist (14), -ant/-ent (13), -ne (13), -dzhiya/-chiya (11), -er/-ier/-ăr (11), conversion (10), -in (7), -l (7), -chik (7), -n (6), -nie (6), -ovach (6), -ko (5), -ak (2), -at (4), -tsiya (-tsiya/-atsiya/-itsiya/-ziya/-siya) (5), -stvo (4), (-uvam) (4), -entsiya (2), -ir (2), -itsa (2), -telka (2), -lo (2), -ba (1), -ezh (1), -ek (1), -ik (1), -ka (1), -lyo (1), -(n)itsa (1), -m (1), – -ot (1), -t (1), -yaga (1), -yay (1) <b>Total 45</b>	-(ă)tor (176), -t/-s (31), – (-i) (20), – (-a) (10), – (-iza) (4), -re (7), – (în- + -a) (1), -ar (5), -ant (7), -or (5), -ier (3), -t + -el (1), -[ăi]cios (3), -ist (2), -ăros (1), -u (1), -ici (1), -ăreț (5), -aș (5), -(ă)toare (4), -ură (1), -(ă)tor+-easă (2), -aci (1), -angiu (1), -nic (1), – (în- + -i) (1), – (-ui) (1), -ație (2), -ațiune (1), -aș (1) <b>Total 31</b>
Body-part	conversion (4), -ka (1), -nie (1) <b>Total 3</b>	(-a) (1) <b>Total 1</b>
By-means-of	-ne (64), -nie (53), conversion (45), -ka (33), – (-iram, -iziram) (29), -tsiya (-tsiya/-atsiya/-itsiya/-ziya/-siya) (28), -(n)ost/-est (8), -tor (-tor/-tyor/-ator/-itor/-ityor) (8), -ina (8), – (-vam/-avam/uvam) (7), -tel (6), -lo (5), -no (5), -ets (3), -ie (3), -iya (3), -lka (3), -ovka (3), – (-(n)icha) (3), – (-na) (3), -stvo (3), -at (2), -entsiya (2), -izăm (2), -achka (1), -ba (1), -er/-ier/-ăr (1), -or/ -yor (1), – (-(y)asam/-(y)osam) (1), – (-stvam) (1) <b>Total 30</b>	-re (98), – (-a) (37), – (-i) (12), -(ă)tor (10), – (-ifica) (6), -ație (8), -t/-s (6), -(ă)tură (6), -eală (4), -ant (2), -or (3), -(ă)toare (4), – (-î) (2), – (în- + -a) (6), -ăciune (2), -tor + -ie (1), -t + -ie (1), – (-ui) (3), – (-iza) (3), -ment (3), -ie (2), -ală (1), – (-ia) (2), – (-(ur)i) (1) <b>Total 25</b>
Destination	– (-ifitsiram) (2) <b>Total 1</b>	– (-a) (1), -ar (1), -ant (1) <b>Total 3</b>

