Turkish WordNet KeNet

Özge Bakay
Boğaziçi University
bakayozge@gmail.com

Özlem Ergelen
Boğaziçi University

Elif Sarmış
Starlang Yazılım Danışmanlık
elif@starlangyazilim.com

Selin Yıldırım
Özyeğin University
selin.yildirim@ozyegin.edu.tr

Atilla Kocabalcıoğlu
Boğaziçi University

Bilge Nas Arıcan
Starlang Yazılım Danışmanlık
bilge@starlangyazilim.com

Merve Özçelik
Starlang Yazılım Danışmanlık
merve@starlangyazilim.com

Ezgi Sanıyar
Starlang Yazılım Danışmanlık
ezgi@starlangyazilim.com

Oğuzhan Kuyrukçu
Starlang Yazılım Danışmanlık
oguzhan@starlangyazilim.com

Begüm Avar
Boğaziçi University

Olcay Taner Yıldız
Özyeğin University
olcay.yildiz@ozyegin.edu.tr

Abstract

Currently, there are two available wordnets for Turkish: TR-wordnet of BalkaNet and KeNet. As the more comprehensive wordnet for Turkish, KeNet includes 76,757 synsets. KeNet has both intralingual semantic relations and is linked to PWN through interlingual relations. In this paper, we present the procedure adopted in creating KeNet, give details about our approach in annotating semantic relations such as hypernymy and discuss the language-specific problems encountered in these processes.

1 Introduction

Information regarding words and meanings are traditionally stored in dictionaries. With the advancement of natural language processing studies, a need for machine-readable dictionaries has arisen (Miller, 1995). In an attempt to answer that need, wordnets which store lexicographic information in a format that is adaptable to modern computing have emerged. Wordnet, in its broader definition, is a highly comprehensive dictionary that is built on distinct word senses along with their definitions. Most of the words in a wordnet are open-class words such as nouns, verbs, adjectives and adverbs. Main building blocks of a wordnet are *synsets*, which are comprised of synonym synset members. Synsets are the distinct units in wordnets and all the mappings including intra- and inter-lingual ones are constructed based on the synsets. In lexical semantics, it is argued that words can be defined based on the relations between them. Adopting this principle, wordnets map semantic relations such as hypernymy, meronymy or antonymy through synsets.

Turkish wordnet of BalkaNet (TR-wordnet) (Bilgin et al., 2004) is the first wordnet created for Turkish. TR-wordnet of BalkaNet includes 14,626 synsets and 19,834 internal semantic relations. In this paper, we present our work on creating a more comprehensive wordnet for Turkish, namely KeNet[1,2] (Ehsani et al., 2018) and discuss the creation of semantic relations such as hypernymy. We present the literature review on wordnets for other languages in Section 2, describe the process of synset construction in KeNet in Section 3, present intralingual semantic relations including hypernymy in Section 4, explain the interlingual mapping of KeNet to Princeton WordNet in Section 5, summarize the challenges we have encountered in all these processes in Section 6 and conclude in Section 7.

2 Literature Review

Constructing a wordnet, whether from scratch or by expanding a previous one, is a labor intensive process that requires several steps and ex-

[1]"Ke" in KeNet comes from "kelime" (word) in Turkish.

[2]KeNet can be freely and publicly downloaded under an open source licence from:
https://github.com/StarlangSoftware/TurkishWordNet
https://github.com/StarlangSoftware/TurkishWordNet-Py
https://github.com/StarlangSoftware/TurkishWordNet-Cy
https://github.com/StarlangSoftware/TurkishWordNet-C#
https://github.com/StarlangSoftware/TurkishWordNet-CPP

