

Using Morphosemantic Information in Construction of a Pilot Lexical Semantic Resource for Turkish

Gözde Gül İsgüder-Şahin

Department of Computer Engineering
Istanbul Technical University
Istanbul, 34469, Turkey
isguderg@itu.edu.tr

Eşref Adalı

Department of Computer Engineering
Istanbul Technical University
Istanbul, 34469, Turkey
adali@itu.edu.tr

Abstract

Morphological units carry vast amount of semantic information for languages with rich inflectional and derivational morphology. In this paper we show how morphosemantic information available for morphologically rich languages can be used to reduce manual effort in creating semantic resources like PropBank and VerbNet; to increase performance of word sense disambiguation, semantic role labeling and related tasks. We test the consistency of these features in a pilot study for Turkish and show that; 1) Case markers are related with semantic roles and 2) Morphemes that change the valency of the verb follow a predictable pattern.

1 Introduction

In recent years considerable amount of research has been performed on extracting semantic information from sentences. Revealing such information is usually achieved by identifying the complements (arguments) of a predicate and assigning meaningful labels to them. Each label represents the argument's relation to its predicate and is referred to as a semantic role and this task is named as semantic role labeling (SRL). There exists some comprehensive semantically interpreted corpora such as FrameNet and PropBank. These corpora, annotated with semantic roles, help researchers to specify SRL as a task, furthermore are used as training and test data for supervised machine learning methods (Giuglea and Moschitti, 2006). These resources differ in type of semantic roles they use and type of additional information they provide.

FrameNet (FN) is a semantic network, built around the theory of **semantic frames**. This theory describes a type of event, relation, or entity with their participants which are called **frame elements (FEs)**. All predicates in same semantic frame share one set of FEs. A sample sentence annotated with FrameNet, VerbNet and PropBank conventions respectively, is given in Ex. 1. The predicate “buy” belongs to “Commerce buy”, more generally “Commercial transaction” frame of FrameNet which contains “Buyer”, “Goods” as core frame elements and “Seller” as a non-core frame element as in Ex. 1. FN also provides connections between semantic frames like inheritance, hierarchy and causativity. For example the frame “Commerce buy” is connected to “Importing” and “Shopping” frames with “used by” relation. Contrary to FN, **VerbNet (VN)** is a hierarchical verb lexicon, that contains categories of verbs based on Levin Verb classification (Schuler, 2006). The predicate “buy” is contained in “get-13.5.1” class of VN, among with the verbs “pick”, “reserve” and “book”. Members of same verb class share same set of semantic roles, referred to as **thematic roles**. In addition to thematic roles, verb classes are defined with different possible syntaxes for each class. One possible syntax for the class “get-13.5.1” is given in the second line of Ex. 1. Unlike FrameNet and VerbNet, **PropBank (PB)** (Palmer et al., 2005) does not make use of a reference ontology like semantic frames or verb classes. Instead semantic roles are numbered from Arg0 to Arg5 for the core arguments.

Ex. 1 [Jess]_{Buyer-Agent-Arg0} bought [a coat]_{Goods-Theme-Arg1} from [Abby]_{Seller-Source-Arg2}¹
Syntax: Agent V Theme {From} Source

This work is licenced under a Creative Commons Attribution 4.0 International License. Page numbers and proceedings footer are added by the organizers. License details: <http://creativecommons.org/licenses/by/4.0/>

¹In PropBank Arg0 is used for actor, agent, experiencer or cause of the event; Arg1 represents the patient, if the argument is affected by the action, and theme, if the argument is not structurally changed.

There doesn't exist a VerbNet, PropBank or a similar semantically interpretable resource for Turkish (except for WordNet (Bilgin et al., 2004)). Also, the only available morphologically and syntactically annotated treebank corpus: METU-Sabancı Dependency Treebank (Eryigit et al., 2011), (Ofizer et al., 2003), (Atalay et al., 2003) has only about 5600 sentences, which has presumably a low coverage of Turkish verbs. VerbNet defines possible syntaxes for each class of verbs. However, due to free word order and excessive case marking system, syntactic information is already encoded with case markers in Turkish. Thus the structure of VerbNet does not fit well to the Turkish language. PropBank simplifies semantic roles, but defines neither relations between verbs nor all possible syntaxes for each verb. Moreover only Arg0 and Arg1 are associated with a specific semantic content, which reduces the consistency among labeled arguments. Due to lack of a large-scale treebank corpus, building a high coverage PropBank is currently not possible for Turkish. FrameNet defines richer relations between verbs, but the frame elements are extremely fine-grained and building such a comprehensive resource requires a great amount of manual work for which human resources are not currently available for Turkish.