Continued

Table 1 – Continued

Semantic label	BG Affixes (number of occurrences)	RO Affixes (number of occurrences)
Event	-ne (2372), conversion (418), -nie (353), -tsiya (-tsiya/-atsiya/-itsiya/-ziya/-siya) (325), -ka (83), - (-iram) (66), - (-vam/-avam/uvam) (51), -ie (48), -stvo (42), -(n)ost/-est (29), -iya (23), -ăk (21), -ezh (18), -ba (15), -(n)itsa (17), - (-na) (9), -ek (8), -ovka (7), - (-(y)asam/- (y)osam) (7), -n (6), -azh (5), -tva (5), -ina (5), -at (4), -nya (4), -entsiya (4), -(n)ik (3), -zăm (3), -ar (3), -da (2), -na (2), -ot (2), -s (2), -t (2), -al (1), -ans (1), -et (1), -ishte (1), -ota (1), -otevitsa (1), -tba (1), -tel (1), -ulka (1), -h (1) <b>Total 45</b>	-re (1174), -t/-s (112), -ație (110), -(-a) (97), -(ă)tură (48), -(e)ală (44), - (-i) (20), -ment (9), -or (1), -ător/ătoare (4), -ăt (1), -et (5), - (-î) (1), -ă (1), - (-ări) (1), -aci (1), -(ă)ciune (3), - (-ui) (6), -ie (8), - (-ifica) (1), -ațiune (1), -iș (1), -uș (1), - (-ia) (1), - (-ua) (1), - (-âi) (1), -e (4), -aj (3), -[a/e/i]nță (4), - (-iza) (3), -iune (1), - (-i)ona (5), -erie (1), - (-ăi) (1) <b>Total 34</b>
Instrument	conversion (26), -tor (-tor/ -tyor/ -ator/ -itor/ -ityor) (15), - (-iram/-iziram) (15), -tel (13), -er/-ier/-ăr (6), -ka (9), -ie (4), -or/ -yor (3), - (n)ik (2), -ach/-yach (2), -lka (2), -l (2), -nie (2), -(n)itsa (1), -ik (1), -la (1), -nya (1), -ovach (1) <b>Total 18</b>	-(ă)tor (21), - (-a) (10), (-iza) (2), -re (1), -t/-s (1), - (-i) (2), -(ă/i)toare (6), -tură (1), - (-î) (1), - (în- + -a) (1), pre- + -ător (1), - (-ui) (2), - (-ia) (1) <b>Total 13</b>
Location	conversion (25), - (-iram) (9), - (-vam/-avam/uvam) (7), -ishte (6), -iya (2), -ne (2), -(n)itsa (1), -ing (1), -ka (1) <b>Total 9</b>	-re (6), - (-iza) (1), - (în- + -i) (1), -ment (1) <b>Total 4</b>
Material	-tel (17), -tor (-tor/ -tyor/ -ator/ -itor/ -ityor) (12), - (-iram/-iziram) (12), -tsiya (-tsiya/-atsiya/-itsiya/-ziya/-siya) (6), - (-ifitsiram) (5), conversion (2), -ka (2), -lka (2), -nie (2), -ant/-ent (2), -at (2), -(n)ik (1), -atoar (1), -ezh (1), -er/-ier/-ăr (1), -ivo (1), - (-osam) (1) <b>Total 17</b>	- (-iza) (2), - (-a) (5), -ant (1), -(ă)tor (3), - (-i) (1), -tură (1), - (-ona) (1) <b>Total 7</b>
Property	-nie (28), -(n)ost/-est (24), conversion (23), -ne (15), -tsiya (-tsiya/-atsiya/-itsiya/-ziya/-siya) (4), -ie (4), -ba (3), -entnost (2), -entsiya (2), -iya (2), - (-iram) (4), - (-(u/o)vam) (4), -ka (1), -ota (1), -stvo (1) <b>Total 15</b>	-re (32), - (-a) (6), - (-i) (2), - (în- + -a) (1), - (-ui) (2), -ment (2) <b>Total 6</b>
Result	conversion (126), - (-iram/-iziram) (53), -ne (46), -tsiya (-tsiya/-atsiya/-itsiya/-ziya/-siya) (40), -nie (32), -ka (19), -(n)ost/-est (18), -at (10), -no (9), - (-asam/yasam) (4), -ets (4), -iya (4), -n (4), -ie (3), -ina (3), -tor (-tor/ -tyor/ -ator/ -itor/ -ityor) (3), -ezh (2), -(n)itsa (2), - (-vam) (2), -azh (1), -al (1), -ar (1), -ba (1), -e (1), -ek (1), -eriya (1), -iy (1), -ing (1), -l (1), -ma (1), -ment (1), -ovka (1), -ura (1), -ăk (1) <b>Total 34</b>	-re (83), - (-a) (26), - (-iza) (13), - (-ifica) (15), -t/-s (11), - (-i) (6), -eală (4), - (în- + -a) (2), -tură (10), -et (2), - (în- + -i) (1), - (-ui) (2), - (-ifia) (2), -ație (1), -ere (1), - (-ona) (1), -ment (3), - (-ua) (1) <b>Total 18</b>
State	-ne (68), -nie (47), conversion (37), -(n)ost/-est (30), -tsiya (-tsiya/-atsiya/-itsiya/-ziya/-siya) (15), -ie (7), -stvo (6), - (-iram/-iziram) (5), -ist (2), -iya (2), - (-osam) (2), - (-uvam) (2), -ka (2), -ika (1), -ota (1), -ăk (1) <b>Total 16</b>	-re (94), - (-a) (5), -(e)ală (3), supra- + -re (1), - (-i) (2), -t/-s (1), - (în- + -a) (1), - (în- + -i) (1), -ie (1), -ment (1), -e (1) <b>Total 10</b>