tensive use of both human labor and automated systems. Since the creation of the first wordnet Princeton WordNet (PWN) in 1995 (Miller, 1995), many other wordnets have been created for several languages (e.g., Finnish WordNet FinnWordNet (Linden and Carlson, 2010), Polish WordNet (Derwojedowa et al., 2008), Norwegian WordNet (Fjeld and Nygaard, 2009), Danish WordNet (Pedersen et al., 2009), French WordNet WOLF (Sagot, 2008)). In addition, multilingual wordnets linking the wordnets of multiple languages have been created. To exemplify, EuroWordNet (EWN) is a multilingual WordNet project that consists several European languages (English, Dutch, Italian, Spanish, German, French, Czech and Estonian) (Vossen, 2007). In EWN, the wordnets were created for each language separately and then linked through an Inter-Lingual-Index based on PWN. BalkaNet, similar to EWN, is a multilingual wordnet project consisting of six Balkan languages (Bulgarian, Czech, Greek, Romanian, Serbian, and Turkish) (Tufis et al., 2004). This project was done to produce a multilingual semantic network, fully compatible with EWN and its extensions.[3]

Different approaches were adopted in creating wordnets and mapping them with those for other languages. Two of the most-commonly used approaches in the literature are *expand approach* and *merge approach*. In the expand approach, a set of synsets from PWN, including their semantic database, are first translated into the target language and then the relations that are transferred from English are checked in a manual fashion. For example, the Finnish WordNet FinnWordNet (Linden and Carlson, 2010) and the French WordNet WOLF (Sagot, 2008) were constructed with this approach. In the merge approach, on the other hand, the first step in creating a new wordnet is to build the intralingual relations from scratch, without any links to English. The monolingual wordnet is then mapped onto English. Exemplary wordnets that were created with this approach are Polish WordNet (Derwojedowa et al., 2008), Norwegian WordNet NorNet (Fjeld and Nygaard, 2009) and Danish WordNet DanNet (Pedersen et al., 2009). In a comparison of these two approaches, it is argued that the expand approach is more practical and less time-consuming since it enables us to have many correct monolingual relations that are extracted from PWN automatically. This automatic extraction of relations from PWN is especially beneficial for languages which show a similar pattern to English in their semantic associations, such as in the case of French (Sagot, 2008).

In the following sections, the details in creating KeNet is presented.

3 Synset Construction

The very first step in constructing KeNet, as in every other wordnet, was to create synsets. Synset can be defined as a group of words sharing the same sense and part of speech (POS). The structure of a sample synset in KeNet is as follows:

```
<SYNSET>
  <ID>TUR10-0038510</ID>
  <synset member>anne<SENSE>2</SENSE>
  </synset member>
 <POS>n</POS>
 <DEF>...</DEF>
 <EXAMPLE>...</EXAMPLE>
</SYNSET>
```

An exemplary set of synsets from KeNet is given in Table 1. In this table, examples of the four most frequent parts of speech in KeNet are listed, i.e., noun, adjective, verb and adverb, respectively. For each of these examples, the first column shows the ID of the synset. The characters that are separated with "-" from the ID gives the POS of the synset (*n* for noun, *v* for verb, *a* for adjective, *adv* for adverb). The second column lists the synset members; the synset members that are listed in the same synset are synonyms. The third column demonstrates the definitions and lastly, the fourth column presents an exemplary sentence (if there is any) including one of the synset members.

Regarding the construction of these synsets, the first version of the database was constructed through mining of the latest Contemporary Dictionary of Turkish (CDT) (2011's print) published by the Turkish Language Institute (TLI) (Ehsani et al., 2018). By convention, CDT marks synonyms by using commas such that synonyms of a word are given after its definition with a separation of comma. To decide on true synonyms that must occur in the same synsets, we sliced the definitions at commas and listed the comma-separated lemmas and the rest of the definitions as candidates of synonyms. Then, those lists were displayed for

[3]Further details about the wordnets that are discussed above and many others can be found on the website for the Global WordNet Association (GWA).

