Semantic Analysis of Verb-Noun Derivation in Princeton WordNet

Verginica Barbu Mititelu
Research Institute for Artificial Intelligence
Romanian Academy
vergi@racai.ro

Svetlozara Leseva
Institute for Bulgarian Language
Bulgarian Academy of Sciences
zarka@dcl.bas.bg

Ivelina Stoyanova
Institute for Bulgarian Language
Bulgarian Academy of Sciences
iva@dcl.bas.bg

Abstract

We present here the results of a morphosemantic analysis of the verb-noun pairs in the Princeton WordNet as reflected in the standoff file containing pairs annotated with a set of 14 semantic relations. We have automatically distinguished between zero-derivation and affixal derivation in the data and identified the affixes and manually checked the results. The data show that for each semantic relation an affix prevails in creating new words, although we cannot talk about their specificity with respect to such a relation. Moreover, certain pairs of verb-noun semantic primes are better represented for each semantic relation, and some semantic clusters (in the form of WordNet subtrees) take shape as a result. We thus employ a large-scale data-driven linguistically motivated analysis afforded by the rich derivational and morphosemantic description in WordNet to the end of capturing finer regularities in the process of derivation as represented in the semantic properties of the words involved and as reflected in the structure of the lexicon.

1 Introduction

The focus of this paper is the study of the derivational patterns between English verb-noun pairs. The perspective adopted is semantic, with two correlated aims: identifying semantic regularities involved in derivation (i.e., semantic relations between the members of a derivational pair) and establishing the semantic conditions in which it occurs (the semantic classes to which the nouns and verbs belong expressed in terms of semantic primitives, or primes (Miller et al., 1990), i.e. language-independent semantic classes).

The study is based on the semantically-labeled derivational pairs identified in the Princeton WordNet (PWN) (Fellbaum, 1998) presented in a standoff file (Fellbaum et al., 2009).

An in-depth study of the regularities will help paint a more detailed picture of the distribution of derivation, be it affixal or zero derivation. Based on the perspective adopted in this work, conclusions will highlight similarities and differences between the so-called "zero" morpheme conversion and the affixes involved in derivation.

Throughout this paper we use the terms *conversion* and *zero derivation* interchangeably to refer to the process of creating new words without any lexical material or, in other words, with the zero affix; *affixal derivation* refers to the morphological process involving the attachment of a non-zero affix to a base form to create a new word; *derivation* is their hypernym, a general term designating the morphological process whereby new words are created either involving affixes or not[1].

2 Related work

Zero derivation has been widely debated and discussed by linguists. Its high productivity in English specifically for creating verbs from words of other parts of speech was noted by researchers (Plag, 1999), as well as the fact that derivatives with overt suffixes are a subset of the possible meanings of converted verbs (Plag, 1999). A variety of meanings involved in conversion was noticed by Clark and Clark (1979), by Cetnarowska (1993), Plag (1999), Lieber (2004), Bauer et al. (2013), to mention but a few. Criteria for establishing the direction of conversion were identified and discussed (Bauer et al., 2013): semantic dependency, frequency, order of coining in the lan-

[1] Besides *conversion* or *zero derivation* and *affixal derivation*, another hyponym of *derivation* is *backformation*, which involves subtracting an affix from a word in order to create a new one.

guage. All studies consider words as entities involved in the process; however, our analysis here is done at the word sense level and is facilitated by the organizing principle of PWN, which takes the word sense as the minimal analysis unit.

3 Morphosemantic Relations in WordNet

The standoff file[2] consists of 16,812 [3] unique verb-noun pairs of which 53.57% represent patterns of affixal derivation and 46.43% are conversions. Each pair is annotated with a morphosemantic relation (out of a set of 14 such relations).

Although not explicitly defined, the meaning of these relations may be inferred from the observation of the data. Below, we sketch out a revised version of a description of these relations proposed by Koeva et al. (2016). Many of the relations have a more or less direct correspondence in the domain of thematic relations; in fact, in the lexicalist approaches in the Generative grammar of the 1980s, V-to-N derivation was accounted for as theta-role assignment from the predicate argument structure of the verb within the word structure of the noun (Müller, 2016), but this is not a one-to-one correspondence as the overview below shows.

3.1 Describing Morphosemantic Relations

An **Agent** is a person (noun.person), a social entity, such as organisations (noun.group), an animal (noun.animal) or a plant (noun.plant) that is capable of acting so as to bring about a result.

An **Instrument** is either a concrete, usually man-made object (noun.artifact), or something abstract, such as a noun with the prime noun.communication, e.g. *debug*:1 – *debugger*:1 ('a program that helps in locating and correcting programming errors') or noun.cognition, e.g. *stem*:4 – *stemmer*:3 ('an algorithm for removing inflectional and derivational endings in order to reduce word forms to a common stem'). It is always implied that the Instrument acts under the volition of an Agent.