Continued

Table 1 – Continued

Semantic label	BG Affixes (number of occurrences)	RO Affixes (number of occurrences)
Undergoer	conversion (65), -ne (28), -nie (23), – (-iram/iziram) (18), -n (11), -tsiya (-tsiya/-atsiya/-itsiya/-ziya/-siya) (10), -(n)ost/-est (9), -at (9), -ka (8), -ba (5), – (-vam/-avam/uvam) (5), -(n)ik (4), -ie (4), -ina (4), -sht (4), -ach/-yach (2), -ek (2), -da (2), -m (2), – (-ifitsiram) (2), – (-(y)asam/-(y)osam/-(d)isam) (2), -ant/-ent (1), -el (1), -entsiya (1), -ivo (1), -iya (1), -ma (1), -n (1), -nya (1), -och (1), -tva (1), -tel (1), -t (1), -āk (1), – (-stvam) (1) <b>Total 35</b>	-re (27), – (-a) (23), -t/-s (15), -ant (1), – (-i) (2), – (-ui) (4), – (-iza) (2), – (-ifica) (1), -ație (1), -ment (1) <b>Total 10</b>
Uses	– (-iram/iziram) (45), -ne (25), conversion (22), -nie (20), -tsiya (-tsiya/-atsiya/-itsiya/-ziya/-siya) (13), -ka (10), -lo (7), -stvo (7), -iya (5), – (-vam/-avam/uvam) (5), -at (4), -ie (3), – (-ifitsiram) (3), -et (2), -iy (2), -ina (2), -lka (2), -ovka (2), -ura (2), – (-(y)asam/-(y)osam/-(d)isam) (4), -(n)ost/-est (1), -ant/-ent (1), -ezh (1), -er/-ier/-ăr (1), -tel (1), -tor (-tor/-tyor/-ator/-itor/-ityor) (1) <b>Total 26</b>	-re (24), – (-a) (23), – (-iza) (4), -t/-s (1), -eală (1), -tor (1), – (-i) (2), -tură (1), – (în- + -a) (1), – (în- + -i) (1), – (-ui) (2), – (-ifica) (1), -ație (1), -ment (3) <b>Total 14</b>
Vehicle	-ach/-yach (1), -er/-ier/-ăr (1), -ovach (1) <b>Total 3</b>	– (-a) (1), -or (1), -er (1) <b>Total 3</b>

Table 1: Derivational affixes in Bulgarian and Romanian wordnets associated with semantic labels

As a consequence of the fact that prefixes normally do not change the part of speech of the stems they are attached to and that we focus on noun-verb pairs, the affixes discussed here are almost exclusively suffixes (with the exception of parasynthetic derivational patterns).

The statistics (see Table 1) show that more affixes are found in the Bulgarian data – 252 noun suffixes (in each of their senses), 38 verbal ones, and 12 cases of conversion. In Romanian there are 91 noun suffixes and 45 verbal ones (plus 26 cases of verbal derivation that are equivalent to conversion in Bulgarian). Besides the quantitative difference between the pairs subject to analysis here (4,590 pairs in Bulgarian and 2,429 pairs in Romanian), the difference in the number of the suffix senses can also be explained in terms of the specifics of the derivational morphology of the two languages. As a Slavic language, Bulgarian has a rich inventory of noun suffixes that outnumber considerably the corresponding Romanian suffixes: compare the three most productive Bulgarian suffixes with a primary agentive reading (-tel, -ach/-yach, and -(n)ik) vs. one such suffix in Romanian (-(ă)tor). Additionally, Bulgarian has adopted many Romance suffixes through the active borrowing of Romance words, so that the Romanian -(ă)tor has an exact equivalent in Bulgarian, the suffix -tor. The verbal aspect in Bulgarian is another linguistic reason for the greater diversity of patterns as both imperfective and perfective stems may be productive in verb–noun derivation and some noun suffixes may attach preferentially or exclusively to either an imperfective or a perfective verb stem, giving rise to different derived words; for example both -ne, which combines only with imperfective stems, and -nie, which usually selects perfective stems, correspond to -re in Romanian.