In this paper, we discuss how the semantic information supplied by morphemes, named as morphosemantics, can be included in the construction of semantic resources for languages with less resources and rich morphologies, like Turkish. We try to show that we can decrease manual effort for building such banks and increase consistency and connectivity of the resource by exploiting derivational morphology of verbs; eliminate mapping costs by associating syntactic information with semantic roles and increase the performance of SRL and word sense disambiguation by directly using morphosemantic information supplied with inflectional morphology. Then, we perform a pilot study to build a lexical semantic resource that contains syntactic information as well as semantic information that is defined by semantic roles both in VerbNet and PropBank fashion, by exploiting morphological properties of Turkish language.

2 Related Work

In study by Agirre et al. (2006) and Aldezabal et al. (2010), the authors discuss the suitability of PropBank model for Basque verbs. In addition to semantic role information, the case markers that are related to these roles are also included in the verb frames. It is stated that including case markers in Basque PropBank as a morphosemantic feature is useful for automatic tagging of semantic roles for Basque language which has 11 case markers. Hawwari et al. (2013) present a pilot study for building Arabic Morphological Pattern Net, that aims at representing a direct relationship between morphological patterns and semantic roles for Arabic language. Authors experiment with 10 different patterns and 2100 verb frames and analyze the structure and behavior of these Arabic verbs. The authors state that the results encourage them for a more comprehensive study. The SRL system for Arabic (Diab et al., 2008) and the light-verb detection system for Hungarian (Vincze et al., 2013) also benefited from the relation between case markers and semantic roles.

Furthermore, there are studies on exploiting morphosemantics in WordNets for different languages. Fellbaum et al. (2007), manually inspects WordNet's verb-noun pairs to find one-to-one mapping between an affix and a semantic role for English language. For example the nouns derived from the verbs with the suffixes *-er* and *-or*, like *invent-inventor* usually results as the agents of the event. However, it is stated that only two thirds of the pairs with this pattern could be classified as agents of the events. More patterns are examined and the regularity of these patterns are shown to be low for English language. In another work (Bilgin et al., 2004), authors propose a methodology, on exploiting morphosemantic information in languages where the morphemes are more regular. They perform a case study on Turkish, and propose application areas both monolingually and multilingually, such as globally enriching WordNets and auto detecting errors in WordNets. In a similar work (Mititelu, 2012), morphosemantic information is added to Romanian WordNet and the proposed application areas in Bilgin et al. (2004) are examined and shown to be feasible.

Previous studies based on building Basque PropBank focus on the building process of Basque PropBank, rather than analysis of the regularity of case markers and the relation between semantic roles and case markers. Furthermore, the study related to building Arabic Morphological Pattern Net, aims to build a separate dataset and map it to other resources such as Arabic VerbNet, WordNet and PropBank. Word-

Net has rich cross-language morphosemantic links however it does not list all arguments of predicates, thus its structure is not convenient for NLP tasks like semantic role labeling. These studies either make use of case markers or derivational morphology of verbs, not both. Moreover, some of them requires extra mapping resources and some are difficult to get utilized for semantic interpretation of sentences. Most important of all, none of the studies investigates Turkish language. To the best of our knowledge, this is the first attempt to build such a lexical semantic resource for Turkish and perform experiments on data to expose the relationship between semantic roles and morphemes known as case markers and valency changers in Turkish.

3 Morphosemantic Features

In morphologically rich languages, the meaning of a word is strongly determined by the morphemes that are attached to it. Some of these morphemes always add a predefined meaning while some differ, depending on the language. However, only regular features can be used for NLP tasks that require automatic semantic interpretation. Here, we determine two multilingual morphosemantic features: case markers and verb valency changing morphemes and analyze the regularity and usability of these features for Turkish.