A **Body-part** is an inalienable part of the body of an Agent expressed by nouns with the prime noun.body (rarely noun.animal or noun.plant).

The relation **Material** may denote a type of inanimate cause (Fellbaum et al., 2009) – substances that may bring about a certain effect: e.g. *inhibit*:2 – *inhibitor*:1 ('a substance that retards or stops an activity'). Besides noun.substance, noun.artifacts (synthetic substances or products) also qualify for the relation, e.g. *depilate*:1 – *depilatory*:2 ('hair removal cosmetics'). In addition, the relation may also express function or purpose, as in *sweeten*:1 – *sweetener*:1 ('something added to foods to make them taste sweeter').

The relation **Vehicle** represents a subclass of artifacts (means of transportation), so the respective synsets have the prime noun.artifact and are generally hyponyms of the synset *conveyance*:3; *transport*:8. Vehicles are distinguished from Instruments as their semantic and syntactic behaviour is more similar to Agents.

The relation **By-means-of** is also associated with two subtypes: on the one hand, it may be thought of as a kind of inanimate cause, e.g. *geyser*:1 ('to overflow like a geyser') – *geyser*:1 ('a spring that discharges hot water and steam') (noun.object), while on the other, it is found in cases where the semantics is not so much causative as enabling or facilitating: consider the pair *certify*:2 ('guarantee payment on; of checks') and *certificate*:2 ('a formal declaration that documents a fact of relevance to finance and investment').

The relation **Event** denotes a processual nominalization and involves nouns such as noun.act, noun.event, noun.phenomenon, noun.process, while ruling out concrete entities such as animate beings, natural (noun.object) or man-made (noun.artifact) objects, etc.

The relation **State** denotes abstract entities: feelings (noun.feeling), cognitive (noun.cognition) and other non-dynamic state-of-affairs, such as synsets with the prime noun.state.

The relation **Undergoer** denotes entities affected by the situation described and roughly corresponds to the thematic role of Patient/Theme.

The relation **Result** involves entities that are produced or come into existence as a result of the situation described by the verb.

The relation **Property** denotes various attributes and qualities. This relation involves primarily nouns with the prime noun.attribute and more rarely noun.location.

The relation **Location** denotes a concrete (natural or man-made) or an abstract location where an event takes place and therefore relates verbs with nouns with various primes – most typically noun.location, but also noun.object, noun.plant,

[2] https://wordnet.princeton.edu

[3] The actual size is 17,739 pairs, but we worked on an improved, more consistent version of the file (Koeva et al., 2016) and report cleaned data.

noun.artifact, noun.cognition, etc.

The relation **Destination** is associated with the primes noun.person, noun.location and noun.artifact, corresponding to two distinct interpretations in terms of the thematic role theory – as a Recipient (noun.person) or as a Goal (noun.artifact, noun.location).

The relation **Uses** denotes a function or purpose of an entity. In many cases, especially with verbs of putting, the entity is directly involved as the Theme of the verb, e.g. *lipstick*:2 ('apply lipstick to') – *lipstick*:1 ('makeup that is used to color the lips'). The relation allows nouns with various primes, both concrete and abstract.

A number of procedures towards the trimming of the morphosemantic relations in the standoff file were carried out previously (Koeva et al., 2016). These involved the disambiguation of 450 cases of multiple assignment, which included both very clear-cut 'bugs', such as the assignment of both Agent and Event to a pair of synsets, as well as ambiguous cases of relations that may be considered as overlapping in scope, such as Instrument and Uses or By-means-of and Instrument. The leading principle in choosing one relation over another was the consideration for the overall logic of the relations' assignment as reflected in the typical attested combinations of semantic primitives (of both verbs and nouns) and relations. Other inconsistencies were also removed following the same guidelines.

The analysis of the morphosemantic relations in light of their correspondence in the domain of thematic roles and their semantic grounding gives insights into the linguistic motivation behind the semantic description of the participants in the semantic structure of verbs and serves as a point of departure for a more fine-grained analysis of the semantics of derivation with respect to classes of words with certain properties, cf. Section 6.

3.2 Relations' Independence and Overlap

As the analysis of the data presented in the previous subsection reveals, some relations cover two distinct meanings: a causative one and a means-or-function-oriented one (consider the examples given for the relations Material and By-means-of). A more detailed approach would thus involve the redefinition and reassignment of relations so that they satisfy uniform criteria, a question which we leave aside for the time being.

On the other hand, not all relations seem to be equally justified. Indeed, Vehicle, as well as Body-part, may qualify as kinds of Instruments. However, both relations are very specifically defined, and the relevant nouns fall into clear-cut semantic classes and combine syntactically with very coherent classes of verbs, such as verbs of controlled motion or vehicle operation (Vehicle) or verbs of gestures and bodily movements (Body-part). Thus, we would rather recognise these relations' membership to a more comprehensive class of relations, rather than discarding them in favour of a greater generalisation by reassigning them as Instruments.