Verbal patterns involve the attachment or removal of one verbal suffix. The noun suffix representative for the respective relation may remain ‘hidden’ as it is present both in the noun and the verb: see the Romanian pair *călători/călător* (“to travel/traveller”), which involves the attachment of the verbal suffix -i to the base noun (the agentive suffix -tor is considered part of the base noun). There are examples, such as *ucenici/ucenic* (“to apprentice/apprentice”) or *grădinări/grădinar* (“to garden/gardener”) in which the verbs are formed from suffixed Slavic (Bulgarian) loan nouns following a Romanian verbal pattern.

Despite the difference in the number of suffixes, the two languages show similarity in the derivational productivity of the morphosemantic relations. The relations with the highest diversity of derivational patterns are Agent (expressed by 45 derivational patterns in Bulgarian and 31 in Romanian) and Event (45 patterns in Bulgarian, 37 in Romanian). They also cumulate the greatest number of occurrences.

Agentive suffixes in Bulgarian are both domestic and loaned, with prevalence of the former, such as *-tel*, *-ach/-yach*, *-nik*, which are also the most productive. Another productive pattern is represented by the suffix for the present active participle (*-sh*) substantivised to express Agent. Another frequent derivational pattern is formed by a noun, usually loaned, that cannot be morphologically segmented and is verbified by adding the suffix *-iram/-iziram* (*tip* (“type”) > *tipiziram* (“to type, to typecast”). To express Agent, Bulgarian uses also suffixes loaned from the Romance languages, mainly French (*-tyor*), from Turkish (*-dzhiyal-chiya*), Russian (*-chik*), and other languages. The Romanian agentive affixes are of various origins (Latin, Slavic, Romance, Hungarian and domestic), with a prevalence of Latin and Romance affixes, among which the most productive one can also be found: *-tor*. Quite frequently the participle (or even the gerund) is used to denote an agent. In both languages agentive nouns are usually derived from verbs, but there is a considerable number of instances (one fourth of the total number of affixes for Romanian) where verbs are derived from agentive nouns.

In Bulgarian the relation Event is most typically expressed by the suffix *-ne*, whose occurrences outnumber the sum of the occurrences of all the other suffixes. The suffix *-nie*, traditionally associated with a resultative meaning, was found to be very productive in expressing Event, too. Conversion is used to form 418 derivational verb–noun pairs in Bulgarian. Event is expressed mostly by domestic suffixes, except for *-tsiya* and its variants, which is among the productive patterns. The suffixes *-ie*, *-stvo*, *-(n)ost/-est*, whose typical meaning is associated with Event, are ranked among patterns with medium productivity. The prevalent eventive suffixes in Romanian are of Latin or Romance origin. The most productive one is the old infinitive formant *-re* reinterpreted as a suffix for deverbal nouns. In Romanian the participle used as a noun is also a productive means of denoting events. The cases where the verb is derived from the noun denoting Event are quite numerous (one third of the affixes).

Other relations expressed by a variety of derivational patterns in Bulgarian and Romanian are By-means-of, Instrument, Result, State, Undergoer, and Uses. Disparity between the two languages is observed in the relations Material and Property. The former is expressed by 17 derivational patterns in Bulgarian vs. 7 in Romanian, and the latter by 15 patterns in Bulgarian vs. 6 in Romanian.