3.1 Declension and Case Marking

Declension is a term used to express the inflection of nouns, pronouns, adjectives and articles for gender, number and case. It occurs in many languages such as Arabic, Basque, Sanskrit, Finnish, Hungarian, Latin, Russian and Turkish. In Table. 1, statistic performed by Iggesen (2013), shows that there are 86

2 cases	3 cases	4 cases	5-7 cases	8-9 cases	10 or more cases
23 languages	9 languages	9 languages	39 languages	23 languages	24 languages

Table 1: Case marking across languages

languages with at least 5 case markings. An exemplary morphological analysis for the Turkish word *evlerinde* “in his houses” is given in Ex. 2. In this analysis, *ev* is inflected with *ler* morpheme for plurality, *i* for third person singular and *(n)de* for locative (LOC) case.²

Ex. 2 *ev* (-ler) (-i) (-nde)
 ev +Noun+ Pl + P3s + LOC

Even though the languages differ, the same case markers are used to express similar meanings with some variation. In order to exemplify this statement, sentences with similar meanings and the same case markers are given in Table 2 for languages Turkish and Hungarian, which have rich case marking systems. Relation between semantic roles and case markers can assist researchers in solving some of the

	NOM	ACC	DAT	LOC	ABL
TR	Ben geldim. I-NOM come-PAST. I came.	Avcı tavşan-ı gördü. Hunter the rabbit-ACC see-PAST. The hunter saw the rabbit .	Jack okul-a gitti. Jack school-DAT go-PAST. Jack went to school .	Ankara'da oturuyorum. Ankara-LOC live-P1s-PRES. I live in Ankara .	Annem-den geldim. Mother-ABL come-P1s-PAST. I came from my mother .
HR	Ági jött. Ági come-PAST Ági came.	Látom a hegy-et . see-P1s mountain-ACC . I see the mountain .	Ági-nak adtam ezt a könyv-et. Ági-DAT give-P1s-PAST book-ACC. I gave this book to Ági .	Budapest-ban lakom. Budapest-LOC live-P1s-PRES. I live in Budapest .	Ági-tól jöttem. Ági-ABL come-P1s-PAST. I came from Ági .

Table 2: Case marking in Turkish and Hungarian

challenging problems in natural language processing. In languages where case markers exist, these

- can be used as features for Semantic Role Labeling,
- can supply prior information for disambiguating word senses,
- can be used in language generation as such: Once the predicate and the sense is determined, the arguments can directly be inflected with the case markers associated with their roles.

²Throughout the paper NOM is used as nominative, ACC as accusative, DAT as dative, LOC as locative, ABL as ablative, COM as comitative.

3.2 Valency Changing Morphemes

The valency of a verb can be defined as the verb’s ability to govern a particular number of arguments of a particular type. “In Turkish, verb stems govern relatively stable valency patterns or prototypical argument frames” as stated by Haig (1998). Consider the root verb *giy* “to wear”. One can derive new verbs from the root *giy* “to wear” such as *giy-in* “to get dressed”, *giy-dir* “to dress someone” and *giy-il* “to be worn”. These verbs are referred to as verb stems and these special suffixes are referred to as valency changing morphemes. Some advantages of valency changing morphemes are

- They exist for many languages.
- They are regular, easy to model and morphological analyzers available for such languages can analyze the valency of the verb stem.
- They are directly related to the number and type of the arguments, which are important for SRL related tasks.

By modeling the semantic role transformation from verb root to verb stem, we can automatically identify argument configuration of a new verb stem given the correct morphological analysis. By doing so, framing only the verb roots can guarantee to have frames of all verb stems derived from that root. This quickens the process of building a semantic resource, as well as automatizing and reducing the human error. In this section we present a pilot study for some available valencies in Turkish language. For the sake of simplicity, instead of thematic roles, argument labeling in the PropBank fashion is used.

Reflexive

As the word suggests, in reflexive verbs, the action defined by the verb has its effect directly on the person/thing who does the action (Hengirmen, 2002). The reflexive suffix triggers the suppression of one of the arguments. In Fig. 1 observed argument shift and in Table 3 some interesting reflexive Turkish verbs are given like *besle* “to feed” and *besle-n* “to eat - feed himself”.

[Kaçağ-1]_{A1} sakla-dı-lar. [Kaçak]_{A0} sakla-n-dı.
 convict-ACC hide-PAST convict hide-REFL-PAST
 [They]_{A0} hid [the convict]_{A1}. [The convict]_{A0} hid(himself).