4 Distribution of Morphosemantic Relations between Affixal and Zero Derivation

The theoretical findings sketched in Section 2 and based on empirical analyses are reflected by the data we work with: on the one hand, zero derivation is found across all the relations under discussion; on the other, conversion is the prevalent process of creating new words for 8 relations (By-means-of, Undergoer, Vehicle, Result, Property, Location, Uses, Body-part), while suffixation is the dominant word-formation technique for 4 relations (Agent, Destination, Material, State); for 2 semantic relations (Instrument, Event) conversion and derivation are in quite strong competition. These data are shown in Table 1 and Figure 1.

For all morphosemantic pairs we analyzed, besides deciding upon the formation process (zero or affixal derivation), we have also automatically identified the affix (and manually validated the data) in the latter case. Thus we were able to establish the frequency of each occurring affix, as well as of the zero affix (∅ henceforth). For a number of relations ∅ is not prevalent, but is the major competitor of the most productive suffix. In Table 1 a comparison between the proportion of ∅ (column 3) and the most frequent affix (column 6) shows four relations clearly dominated by affixal derivation (State, Agent, Destination, and Material); however, for two relations (State and Agent) ∅ is the second most frequent affix. Further, the results for Instrument and Event demonstrate balance between conversion and affixation.

An interesting case is that of the Vehicle relation which is morphologically represented either by ∅ (57 cases) or by the suffix *-er* (37 cases). Sim-

Relation name	No. of Ø-deriv.	%	Most freq. aff.	No.	%	2nd most freq. aff.	No.	%	Rest, %	Total
Uses	655	87.92	-ation	31	4.16	-ify	19	2.55	5.37	745
Location	220	80.88	-ation	23	8.46	-er	14	5.15	5.51	272
Undergoer	664	76.85	-ation	87	10.07	-ee	36	4.17	8.91	864
Result	882	63.59	-ation	301	21.70	-ify	60	4.33	10.38	1,387
Property	190	62.09	-ation	58	18.95	-ence	25	8.17	10.78	306
Vehicle	57	60.64	-er	37	39.36	-	-	-	0.00	94
By-means-of	677	59.54	-er	155	13.63	-ation	195	17.15	9.67	1,137
Body-part	40	57.14	-er	28	40.00	-ate	2	2.86	0.00	70
Event	3,544	46.34	-ation	3,328	43.52	-ment	387	5.06	5.07	7,647
Instrument	352	45.30	-er	403	51.87	-ise	14	1.80	1.03	777
State	168	32.75	-ation	237	46.20	-ment	61	11.89	9.16	513
Agent	351	12.10	-er	2,491	85.90	-ation	19	0.66	1.34	2,900
Destination	2	6.90	-ee	25	86.21	-ify	2	6.90	0.00	29
Material	3	4.23	-er	58	81.69	-ise	5	7.04	7.04	71
TOTAL	**7,805**	**46.43**								

Table 1: Distribution of conversion and affixal derivation in PWN after changes were performed. The number of unique verb-noun derivational pairs labeled with morphosemantic relations totals 16,812.