The relations represented by a small number of occurrences and derivational patterns are Location, Destination, Body-part, and Vehicle. Due to the lack of evidence, they are of little importance for the general analysis, yet some observations can be made. In Bulgarian Vehicle is expressed by the loaned suffix *-erl/-ierl-ăr*, which is the equivalent of *-er* in Romanian, and the Slavic suffix *-ach/-yach* and its variant *-ovach*. Two typical Location suffixes occur in the Bulgarian data as well – *-ishte* and *-nitsa*.

## 6. Derivational Patterns. The Nature and Properties of Derivational Relations

In this study we analyse mainly noun suffixes as they are the bearers of the semantics of the relations under discussion. Verb suffixes in both languages have mostly grammatical functions. In Bulgarian, typically, they either imperfectivise a perfective verb, or perfectivise an imperfective verb, or are used to derive a verb from a word pertaining to a different part of speech. The verb suffixes occurring in the Romanian data set always create verbs from other parts of speech. Besides the established noun suffixes, we look at certain participial and adjectival suffixes as participles and adjectives can be substantivised.

The data below are based on noun and verb synsets with equivalents in the PWN. Therefore, the results are not conclusive either with respect to the language system or to the parts of speech involved.

In the data we have analysed (Table 2) there is a large number of monosemous affixes: 32 for Bulgarian and 45 for Romanian, associated mostly with the labels Event (18 in Romanian: *-erie*, *-anță*, *-aj*, etc., and 7 in Bulgarian: *-ulka*, *-tba*, *-otevitsa*, etc.), and Agent (18 in Romanian: *-aci*, *-angiu*, *-nic*, etc., and 13 in Bulgarian: *-chik*, *-ar/-yar*, *-chiyal-dhziya*, *-in*, *-lyo*, etc.). Several other relations – Material, Result, Undergoer, Property, State and Instrument in Bulgarian, and By-means-of, Instrument, State, Result and Vehicle in Romanian are represented by one or a couple of unambiguous suffixes.

Number of relations	1	2	3	4	5	6	7	8	9	10	11	12	13
Number of Bg suffixes	32	19	6	7	6	4	3	2	2	-	2	1	1
Number of Ro suffixes	45	7	5	2	4	4	1	3	1	2	-	-	-

Table 2: Number of Relations across suffixes

Polysemous suffixes are usually associated with clusters of relations with one of them being the default reading (estimated in terms of number of instances). For example, suffixes which primarily express the relation Agent can also express relations denoting inanimate agents and causes, such as Instrument, Material, By-means-of. The relations Vehicle and Body-part typically should also be included in this group, but the number of instances is too small so we defer judgement. The relation Uses, which denotes a function or a purpose, is also often expressed by agentive suffixes. In certain cases the same suffix denotes both Agent and Undergoer depending on whether the verb is unergative or unaccusative (Fellbaum et al. 2009), e.g., *demonstrant* (“demonstrator”), and *mutant* (“mutant”), respectively.

Another large part of the suffixes typically express relations such as Event, Result and / or other relations involving the process or result of an action, state or another kind of situation, such as State and Property. A relatively frequent relation associated with this type of suffixes is Undergoer, which in this case denotes patients. The relation Uses and By-means-of can also be expressed by event-like suffixes.

In Table 3 below, we show Bulgarian and Romanian suffixes primarily associated with the relations Agent and Event and their other senses expressed by the respective relations.

Language	Suffix	Default semantic value	Other semantic values
Bg	-tel	Agent (169)	Material (17), Instrument (13), By-means-of (6), Undergoer (1), Uses (1)
Bg	-tor	Agent (42)	Instrument (15), Material (12), By-means-of (8), Result (3), Uses (1)
Bg	-tsiya	Event (325)	Result (40), By-means-of (28), State (15), Uses (13), Undergoer (10), Material (6), Agent (5), Property (4)
Bg	-ne	Event (2372)	State (68), By-means-of (64), Result (46), Undergoer (28), Uses (25), Property (15), Agent (13), Location (2)
Bg	-nie	Event (353)	By-means-of (53), State (47), Result (32), Property (28), Undergoer (23), Uses (20), Agent (6), Instrument (2), Material (2), Body-part (1)
Ro	-tor	Agent (180)	Instrument (27), By-means-of (14), Event (3), Material (3), Uses (1)
Ro	-re	Event (1173)	By-means-of (98), State (94), Result (84), Property (32), Undergoer (27), Uses (24), Agent (7), Location (6), Instrument (1)
Ro	-ți(un)e	Event (111)	By-means-of (8), Agent (3), Undergoer (1), Result (1), Uses (1)
Ro	-t/s	Event (112)	Agent (30), Undergoer (15), Result (11), By-means-of (6), Instrument (1), State (1), Uses (1)
Ro	-(ă)tură	Event (48)	Result (10), By-means-of (6), Instrument (1), Material (1), Uses (1)