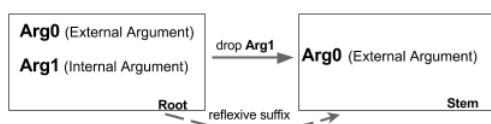


Figure 1: Argument transformation caused by reflexive suffix.

Root	Stem
<i>giy</i> (to wear)	<i>giy-in</i> (to get dressed)
<i>hazırla</i> (to prepare)	<i>hazırla-n</i> (to get ready)
<i>koru</i> (to protect)	<i>koru-n</i> (to protect himself)
<i>öv</i> (to praise)	<i>öv-ün</i> (to boast)
<i>sakla</i> (to hide)	<i>sakla-n</i> (to hide himself)
<i>besle</i> (to feed)	<i>besle-n</i> (to eat)

Table 3: Examples of reflexive verbs in Turkish

Reciprocal

Reciprocal verbs express actions done by more than one subject. The action may be done together or against each other. Reciprocal verbs may have a plural agent or two or more singular co-agents conjoined where one of them marked with COM case as shown in Fig 2. In both cases, the suppression of one of the arguments of the root verb is triggered. We have observed that the suppressed argument may be in different roles (patient, theme, stimulus, experiencer, co-patient), but usually appears as Arg1 and rarely as Arg2. In Table 4, a small list of reciprocal verbs is given. Some semantic links are easy to see, whereas the link between *döv* “to beat” and *döv-üş* “to fight” is not that explicit.

[Oğlan]_{A0} [kız-ı]_{A1} öp-tü. [Çift]_{A0} öp-üş-tü.
 boy girl-ACC kiss-PAST couple kiss-RECIP-PAST
 [The boy]_{A0} kissed [the girl]_{A1}. [They]_{A0} kissed.

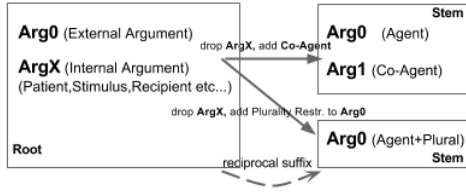


Figure 2: Argument transformation caused by reciprocal suffix.

Root	Stem	Meaning
küs (to offend)	küs-üş (to get cross)	with each other
öde (to pay)	öde-ş (to get even)	with each other
öp (to kiss)	öp-üş (to kiss)	with each other
sev (to love)	sev-iş (to make love)	with each other
döv (to beat)	döv-üş (to fight)	with each other
tanı (to know)	tanı-ş (to get to know)	each other

Table 4: Examples of reciprocal verbs

Causative

Causative category is the most common valence-changing category among Bybee’s (1985) world-wide sample of 50 languages. Contrary to other morphemes, causative morpheme introduces of a new argument called causer to the valence pattern. In most of the languages, only intransitive verbs are causativized (Haspelmath and Bardey, 1991). In this case, as shown in Fig. 3 the causee becomes the patient of the causation event. In other words, the central argument of the root verb, (Arg0 if exists, otherwise Arg1), is marked with ACC case and becomes an internal argument (usually Arg1) of the new causative verb. Some languages can have causatives from transitive verbs too, however the role and the mark of the causee may differ across languages. For the languages where the causee becomes an indirect object, like Turkish and Georgian, the central argument, Arg0 of the root verb, when transformed into a verb stem, receives the DAT case marker and serves as an indirect object (usually as Arg2), while Arg1 serves again as Arg1. This pattern for transitive verbs is given in Fig. 3. Some implicit relations exist in Table 5 such as *öl* “to die”, and cause someone to die *öl-dür* “to kill”. Transformation for intransitive verb *laugh* and transitive verb *wear*, is causativized as follows:

[Kız]_{A0} gül-üyor.
girl laugh-PROG
[The girl]_{A0} is laughing.

[Oğlan]_{A0} [kız-ı]_{A1} gül-**dür**-üyor.
boy girl-ACC gül-CAUS-PROG
[The boy]_{A0} is making [her]_{A1} laugh.

[Kız]_{A0} [mont-u-nu]_{A1} giy-di.
girl coat-POSS3S-ACC put+on-PAST
[The girl]_{A0} put on [her coat]_{A1}.

[Oğlan]_{A0} [kız-a]_{A2} [mont-u-nu]_{A1} giy-**dür**-di.
boy girl-DAT coat-POSS3S-ACC put+on-CAUS-PAST
[The boy]_{A0} had [the girl]_{A2} put on [her coat]_{A1}.