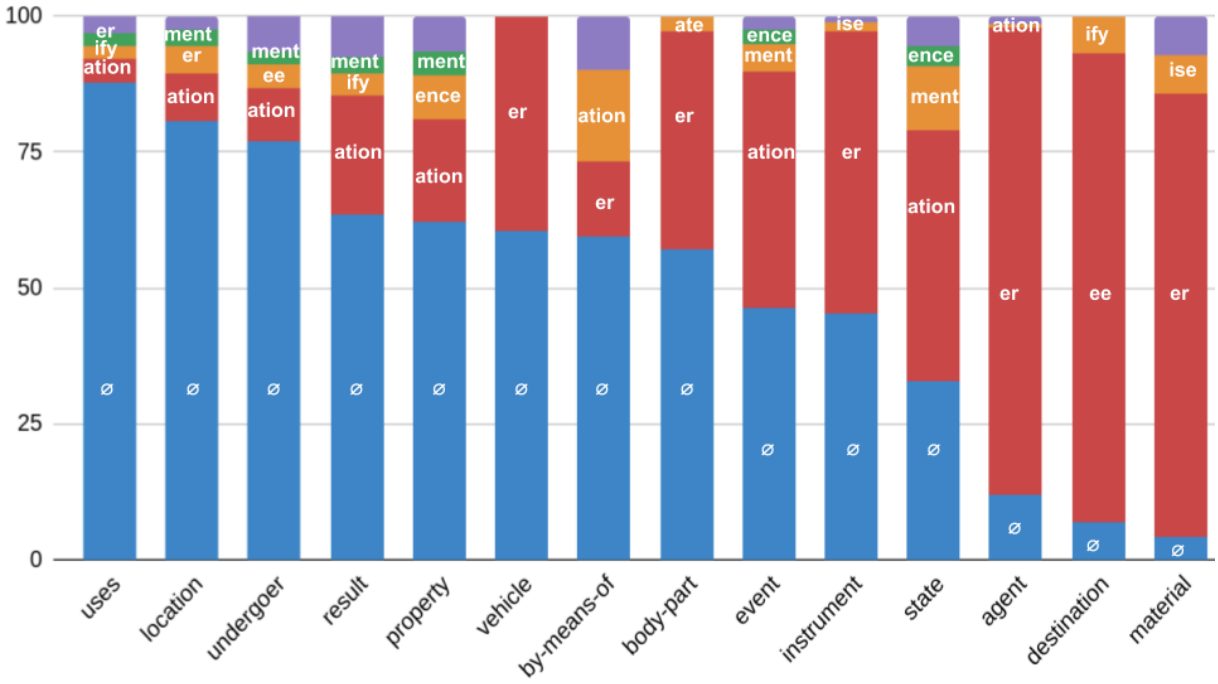

Figure 1: Competition between conversion (blue), the most frequent affix for each relation (red), the second most frequent affix (orange), the third most frequent one (green) and other affixes (purple).

ilarly, the relation Destination displays the suffix *-ee* (in 25 cases), Ø (in 2 cases) and verb suffix *-ify* (in other 2 cases). Both relations are scarce in the data under study, as the figures show.

The data presented clearly show that each relation is dominated by one or two affixes at most (including Ø), and that zero derivation plays an important role in creating new nouns and verbs, at least in the dataset under discussion. Based on our findings, *-er* and *-ation* are the most productive suffixes in these pairs, followed by Ø.

5 Morphosemantic Relations and Derivational Models

Besides the semantic perspective already incorporated by the morphosemantic relations (see Section 3), another perspective relevant for the process of creating new words is the tendency of certain semantic primes of serving either as derivation bases or as derivation results. We illustrate these trends below for each of the 14 relations.

5.1 Agent

The prevalent semantic prime of the nouns acting as Agents is noun.person, and this is no surprise. As opposed to the homogenous semantics of nouns, the verbs in these pairs belong to 14 (out of the 15) primes: the most productive ones with affixal derivation are:

• verb.communication (454[4]): *accuse*:1 – *accuser*:1, *announce*:1 – *announcer*:1;

• verb.contact (326): *carve*:1 – *carver*:1, *butt*:1 – *butter*:3;

• verb.social (318): *betray*:5 – *betrayer*:2;

while the most productive with conversion are:

• verb.social (97): *chairman*:1 – *chairman*:2, *knight*:1 – *knight*:3, *emcee*:1 – *emcee*:2;

• verb.communication (52): *blabber*:1 – *blabber*:2, *advocate*:1 – *advocate*:3;

• verb.possession (32): *auctioneer*:1 – *auctioneer*:2;

Some primes occur with both types of derivation, as noticed in these examples.

5.2 By-means-of

In the case of this morphosemantic relation, diverse verb and noun primes (11 and 10, respectively) are encountered in the data. The most frequent prime pair is verb.contact – noun.artifact, with 153 occurrences, 125 of them with conversion: *barricade*:3 – *barricade*:5, *chain*:1 – *chain*:3, *cushion*:1 – *cushion*:4. Other frequent pairs with conversion are:

• verb.communication – noun.communication (93): *alibi*:1 – *alibi*:3, *email*:2 – *email*:3, *gesture*:2 – *gesture*:4;

• verb.motion – noun.artifact (30): *bridge*:3 – *bridge*:5, *railroad*:1 – *railroad*:5, *sluice*:2 – *sluice*:5;

• verb.creation – noun.artifact (26): *festoon*:1 – *festoon*:2, *ornament*:1 – *ornament*:3, *cantilever*:1 – *cantilever*:3.

In the case of affixal derivation, the dominant prime pairs are:

• verb.communication – noun.communication (70): *impeach*:2 – *impeachment*:1, *confess*:1 – *confession*:2;

• verb.contact – noun.artifact (28): *decorate*:1 – *decoration*:2, *stop*:7 – *stopper*:2.

[4]The numbers in brackets indicate the number of synset pairs.

5.3 Destination

This relation is represented in the data with only 17 pairs. It is interesting that the prime noun.person is very well represented in combination with several verb primes and that the V-to-N derivation is dominated by the *-ee* suffix:

• verb.possession – noun.person (5): *grant*:5 – *grantee*:2, *trust*:5 – *trustee*:2;

• verb.communication – noun.person (4): *promise*:1 – *promisee*:1, *send*:2 – *sendee*:1;

• verb.social – noun.person (2): *patent*:5 – *patentee*:1, *retire*:7 – *retiree*:1;

• verb.motion – noun.person (2): *refer*:6 – *referee*:3.