Table 1: Exemplary Synsets

Synset ID	Synset Members	Definition	Example Sentence
TUR10-0000030-n	su ab âb ”water”	Hidrojenle oksijenden oluşan, oda sıcaklığında sıvı durumunda bulunan, renksiz, kokusuz, tatsız madde	
TUR10-0000220-a	abajurlu ”with lampshade”	Abajuru olan	Üstünde lacivert abajurlu, parlak bir madenden lamba.
TUR10-0000350-v	abanmak ”to lean over”	Eğilerek bir şeyin, bir kimsenin üzerine kapanmak	Efendi, sen de ne üstüme abanıyorsun?
TUR10-0000520-adv	abartısız mübalağasız ”without exaggeration”	Abartmadan, abartısız olarak, mübalağasız bir biçimde	

linguistically-informed human annotators who decided on the synonymy relation between the lemmas and the definitions. 49,774 pairs were annotated at the end of this phase. Although some of them were included as separate entries in CDT, passivized and causativized forms of verbs were deleted from KeNet as they share the same root with their active forms.

Although the vast majority of the synsets were constructed during this process, there was a need for follow-up procedures to improve the organization of the current synsets. Since the main problem encountered in synset construction was the semantic relatedness of the synset members, two other procedures were followed in order to control the synonymy relations within the synsets: the *merge* process and the *split* process. These two processes are discussed next.

3.1 Merge Process

In the merge process, different synsets that should be grouped together were identified and grouped as a single synset. Three things were crucial while merging the synsets: (i) having a single and unique definition for each synset, (ii) having true synonyms as synset members in each synset and (iii) having a representative first synset member in each synset. Firstly, the synsets that were created by combining the synset members with identical senses had as many definitions as the number of synset members in them since the definitions were also merged while merging the synset members. The definitions of the merged synsets were initially combined with a pipe symbol in between them. A new definition for each merged synset was written so that each synset had a single and unique definition that covers the meaning of all its synset members. None of the synset members of a synset appeared in its definition. In this process, new definitions for 10,612 number of synsets were written by the human annotators.

Secondly, some synsets were found to include unrelated synset members. Therefore, another goal of the merge process was to include only the synset members that were synonyms. 1,144 number of synsets with unrelated synset members that had been identified in other parts of the work were transferred to the split process (see Section 3.2 for details). Additionally, there were cases where the synsets were missing some of the necessary synset members. Whereas some of these missing synset members were present in KeNet, some were not. Those that were already present in KeNet were merged with the current synsets. Those that were not present in KeNet were added as distinct synset members in the existing synsets. At the end of this process, 122 number of synsets were enriched with new synset members.

Lastly, the order of the synset members in the synsets was checked in this process. Due to time limitation, only the first synset member was checked. The most frequently-used synset members in the synsets were placed as the first to appear in order to have a representative display. The ordering of the rest of the synset members was noted as a future task.

3.2 Split Process

In the split process, the synsets that included synset members with different senses were split and separate synsets were created for each group of related synset members. In order to fix this

Table 2: Number of Synsets in KeNet

Part of Speech	# of Synsets
Nouns	44,074
Verbs	17,791
Adjectives	12,416
Adverbs	2,550
Interjections	3342
Pronouns	68
Conjunctions	60
Postpositions	29
Total	77,330

problem, we created a pool where we collected all the synsets that had unrelated synset members. We displayed these synsets on Google Sheets. Linguistically-informed human annotators then split these wrongly-merged synsets and wrote new definitions for the newly-created ones.

There were three main reasons for the wrong mergings: meaning-related drifts, POS-related drifts and morphology-related problems (Bakay et al., 2019c). For meaning-related drifts, the synset members that were semantically related but not true synonyms, e.g., nouns with close meanings such as *dere* "brook" and "*ırmak nehir* "river", had been mistakenly conjoined. For POS-related drifts, synset members which were semantically related but had different POS, e.g., a noun and an adjective with a similar meaning or coming from the same root such as *güç* "difficult" and *güç* "strength", had been mistakenly combined. Lastly, for morphology-related drifts, morphologically-related synset members, e.g., verbs that are morphologically related but have different meanings and different argument structures such as *sopalamak* "to beat somebody" and *sopalanmak* "to get beaten by somebody", had been mistakenly grouped under the same synsets. These synsets were split and different synsets were created for each group.