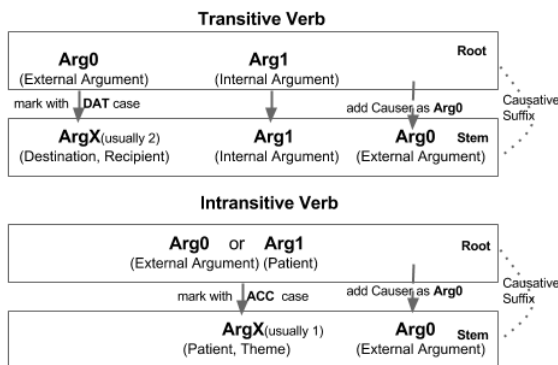


Figure 3: Argument transformation caused by causative suffix.

Root	Stem
ye (to eat)	ye-dir (to feed someone)
öl (to die)	öl-dür (to kill someone)
düş (to fall)	düş-ür (to drop sth.)
sür (to continue)	sür-dür (to resume)
oku (to read)	oku-t (to make someone read)
birleş (to join)	birleş-tir (to integrate)
yan (to get burnt)	yak (to set on fire)

Table 5: Examples of causative verbs

3.3 Application Areas

Semantic Role Labeling (SRL)

Semantic Role Labeling task is to identify the predicates and its arguments in the sentence, and then assign correct semantic roles to identified arguments. In Table 6, English sentences with different syntactic realizations and their translation into Turkish are given among with thematic roles annotated with

VN convention.³ In the second column, all words written in bold represent the arguments in destination roles. English sentences can not describe a common syntax for the destination role; different prepositions such as *into*, *at*, *onto* precedes the argument. However, in Turkish sentences they are always marked with dative case. Similarly, in the last column of Table 6, source and initial location roles are emphasized. Again, it is hard to find a distinguishing feature that reveals these roles in English sentences. There may be different prepositions *out of*, *from* or no preposition at all, before the argument in one of these roles, but they are naturally marked with ablative case in Turkish sentences.

Lang	Destination	Source
#1.En	She _{Ag} loaded boxes _{Th} into the wagon _{Dest} .	He _{Ag} backed out of the trip _{Sou} .
#1.Tr	Kutuları _{Th} vağon-a _{Dest-DAT} yükledi.	Seyahat-ten _{Sou-ABL} vâzgeçti.
#2.En	She _{Ag} squirted water _{Th} at me _{Dest} .	The convict _{Ag} escaped the prison _{IniLoc} .
#2.Tr	Ban-a _{Dest-DAT} su _{Th} fişkırttı.	Mahkum _{Ag} hapis-ten _{IniLoc-ABL} kaçtı.
#3.En	Paint _{Th} sprayed onto the wall _{Dest} .	He _{Ag} came from France _{IniLoc} .
#3.Tr	Duvar-a _{Dest-DAT} boya _{Th} püskürtüldü.	Fransa'dan _{IniLoc-ABL} geldi.

Table 6: Relation between case markers and semantic roles.

A subtask of automatic semantic role labeling is determining which features to extract from semantically annotated corpora. In recent studies, argument’s relative position to predicate (before, after) and voice of the sentence (passive, active) were experimented as features for automatic SRL (Wu, 2013). However, there exist many features and finding the best features requires feature engineering and again extra time. These toy examples suggest that there may be a correlation between case markers and semantic roles. If that is the case, the SRL task can be reduced to predicate and argument identification task, since the labeling will be automatically or semi-automatically done by using case markers as features.

Word Sense Disambiguation

The task of finding the meaning of a word in the context in question is called word sense disambiguation. In Table 7 three senses of Turkish verb lemma *ayır* and their arguments with case markers are given. In the first sense, the arguments are marked with ACC and DAT, with ABL and NOM in the second and with ACC, ABL in the third. The second and the third senses are similar. The action of reserving is usually performed on an indefinite object which usually appears in NOM form, where separating is applied on a certain object that is usually marked with ACC case. After the arguments are identified, one can easily detect the sense of the verb “ayır” by looking at arguments’ case markings.

	ayır.01 - To divide, split into pieces
#1.En	[He/she] _{Ag} divided [the apple] _{Pat} [into four] _{Dest} .
#1.Tr	[Elmay-ı] _{Pat-ACC} [dörd-e] _{Dest-DAT} ayırdı.
	ayır.02 - To keep, reserve (get-13.5.1)
#2.En	[I] _{Ag} reserved [a table] _{Th} [from the restaurant] _{Sou} .
#2.Tr	[Restoran- dan] _{Sou-ABL} [masa] _{Th-NOM} ayırdım.
	ayır.03 - To separate (separate-23.1)
#3.En	[I] _{Ag} separated [the yolk] _{Pat1} [from the white] _{Pat2} .
#3.Tr	[Sarısin-ı] _{Pat1-ACC} [beyazın- dan] _{Pat2-ABL} ayırdım.