5.4 Instrument

The prevalent prime pair is verb.contact – noun.artifact (398). Whereas nouns are mostly artifacts, the verbs are diverse: all 15 primes occur with this morphosemantic relation; some primes prevail with affixal derivation:

• verb.change (97): *deice*:1 – *deicer*:1;

• verb.motion (32): *elevate*:2 – *elevator*:1;

• verb.communication (24): *prompt*:2 – *prompter*:1, *buzz*:1 – *buzzer*:1, *page*:1 – *pager*:1.

Other primes tend to occur with conversion:

• verb.contact (239): *catapult*:2 – *catapult*:4;

• verb.creation (25): *crayon*:1 – *crayon*:2;

• verb.competition (14): *seine*:1 – *seine*:2.

5.5 Undergoer

Diverse noun and verb primes are implicated in pairs labeled with this morphosemantic relation, but the most frequent one is verb.communication – noun.communication: 77 occurrences out of which 42 are conversions (*compliment*:1 – *compliment*:3) and 35 are affixal derivations (*communicate*:1 – *communication*:1). Other prevalent prime pairs with conversion are:

• verb.possession – noun.possession (50): *store*:1 – *store*:6;

• verb.contact – noun.artifact (36): *veneer*:1 – *veneer*:3.

There are primes occurring only with conversion, never with affixal derivation:

• verbal primes: verb.competition: 17 with noun.animal (*rabbit*:1 – *rabbit*:2), 11 with noun.artifact (*bomb*:1 – *bomb*:3), 5 with noun.food (*prawn*:1 – *prawn*:3); verb.stative: 6 with noun.artifact (*overhang*:1 – *overhang*:3), etc.;

• noun primes: noun.animal: 17 with verb.competition (see above); 11 with verb.contact (*snail*:1 – *snail*:2); noun.plant: 11 with verb.change (*burr*:1 – *burr*:5), 7 with verb.contact (*mushroom*:2 – *mushroom*:7); noun.body: 10 with verb.body (*spit*:1 – *spit*:7), 6 with verb.contact (*transplant*:4 – *transplant*:7), etc.

Noun.person is the only noun prime for which derivation is more productive than conversion in the case of this morphosemantic relation: *address*:2 – *addressee*:1 (verb.communication), *employ*:2 – *employee*:1 (verb.social), *pay*:4 – *payee*:1 (verb.possession).

5.6 Vehicle

The prime pair verb.motion – noun.artifact is unsurprisingly the most frequent one among the pairs annotated as Vehicle: *baloon*:2 – *balloon*:3, *taxi*:2 – *taxi*:3. In the case of affixal derivation, another pair is notable: verb.competition – noun.artifact: *fight*:3 – *fighter*:1, *bomb*:1 – *bomber*:1.

5.7 Result

This relations involves great diversity in terms of both verb and noun primes. Among the most frequent prime pairs we find verb.creation – noun.artifact (82 occurrences, mostly conversions): *corduroy*:1 – *corduroy*:3. Other typical prime pairs with conversion include:

• verb.contact – noun.artifact (66): *bale*:1 – *bale*:2;

• verb.communication – noun.communication (36): *petition*:1 – *petition*:2;

Affixal derivation is frequently found with:

• verb.change – noun.substance (42): *calcify*:2 – *calcium*:1;

• verb.change – noun.attribute (37): *pinkify*:1 – *pink*:5;

• verb.change – noun.state (32): *calcify*:2 – *calcification*:3.

5.8 Body-part

This relation offers a fragmented picture in which 9 verb primes combine with 4 noun primes. As the relation is poorly represented, these prime pairs display only less than a handful of examples and we do not exemplify nor discuss them here.

5.9 Material

This relation displays the conglomeration of the pairs under 3 verb primes (verb.change, verb.contact, verb.body) and 2 noun primes (noun.artifact, noun.substance). The combination verb.change – noun.substance is the best represented (49 pairs): *opalize*:1 – *opal*:1.

5.10 Property

A relatively diverse set of 8 verb primes, the most productive of them being verb.change, verb.motion, verb.stative, combine with 2 noun primes, mostly with the prime noun.attribute and only a few pairs with noun.location. The most frequent prime pair is verb.change – noun.attribute (63, evenly distributed between zero and affixal derivation): *black*:4 – *black*:18, *cool*:1 – *cool*:11, *appear*:1 – *appearance*:4. Other frequent pairs with conversion are:

• verb.motion – noun.attribute (20): *slant*:3 – *slant*:5;

• verb.cognition – noun.attribute (14): *distrust*:1 – *distrust*:2;

• verb.contact – noun.attribute (11): *polish*:3 – *polish*:4.

Affixal derivation is more productive with the pairs:

• verb.change – noun.attribute (32): *align*:1 – *alignment*:2;

• verb.stative – noun.attribute (16): *abound*:1 – *abundance*:1.