4 Semantic Relations

Currently, there are 77,330 synsets, 109,049 synset members and 80,956 distinct synset members in KeNet. The POS categories that are included are nouns, adverbs, adjectives, adverbs, interjections, pronouns, postpositions and conjunctions (see Table 2 for numbers). Regarding the number of words in synset members, although the majority of the synset members are one- (72,436 - 66.48%) or two-word (31,705 - 29.36%) synset members, there are synset members including up to seven words. Lastly, 19,776 number of synsets have exemplary sentences (25.57%). Including an exemplary sentence for each synset was noted as future work.

In KeNet, eight intralingual semantic relations were included: hypernymy, derivational relatedness, domain topic, part holonymy, antonymy, instance hypernymy, member holonymy, substance holonymy and attribute (see Table 3 for examples and the current number of matchings for these relations). For all these relations, the main word class that was annotated was nouns whereas antonymy and attribute were mainly annotated for adjectives.

There can be various approaches to constructing semantic relations in a wordnet such as translating an already constructed wordnet from another language into the target language or building one from scratch. Both approaches have their advantages and disadvantages, which will change drastically from one language to another. Translating a previously constructed wordnet into another language, while seems to be the easier approach, comes with a lot of disadvantages, especially in languages like Turkish which are morphologically and syntactically different from English.

In KeNet, in the creation of all these eight semantic relations, we consulted the semantic relations in the English WordNet PWN, but none of the relations were automatically translated from English. That is, in constructing the semantic relations, possible relations between Turkish synsets were listed based on their dictionary translations in English PWN and the relations between the English synsets in PWN. Then, human annotators checked these relations manually; the correct relations were added to KeNet whereas the incorrect ones were eliminated. For example, as Table 4 shows, two candidate antonymy relations for the Turkish synset *ağır* "heavy" were listed: *yeğni hafif* "light" and *hafif* "light". These candidate antonyms were extracted based on the English translations of the Turkish synsets *ağır* and *hafif* and the existing antonymy relations between their English equivalents "heavy" and "light". For this example, the antonymy relation between *ağır* and *yeğni hafif* were correct and it was added to KeNet, but the antonymy relation between *ağır* and *hafif* were not correct and it was not kept. This procedure was followed for all the semantic rela-

Table 3: Semantic Relations in KeNet

Semantic Relation	Example	# of Mappings
Hypernymy	gürgen - ağaç "alder tree - tree"	45,389
Derivational Relatedness	kitap - kitaplık "book - bookshelf"	39,682
Domain Topic	işlemci - bilgisayar bilimi "processor - computer science"	15,366
Part Holonymy	kulp - bardak "handle - glass"	2,718
Antonymy	sıcak - soğuk "hot - cold"	1,884
Instance Hypernymy	Antartika - kıta "Antarctica - continent"	1,345
Member Holonymy	ebeveyn - aile "parent - family"	862
Substance Holonymy	hidrojen - su "hydrogen - water"	367
Attribute	ılık - sıcaklık "warm - heat"	226

Table 4: Building Antonymy Relations in KeNet based on PWN

TR synset	Sense	TR synset	Sense	ENG synset	Sense	ENG synset	Sense
ağır	tartıda çok çeken	yeğni hafif	tartıda ağırlığı az gelen	heavy	of comparatively great physical weight or density	light	of comparatively little physical weight or density
ağır	tartıda çok çeken	hafif	kalınlığı veya yoğunluğu az olan	heavy	of comparatively great physical weight or density	light	of comparatively little physical weight or density

tions. However, building hypernymy relation was more complicated and it included some additional steps. The details of hypernymy relation in keNet is presented next.