Table 3: Semantic labels for corresponding productive suffixes

The Agent reading is the default one for the Bulgarian suffixes *-tel*, *-tor*, *-(n)ik*, *-antlent*, *-arl-yar*, *-achl-yach*, *-erl-ierl-ăr*, *-ets*, *-ist*, *-orl-yor*, *-dzhiya*, *-ak*, and the Romanian suffixes *-tor*, *-ant*, *-ar*, *-or*. The Bulgarian suffixes *-ne*, *-nie*, *-ba*, *-ezh*, *-ie*, *-iya*, *-ka*, *-stvo*, *-tsiya*, *-ăk*, and the Romanian *-re*, *-ație*, *-tură*, *-eală*, *-t/s* are most often associated with Event.

Event-type suffixes may adopt agentive meanings through metaphorical extension of an activity to a

body of people who are responsible for or carry out the activity, e.g., *-stvo*, *-nie* and *-tsiya* in *răkovodstvo*, *upravlenie*, and *administratsiya* (all meaning “administration” in the sense of a governing body). The Romanian *administrație* (“administration”) and *organizație* (“organization”) are similar examples. The reverse process – the extension of the meaning of agentive suffixes to event-type meanings – is rarely observed in Bulgarian (*-ik* in *plesnik* “a smack, smacking”).

A productive pattern in Bulgarian involves participle substantivisation. Active participles usually express Agent (*pregovaryasht* (“negotiator”), *otselyal* (“survivor”)), while passive participles are mostly associated with Undergoer (*intervyuiran* (“interviewee”)), Result (*izgoreno* (“a burn, burn wound”)), Event (*razlyano* (“a spill, spilled liquid”) and By-means-of (*zakārpeno* (“mend, patch, darn”)). The agentive reading of the passive participle is quite untypical for Bulgarian and the 6 instances registered in our data are due either to reflexive/middle interpretation, as in *pristrastyia selpristrasten* (“to addict/an addict”), *uhilya seluhilen* (“grin/grinner”), or to incorrect relation assignment in PWN (Agent instead of Undergoer), as in *zapodozralzapodozryan* (“to suspect/a suspect”). Since Romanian does not distinguish between active and passive participles, the participle (ending in *-t/s*) descending from the Latin perfect passive participle has assumed functions typical for both the active and the passive participles in Bulgarian. This is reflected in the morphosemantic relations denoted by Romanian participles, which to the exception of several single instances cover the same relations as in Bulgarian: Event (*plagiat* (“plagiarization”)), Agent (*conjurat* (“conspirator”)), Undergoer (*intervievat* (“interviewee”)), Result (*bubuit* (“thunder”)) and By-means-of (*certificat* (“a certificate”)).

Table 3 shows the distribution of several suffixes related by origin or function in the two languages. For example, the deverbal suffixes *-ne* in Bulgarian and *-re* in Romanian are functional equivalents and express very similar sets of morphosemantic relations. Their distribution in terms of frequency across relations differ, with *-ne* showing a stronger preference for the Event reading than *-re*. The Event-type suffix *-tsiya* in Bulgarian, which has common origin and meaning with *-ți(un)e* in Romanian, expresses a broader range of relations and has more even distribution across relations than its Romanian counterpart. The Latin/Romance agentive suffix *-tor* has developed identical meanings in Bulgarian and Romanian. The only differences that we found are 3 instances of Result in Bulgarian expressed by the literals *emulgator* and *emulsifikator* (“emulsifier”) whose Romanian counterpart is formed by the suffix *-ant* (*emulsifiant*), and 3 instances of Event in Romanian.