5.11 Location

Diverse verb primes, among which the most productive ones are verb.motion, verb.contact, verb.stative, combine with nouns with primes such as noun.artifact, noun.location, noun.object, to express this relation. The most frequent prime pair is verb.contact – noun.artifact (39, mostly conversions[5]): *cabin*:1 – *cabin*:3, *closet*:1 – *closet*:2. Other frequent pairs are:

• verb.motion – noun.location (24): *port*:6 – *port*:14;

• verb.contact – noun.location (23): *park*:1 – *park*:7;

• verb.motion – noun.artifact (19): *corner*:1 – *corner*:4;

• verb.stative – noun.location (17): *bivouac*:1 – *bivouac*:3;

• verb.stative – noun.artifact (16): *lodge*:4 – *lodge*:5.

[5] Actually, examples of affixal derivation are very sparse with this relation.

5.12 Uses

Diversity of verb and noun primes characterizes this relation. The most frequent prime pair is verb.contact – noun.artifact (with over a hundred conversions and no affixal derivation): *carpet*:1 – *carpet*:4, *girth*:1 – *girth*:2. Other frequent pairs involve mainly conversion and they are:

• verb.possession – noun.artifact (57): *armor*:2 – *armor*:3;

• verb.contact – noun.substance (55): *asphalt*:1 – *asphalt*:3;

• verb.communication – noun.communication (44): *autograph*:1 – *autograph*:2;

• verb.body – noun.artifact (39): *bonnet*:1 – *bonnet*:2.

With affixal derivation a relatively frequent pair is verb.communication – noun.communication (13): *attest*:3 – *attestation*:1, while other pairs have only a few examples.

5.13 State

Many of the verb primes are involved in this relation, the most productive ones being: verb.change, verb.emotion, verb.social, verb.stative. Out of the several abstract noun primes, 2 occur more often: noun.state, noun.feeling. Affixation is more productive than conversion, but the dominant prime pairs are the same for both types of derivation:

• verb.emotion – noun.feeling (80): *abash*:1 - *abashment*:1, *joy*:2 – *joy*:4;

• verb.change – noun.state (86): *afflict*:1 – *affliction*:3, *decay*:1 – *decay*:8;

• verb.emotion – noun.state (48): *deject*:1 – *dejection*:1, *despair*:1 – *despair*:3.

5.14 Event

The most frequent prime pair is verb.communication – noun.communication and with this, the competition between derivation and conversion is the strongest (363 vs. 361). The most productive pairs differ for the two types of derivation. With affixal derivation they are:

• verb.change – noun.act (593): *alter*:3 – *alteration*:1;

• verb.social – noun.act (421): *abolish*:1 – *abolition*:1;

• verb.change – noun.process (283): *adapt*:1 – *adaptation*:3;

The most frequent pairs with conversion are:

• verb.motion – noun.act (423): *amble*:2 – *amble*:1;

• verb.contact – noun.act (337): *clasp*:2 – *clasp*:1;

• verb.competition – noun.act (126): *cricket*:2 – *cricket*:1.

6 Discussion

The presented data must be interpreted with a view to the PWN organization principles: all pairs contain words considered with only one of their possible meanings; i.e. the same pair of words may be found several times, labeled either with the same semantic relation or with a different one: e.g., the verb *net* and the noun *net* occur as a pair three times: once labeled as Instrument (for the meanings 'catch with a net' and 'a trap made of netting to catch fish or birds or insects', respectively), and twice as Result: the verb meaning 'yield as a net profit' and the noun denoting 'the excess of revenues over outlays in a given period of time (including depreciation and other noncash expenses)', and the verb meaning 'construct or form a web, as if by weaving' with the noun denoting 'an open fabric of string or rope or wire woven together at regular intervals'. Not all senses of the words can enter a morphosemantic relation with all senses of another word: e.g., the verb *net* has four senses in PWN, the homonymous noun has six senses, but the only morphosemantic relations between them are the three mentioned above.

On the other hand, the PWN files include 4,520 noun-verb derivational pairs that do not occur in the standoff file: e.g.: *carbon* and *carbonate* are linked by a derivational relation in the PWN, but they were not included in the standoff file.

Some pairs in the data are not direct derivatives: consider the homonymous verbs and nouns *black* or *green*, where both are derived from the corresponding adjectives. This is not the case with colors only: e.g. the verb and the noun *equal* are both derived from the respective adjective, too[6].

An interesting topic for research is the direction of conversion. There are examples of each direction among the pairs labeled with the same semantic relation: e.g. among the pairs labeled as Agent, we find nouns created from verbs by means of conversion, such as *snoop*, as well as verbs converted from nouns, such as *mouth*. There are cases when, for the same pair of primes, affixation goes in one direction, while zero derivation goes in the opposite one: e.g. for the prime pair verb.possession –

[6] According to data in https://www.etymonline.com

noun.person, nouns are derived from verbs (*auctioneer* from *auction*) and verbs are created from nouns via zero derivation (*auctioneer*). These observations need to be explored in more detail.