4.1 Hypernymy

For now KeNet has a semantic hierarchy tree only for the noun category. In this section, we explain how we built the hypernymy relations only for the nouns in KeNet.

We started building hypernymy relations based on the Turkish Estate WordNet (Parlar et al., 2019) and Turkish Tourism WordNet (Arican et al., 2020) because these wordnets were built based on KeNet but they were much smaller than KeNet in terms of their scope due to being domain-dependent. Both Turkish Estate WordNet and Turkish Tourism WordNet had synsets that were and were not present in KeNet. Thus, we first created two lists on Google Sheets: a list with the synsets that occurred in both Turkish Estate WordNet and KeNet and another with those that occurred in both Turkish Tourism WordNet and KeNet. This enabled us to focus on a smaller size of synsets from KeNet in the first place that belonged to the same domain. Then, the corresponding English synsets from PWN were then found for the Turkish synsets in these lists by human annotators and placed next to the Turkish synsets. Based on the hypernymy relations between the corresponding English synsets, hypernymy relations between the Turkish synsets were created. This enabled us to have small hierarchical trees for the synsets in KeNet.

After building some preliminary hypernymy relations in domain-dependent wordnets, we decided to start forming the comprehensive hierarchical tree from the top. Therefore, our second step was to find the nodes that would occur on the top of the hierarchical tree. In order to find these nodes, we extracted a list of approximately 700 words that repeated the most in the hypernymy relations in Turkish Estate WordNet and Turkish Tourism WordNet. When we had the

Table 5: Constructing Hypernymy Relations in KeNet based on PWN

EN ID	EN synset member	EN Definition	TR synset member	TR Definition1	TR Definition2
ENG31-00001740-n	entity	that which is perceived or known or inferred to have its own distinct existence	varlık	Doğumla ölüm arasında yaşanan süre	Her türlü taşınır ve taşınmaz maddi varlık
ENG31-00001930-n	physical entity	an entity that has physical existence	fiziksel varlık		
ENG31-00002684-n	object physical object	a tangible and visible entity	nesne	Belli bir ağırlığı ve hacmi, rengi, maddesi olan her türlü cansız varlık, şey, obje	Geçişli fiili bütünleyen yalın veya belirtme durumunda bulunan tümleç
ENG31-00003553-n	whole unit	an assemblage of parts that is regarded as a single entity	bütün	Birlik, birleşmiş olma durumu	Bölünmezliği içeren yalın bütün
ENG31-00022119-n	artifact artefact	a man-made object taken as a whole	yapı	Türlü amaçlarla kullanılan, insan yapısı, taşınabilir cansız nesnelerin bütünü	Yapılmakta olan konut, yol, köprü vb. inşaat
ENG31-04348764-n	structure construction	a thing constructed	yapı	Yapılmakta olan konut, yol, köprü vb. inşaat	Canlı bir varlığın ruh veya beden özelliklerinin tümü, bünye, strüktür
ENG31-03302664-n	establishment	a public or private structure including buildings and equipment for business or residence	kurum	Bir kurum veya kuruluşun yönetildiği yer veya makam	Ocak bacalarında biriken veya çevrede savrulan kalın is
ENG31-03959296-n	place of business	an establishment where business is conducted, goods are made or stored	işletme	İstihdam edilen kişilerin çalıştığı, üretimin yapıldığı yer	İş yeri
ENG31-03753653-n	mercantile establishment retail store	a place of business for retailing goods	satış noktası	Perakende satış yapan esnafın, küçük zanaat sahiplerinin satış yaptıkları veya çalıştıkları yer	

list for the most repetitive synsets in these wordnets, human annotators formed hypernymy relations among these synsets. At the end of this process, the majority of the nodes that would appear on top of the hierarchical tree in KeNet, e.g., *varlık* ”entity” (see Table 5) was formed. In this process, we also noticed that some of the uppermost nodes that were present in PWN did not have equivalents in Turkish. For example, there was no corresponding synset for ”physical entity” in KeNet (see Table 5). When such synsets were crucial to have in the hierarchical tree, new synsets were created for those in KeNet by translating them from PWN to Turkish.