Table 7: Relation between case markers and word senses

4 Methodology

We have performed a feasibility study for using morphosemantic features in building a lexical semantic resource for Turkish. As discussed in Section 3.2, we assume we can automatically frame a verb (e.g. *sakla* – *n(reflexive)*) that is derived with a regular valency changing morpheme (e.g. *n*), if the argument configuration of the root verb (e.g. *sakla*) is known. Hence, we have only framed root verbs. We have framed 233 root verbs and 452 verb senses. We have calculated the total number of valency changing morphemes as 425. This means 425 verbs can be automatically framed by applying the valency patterns to 233 root verbs. In this analysis we have only considered one sense of the verb since there may be cases where valency changing morpheme can not be applied to another sense of the verb. This can

³Throughout the paper Ag is used as agent, Th as theme, Dest as destination, Sou as source, Pat as Patient, IniLoc as initial location.

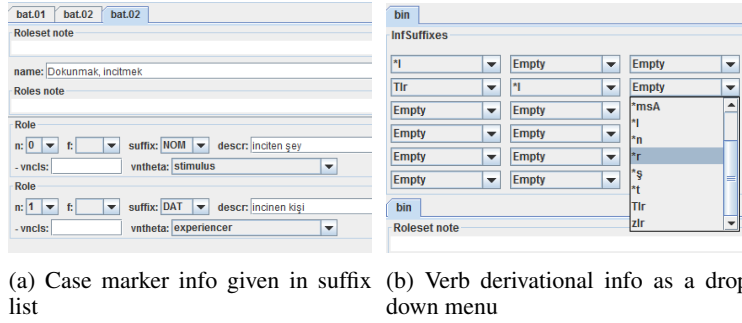


Figure 4: Cornerstone Software Adjusted for Turkish

not be automatically determined. Moreover, a verb stem may have multiple senses. In that case automatically extracted argument transformation may be wrong, because the verb stem may have a completely different meaning.

Turkish is not among rich languages by means of computational resources as discussed before. Turkish Language Association (TDK) is a trustworthy source for lexical datasets and dictionaries. To run this pilot study, we have used the list of Turkish root verbs provided by TDK and the TNC corpus⁴. The interface built for searching the TNC corpus gives the possibility to see all sentences that were built with the verb the user is searching for (Aksan and Aksan, 2012). The senses of the verbs and case marking of their arguments are decided by manually investigating the sentences appear in search results of the TNC corpus. Then, the arguments of the predicates are labeled with VerbNet thematic roles and PropBank argument numbers, by checking the English equivalent of Turkish verb sense. This process is repeated for all verb senses.

For framing purposes, we have adjusted an already available open source software, cornerstone (Choi et al., 2010)⁵. To supply case marking information of the argument, a drop down menu containing six possible case markers in Turkish is added as shown in Fig 4a. Finally, another drop down menu that contains all possible suffixes that a Turkish verb can have is added, shown in Fig 4b. Theoretically, the number of possible derivations may be infinite for some Turkish verbs, due to its rich generative property. However, practically the average number of inflectional groups in a word is less than two (Ofrazier et al., 2003). TDK provides a lexicon⁶ for widely used verb stems derived from root verbs by a valency changing morpheme. To avoid framing a nonexisting verb, we have used a simple interface shown in Fig 4b to enter only the stems given by TDK. An example with the Turkish verb *bin* “to ride” is given in Fig 4b. The first line defines that one can generate a stem *bin-il* “to be ridden by someone” from the root *bin* by using the suffix *l*. Similarly, second line illustrates a two layer derivational morphology, which can be interpreted as producing two verbs: *bin-dir* “cause someone to ride something” and *bin-dir-il* “to be caused by someone to ride something”.