Researchers, see mainly Clark and Clark (1979) and Plag (1999), have aligned derivational semantics (zero derivation in particular) with the semantics of verb classes. Bauer et al. (2013) discuss the predictability of the semantics of nominalizations especially those denoting "an instance or a state aspectual meaning". Such information can also be drawn from our data, but taking the form of clusters of hyponyms that belong to the same region of the wordnet structure (the same subtree). The more detailed analysis of the data leads us to conclude that the clusters give a more profound insight into the semantic conditions on derivation than general classes as it is clusters that provide the structured part of the lexicon involved.

Relation: V prime – N prime pair (total #)		
Cluster root	**No. cases**	**%**
Agent: verb.body – noun.person (109)		
{change:1}	24	22.02
{act:1}	14	12.84
Agent: verb.change – noun.person (140)		
{change:1}	69	49.29
Agent: verb.cognition – noun.person (168)		
{think:3}	71	22.26
Agent: verb.communication – noun.person (506)		
{act:1}	234	46.25
{express:2}	66	13.04
{think:3}	46	9.09
Agent: verb.consumption – noun.person (69)		
{consume:2}	39	56.52
Agent: verb.creation – noun.person (205)		
{make:3}	127	61.95
Agent: verb.motion – noun.person (286)		
{go:1}	149	52.10
{move:3}	34	11.89
Agent: verb.perception – noun.person (74)		
{perceive:1}	18	24.32
{watch:1}	15	20.27
{show:4}	7	9.46
Agent: verb.possession – noun.person (250)		
{transfer:5}	82	32.80
{take:21}	25	10.00
{show:4}	7	9.46

Table 2: Some significant clusters within the morphosemantic relation of Agent.

Table 2 shows the overall number of occurrences for the most numerous combinations of verb primes with the prime noun.person for the relation Agent. The most meaningful clusters are represented as the root verb synsets to whose tree the verbs in the clusters belong, the nouns being in the subtree of *person*:1. The table shows each cluster's share as absolute numbers and as percentage of the number of the prime pair occurrences.

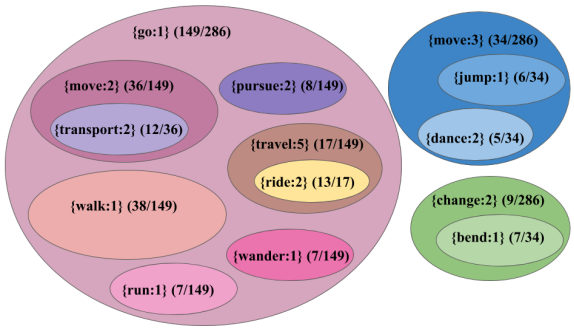

Figure 2: Clusters of verbs for the relation of Agent within the prime of verb.motion.

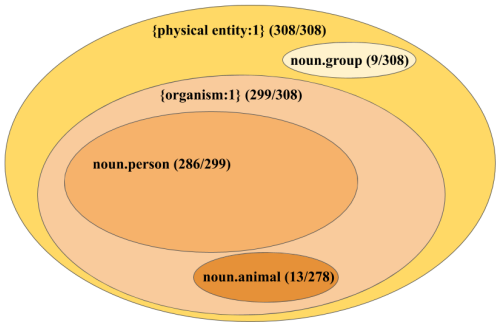

Figure 3: Clusters of nouns for the relation of Agent within the prime of verb.motion.

Figure 2 exemplifies the distribution of synsets belonging to the prime verb.motion which are involved in the relation Agent (see Table 2). More than half of the synsets (149 out of 286) are hyponyms of the synset *travel*:1; *go*:1; *move*:1; *locomote*:1 (the embedded bubbles). Each of these bubbles is the root of a smaller subtree and is also represented numerically. The blue and green bubbles exemplify outliers in other subtrees.

Figure 3 provides a similar representation of the noun synsets involved in the Agent relation with verbs of motion (verb.motion). Most of them are nouns designating persons (286 out of 308), with a very small number of synsets from other primes.

As not all derivational relations are assigned a morphosemanticl label (see above, this Section), the question of the predictability of the morphosemantic relations arises. Our analysis of several samples of the data shows that the relations are predictable only to a certain extent. A semantic relation of one of the 14 types could be automatically assigned to pairs of word senses that are derivationally related in PWN but lack a semantic

label (as in the standoff file) if they occur in subtrees where a certain relation is already assigned to other pairs: consider the pairs in the standoff file: *nickel*:1 – *nickel*:1 , *silver*:1 – *silver*:1, *copper*:1 – *copper*:1 and *chrome*:1 – *chrome*:1. These four verbs are hyponyms of *cover*:1, while the nouns are hyponyms of *metal*:1; the semantic relation these pairs are annotated with is Uses. However, there are other such pairs in PWN, e.g.: *aluminium*:1 – *aluminize*:1, where the verb is a hyponym of *cover*:1 and the noun is a hyponym of *metal*:1 and there is a derivational relation between them, but it is not labeled morphosemantically. Given the four examples above, we can infer that the right semantic label for this pair (and other similar ones) is Uses. In other cases inferring a semantic relation may not be so trivial, but at least the number of possibilities will be greatly reduced and manual validation will be facilitated. Besides labeling new pairs, such regularities can also help to easily spot oddities in the data and correct them.