## 7. Applications and Further Work

We plan to expand our work by further identifying derivationally related literals and semantically related synsets that have not been discovered so far due to imperfections in the recognition algorithms or because the derivationally related pairs are not morphologically related in English.

The benefit of adding new relations is two-fold – it will enable us to increase the connectivity of the wordnet synsets on the one hand, and to establish procedures for semi-automatic expansion with new synsets, on the other. A possible source of new relations that is worth exploring are other wordnets that have implemented (possibly other sets of) morphosemantic relations. The ones implemented by us in the Bulgarian and the Romanian WordNet can also be transferred to wordnets for other languages.

In the context of automatic labelling of morphosemantic relations this study can be helpful in the task of disambiguating the suffixes of words from the same synset (in a monolingual setting) or from corresponding synsets (in a multilingual setting) on the basis of their senses and semantic restrictions.

From the applications perspective, marking morphosemantic relations explicitly in individual and aligned wordnets can prove useful in text processing and information retrieval both in a monolingual and in a multilingual context. For instance, Barbu Mititelu (2013) showed how marking derivational relations in RoWordNet can help improve a task of Question Answering that makes use of lexical links. Extending that experiment and imagining a cross-language Question Answering system, the resource created by us can help identify and subsequently transfer relations between words that are morphosemantically unrelated in one language, but are related in another.



## Acknowledgements

The present paper was prepared within the project *Integrating New Practices and Knowledge in Undergraduate and Graduate Courses in Computational Linguistics* (BG051PO001-3.3.06-0022) implemented with the financial support of the Human Resources Development Operational Programme 2007 - 2013 co-financed by the European Social Fund of the European Union. The authors take full responsibility for the content of the present paper.

## References

- Barbu Mititelu, V. (2012). Adding Morpho-semantic Relations to the Romanian Wordnet. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012)*, pages 2596–2601.
- Barbu Mititelu, V. (2013). Increasing the Effectiveness of the Romanian Wordnet in NLP Applications. *Computer Science Journal of Moldova*, 21(3):320–331.
- Bilgin, O., Cetinoglu, O., and Oflazer, K. (2004). Morphosemantic Relations in and across Wordnets – A Study Based on Turkish. In *Proceedings of the Second Global Wordnet Conference (GWC 2004)*, pages 60–66.
- Dimitrova, T., Tarpomanova, E., and Rizov, B. (2014). Coping with Derivation in the Bulgarian WordNet. In *Proceedings of the Seventh Global Wordnet Conference (GWC 2014)*, pages 109–117.
- Fellbaum, C., Osherson, A., and Clark, P. (2009). Putting Semantics into WordNet’s “Morphosemantic” Links. In *Proceedings of the Third Language and Technology Conference, Poznan, Poland*. [Reprinted in: *Responding to Information Society Challenges: New Advances in Human Language Technologies. Springer Lecture Notes in Informatics*], volume 5603, pages 350–358.
- Koeva, S., Krstev, C., and Vitas, D. (2008). Morpho-semantic Relations in Wordnet – A Case Study for Two Slavic Languages. In *Proceedings of the Fourth Global WordNet Conference (GWC 2008)*, pages 239–254.
- Koeva, S. (2008). Derivational and Morphosemantic Relations in Bulgarian Wordnet. *Intelligent Information Systems*, pages 359–368.
- Pala, K. and Hlaváčková, D. (2007). Derivational Relations in Czech Wordnet. In *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing*, pages 75–81.
- Piasecki, M., Szpakowicz, S., and Broda, B. (2009). *A Wordnet from the Ground up*. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej.
- Stoyanova, I., Koeva, S., and Leseva, S. (2013). WordNet-based Cross-language Identification of Semantic Relations. In *Proceedings of the 4th Biennial International Workshop on Balto-Slavonic Natural Language Processing*, pages 119–128.