5 Experiments and Results

In Table 8, number of co-occurrences of each thematic role with each case marker are given. Since in PropBank only Arg0 and Arg1 have a certain semantic interpretation, we have used VerbNet thematic roles in our analysis. Some roles look highly related with a case marker, while some look arbitrary. Results can be interpreted in two ways: 1) If the semantic roles are known and case marker information is needed, Agent will be marked with NOM, Destination with DAT, Source with ABL and Recipient with DAT case with more than 0.98 probability, furthermore Patient and Theme can be restricted to NOM or ACC cases; 2) If case markers are known and semantic role information is needed, only restrictions and prior probabilities can be provided. Highest probabilities occur with COM-instrument, LOC-location, DAT-destination, ACC-Theme and NOM-Agent pairs. We have applied our proposed argument trans-

⁴TNC corpus is a balanced and a representative corpus of contemporary Turkish with 50 million words

⁵Cornerstone is also used for building English, Chinese and Hindi/Urdu PropBanks.

⁶This lexicon is not computationally available

	NOM	ACC	DAT	LOC	ABL	COM	Total	Explanation
Agent	318	0	1	0	0	0	319	Human or an animate subject that controls or initiates the action.
Patient	36	34	0	0	0	0	70	Participants that undergo a state of change.
Theme	101	117	14	0	7	1	240	Participants in a location or experience a change of location
Beneficiary	1	2	5	0	0	0	8	Entity that benefits negatively or positively from the action.
Location	0	0	2	6	0	0	8	Place or path
Destination	1	0	66	0	0	0	67	End point or direction towards which the motion is directed.
Source	0	0	0	0	29	0	29	Start point of the motion.
Experiencer	13	5	4	0	0	0	22	Usually used for subjects of verbs of perception or psychology.
Stimulus	8	2	4	0	2	0	16	Objects that cause some response from Experiencer.
Instrument	0	0	0	0	0	10	10	Objects that come in contact with an object and cause a change.
Recipient	0	1	13	0	0	0	14	Animate or organization target of transfer.
Time	1	0	2	2	0	0	5	Time.
Topic	0	1	3	0	2	0	6	Theme of communication verbs.
Total	479	162	114	8	40	11	814	

Table 8: Results of Semantic roles - Case Marking

	#Intransitive	#Transitive	#Hold	#!Hold	Total
Reflexive	0	20	20	0	20
Reciprocal	8	18	26	0	26
Causative	26	11	37	0	37

Table 9: Results of Argument Transformation

formation on verbs with different valencies, and compared the argument configurations of the roots and stems. In Table 9, rows represent the valency changes applied to verb root, where Intransitive column contains the number of intransitive verbs that the pattern is applied to, and Transitive similarly. The #Hold column shows the number of root verbs for which the proposed patterns hold, and #!Hold shows the number of times the pattern can not be observed. Reflexive pattern can only be applied to transitive verbs, while others can be applied to both. Experiments are done for reflexive, reciprocal and causative forms. Our preliminary results on a small set of root verbs show that proposed argument transformation can be seen as a regular transformation.

6 Conclusion and Future Work

In this study, we presented a pilot study for building a Turkish lexical semantic resource for 452 verb senses by making use of two morphosemantic features that appear to be useful for challenging NLP tasks. Our experimental results on 814 arguments showed that the first feature, case markers, are not arbitrarily linked with a semantic role. This brings us to a conclusion that they can be a distinguishing feature for SRL, word sense disambiguation and language generation tasks. We ran some experiments for the second feature, valency changing morphemes and observed that the transformation of the argument structures of root to stem follows a specific pattern, hence proposed transformation seems to be regular and predictable. The results suggest that argument configuration of the root verb may be enough to label any verb stem derived with valency changing morphemes. This gives us the ability to build a semantic resource in a shorter time and reduce the human error, as well as provide a direct relationship like "causativity", "reflexivity" and "reciprocity" between verbs except for some problematic cases explained in Sect. 5. To conclude, this study encourages us to continue using morphosemantic features and increase the size of this resource.

7 Acknowledgements

We thank Gülşen Eryiğit and our three anonymous reviewers for insightful comments and suggestions that helped us improve this work.