Regular polysemy is reflected in morphosemantic relations, especially as from a contemporary point of view a verb's sense may be considered to be related to more than one (closely) related noun senses or vice versa. Such an example is found with nouns of the class noun.artifact (mostly containers) and nouns denoting the quantity that the respective container holds, e.g. *barrel*:2, *cask*:2 ('a cylindrical container that holds liquids') and *barrel*:4, *barrelful*:1 ('the quantity that a barrel (of any size) will hold'). Each of the two synsets is related to *barrel*:1 ('put in barrels') by means of the relations Location and Undergoer, respectively. Regular polysemy reveals how regularities between related meanings in the nominal or the verbal domain are reflected in the semantics of the relation in verb-noun pairs.

Observations on structured parts of the lexicon such as the ones discussed above enable us also to predict missing relations, both morphosemantic and derivational. Consider *jar*:5 ('place in a cylindrical vessel') and the noun synsets *jar*:1 ('a vessel (usually cylindrical') and *jar*:2, *jarful*:1 ('the quantity contained in a jar'). Although only the Undergoer relation is encoded, the Location relation is easily predictable on the basis of the *barrel* example above. Exploring further the hyponyms of the synset *containerful*:1 ('the quantity that a container will hold'), we discover that 25 out of its 67 hyponyms have corresponding verbs, but only 3 of the verbs are appropriately linked to the noun synsets denoting the respective quantity and artifact (in a like manner to *barrel*) – the remaining verbs lack one or both morphosemantic relations or even the derivational ones. In such a way, we are able to tackle the inconsistencies in derivational and morphosemantic relations throughout this and other parts of the PWN structure.

7 Conclusions

Our study based on the PWN standoff file consisting of noun-verb pairs labeled with one of a set of 14 semantic relations shows the distribution of zero and affixal derivation within the data, at a general level, as well as with respect to each such relation. We have also presented the most frequent affixes by means of which words are created in the subgroups represented by relations labeled identically and showed that the zero affix is among the most frequent ones for each such subgroup: for some relations it is the prevalent affix and for others it competes with the prevalent one. The semantics of these pairs was further enriched with information about the semantic primes of each word in the pair and several noun-verb prime combinations proved more frequent in some subgroups, with some of the combinations even being specialised for a certain type of derivation.

We intend to augment the work with other pairs extracted from the PWN files and already linked by a derivational relation. We envisage a better representation of certain affixes (especially verbal ones) that are sparse in the standoff file.

Our work can be extended to derivational relations for other languages using the corresponding wordnets. Since the semantic dimension of morphosemantic relations is transferable across languages using the interlingual indexing within PWN, it facilitates the study of derivation across languages and possibly in comparison as well.

Acknowledgments

This work was carried out in the project *Enhancing Multilingual Language Resources with Derivationally Linked Multiword Expressions* between the Institute for Bulgarian Language of the Bulgarian Academy of Sciences and the Research Institute for Artificial Intelligence of the Romanian Academy.

References

L. Bauer, R. Lieber, I. Plag. 2013. The Oxford Reference Guide to English Morphology. Oxford University Press.

B. Cetnarowska. 1993. The Syntax, Semantics and Derivation of Bare Nominalisations in English. Katowice. Uniwersytet Slaski.

E. Clark, and H. H. Clark. 1979. When Nouns Surface as Verbs. *Language*. 55 (4).

Ch. Fellbaum (ed.). 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Ch. Fellbaum, A. Osherson, and P. E. Clark. 2009. Putting Semantics into WordNet's 'Morphosemantic' Links. In: Z. Vetulani, H. Uszkoreit (eds) *Human Language Technology. Challenges of the Information Society. LTC 2007*. Lecture Notes in Computer Science, vol 5603. Springer, Berlin, Heidelberg.

S. Koeva, S. Leseva, I. Stoyanova, T. Dimitrova, M. Todorova. Automatic Prediction of Morphosemantic Relations. In: *Proceedings of GWC 2016*. Bucharest.

R. Lieber. 2004. Morphology and Lexical Semantics. Cambridge University Press.

G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, K.J. Miller. 1990. Introduction to WordNet: An Online Lexical Database. *International Journal of Lexicography,* 3 (4), pp. 235 – 244.

P. O. Müller 2016. Rules and Restrictions in Word-Formation I. General Aspects. *Word-Formation. An International Handbook of the Languages of Europe* De Gruyter Mouton.

I. Plag. 1999. *Morphological productivity: structural constraints in English derivation*. Berlin, New York: Mouton de Gruyter.