References

- Eneko Agirre, Izaskun Aldezabal, Jone Etxeberria and Eli Pociello 2006. A Preliminary Study for Building the Basque PropBank. *In LREC 2006*, Genoa
- Yeşim Aksan and Mustafa Aksan 2012. Construction of the Turkish National Corpus (TNC). *In LREC 2012*, Istanbul

- Izaskun Aldezabal, María Jesús Aranzabe, Arantza Díaz de Ilarraza Sánchez and Ainara Estarrona. 2010. Building the Basque PropBank. *In LREC 2010*, Malta
- Nart B. Atalay, Kemal Oflazer and Bilge Say. 2003. The Annotation Process in the Turkish Treebank. *In Proceedings of the EACL Workshop on Linguistically Interpreted Corpora. Budapest*
- Orhan Bilgin, Özlem Çetinoğlu and Kemal Oflazer. 2004. Building a wordnet for Turkish. *Romanian Journal of Information Science and Technology*, 7.1-2 (2004): 163-172.
- Orhan Bilgin, Özlem Çetinoğlu and Kemal Oflazer. 2004. Morphosemantic relations in and across Wordnets. *In Proceedings of the Global Wordnet Conference*, pp. 60-66. 2004.
- Joan L. Bybee. 1985. *Morphology: A Study of the Relation between Meaning and Form. Typological Studies in Language 9 Amsterdam, Philadelphia: Benjamins*
- Jinho D. Choi, Claire Bonial and Martha Palmer. 2010. Propbank Frameset Annotation Guidelines Using a Dedicated Editor, Cornerstone. *In LREC 10*, Malta
- Mona Diab, Alessandro Moschitti and Daniele Pighin. 2008. Semantic Role Labeling Systems for Arabic Language using Kernel Methods *In Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies.*, 2008
- Gülşen Eryiğit, Tugay İlbay and Ozan A. Can. 2011. Multiword Expressions in Statistical Dependency Parsing. *In Proceedings of the Workshop on Statistical Parsing of Morphologically-Rich Languages SPRML at IWPT, Dublin*
- Christiane Fellbaum, Anne Osherson and Peter E Clark. 2007 Putting Semantics into WordNet's "Morphosemantic" Links. *Computing Reviews*, 24(11):503–512.
- Ana-Maria Giuglea and Alessandro Moschitti. 2006. Semantic Role Labeling via FrameNet, VerbNet and PropBank. *In Proceedings of the 21st International Conference on Computational Linguistics*, pp. 929-936. 2006.
- Geoffrey Haig. 1998. *Relative Constructions in Turkish. Otto Harrassowitz Verlag.*
- Martin Haspelmath, Thomas M. Bardey 1991. *Valence change. HSK-Morphology. A Handbook on Inflection and Word Formation; ed. by G. Booij, C. Lehmann and J. Mugdan, MPI Leipzig, Universität Mainz*
- Abdelati Hawwari, Wajdi Zaghouni, Tim O’Gorman, Ahmed Badran and Mona Diab. 2013. Building a lexical semantic resource for Arabic morphological Patterns. *In Communications, Signal Processing, and their Applications (ICCSPA)*
- Mehmet Hengirmen. 2004. *Türkçe Dilbilgisi. Engin Yayınevi*
- Oliver Iggesen 2013. *Number of cases. In World atlas of language structures online, ed. Matthew S. Dryer and Martin Haspelmath, Leipzig: Max Plank Institute for Evolutionary Anthropology Available online at <http://wals.info/chapter/49>*
- Verginica B. Mititelu. 2012. Adding Morpho-semantic Relations to the Romanian Wordnet. *In LREC 2012, İstanbul*
- Kemal Oflazer, Bilge Say, Dilek Z. Hakkani-Tür and Gökhan Tür. 2003. Building a Turkish Treebank. *Invited chapter in Building and Exploiting Syntactically annotated Corpora, Anne Abeille Editor, Kluwer Academic Publishers*
- Martha Palmer, Paul Kingsbury and Daniel Gildea. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *In Computational Linguistics*, 31(1):71–106
- Karin K. Schuler 2006. *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon* PhD diss., University of Pennsylvania
- Veronika Vincze, István Nagy T. and János Zsibrita. 2013. Learning to detect english and hungarian light verb constructions. *ACM Trans. Speech Lang. Process.*, 10, 2, Article 6 (June 2013), 25 pages
- Shumin Wu. 2013. Semantic Role Labeling Tutorial: Supervised Machine Learning methods. *In Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*