Thirdly, as in the construction of other semantic relations, possible hypernymy relations between the synsets in KeNet were extracted based on their dictionary translations in English PWN and the relations between the synsets in PWN. A list of these possible hypernymy relations was listed on Google Sheets and checked by human annotators one by one. The relations that were correct were added to KeNet. This step again allowed us to have small hierarchical trees that were to be combined in order to form a single hierarchical tree.

The forth step was thus to place these small trees that were created in the first and the third steps under the topmost nodes. To exemplify, in this step, the synsets *nesne* "object" and *bütün* "whole unit" that had a hypernymy relation between from earlier processes was placed under *fiziksel varlık* "physical entity", as shown in Table 5.

After the placement of small hierarchical trees into a single tree, we were left with free-standing synsets that were not currently attached to any node in the hierarchical tree. In the final step, these free-standing synsets were listed in a java program where they were attached to their hypernyms one by one by human annotators. The biggest problem at this stage was that there were no guides to follow and the annotators had to decide where to place free-standing synsets in the tree.

Several strategies have been employed by our team to successfully place the free-standing items. The first strategy was to rely on the definitions of the synsets. For instance, a free-standing synset such as *etimoloji* "etymology", which is defined as "a branch of linguistics", would be placed under *dilbilim* "linguistics". Another useful approach was to refer to PWN to see where the corresponding English synset was placed in PWN. Following from the previous example, if the synset "etymology" was placed under 'linguistics' in PWN, its counterpart in Turkish, i.e., *etimoloji* "etymology", can be placed under the equivalent of its hypernym in Turkish, i.e., *dilbilim* "linguistics". A third strategy was to perform a domain-specific top-down analysis. That is, when we encountered a synset in the tree that could possibly host several hyponyms, we searched for its possible hyponyms in the list of free-standing items and placed them under their hypernyms. For example, when we came across with the synset *dilbilim* "linguistics", we looked for its possible hyponyms such as syntax, semantics or phonology and placed them under it. This was especially useful for domain-related synsets. The last strategy was to consider the sister synsets of the synset in question. If we were not sure of the correct hypernym of a given synset, but placed its sister synset in the hierarchical tree in earlier stages, we would find the hypernym of its sister synset and place the synset in questions under its hypernym. Again, following from the same example, knowing that *sentaks* "syntax" and *etimoloji* "etymology" were sister synsets, a simple search for the hypernym of *sentaks* "syntax" would provide us with the correct hypernym for the synset in question: *dilbilim* "linguistics".

In addition to these strategies, one advantage we had was that as the same team members worked on the same hierarchy for extended periods of time, e.g., 15 hours per week for 5-6 months, they became familiar with the general structure of the tree and placing the free-standing synsets into the tree became easier with practice.

5 Interlingual Relations

With the creation of wordnets in several languages, the idea of matching these wordnets to English and/or to one another has gained importance since the linking of wordnets across languages would enable us to use these resources in machine translation.

As discussed in Section 1, there are two available wordnets for Turkish, which are TR-wordnet of BalkaNet and KeNet. Having been created as part of BalkaNet (Tufis et al., 2004), TR-wordnet of Balkanet was integrated to PWN through an interlingual index (ILI) (Bilgin et al., 2004). As opposed to TR-wordnet, in our work, we used the merge approach and matched the synsets in KeNet with those in PWN manually (Bakay et al., 2019b). Additionally, as in building intralingual semantic relations, we extracted candidate multilingual relations based on dictionary translations and listed these potential interlingual relations on Google Sheets. Two human annotators checked the accuracy of these candidate relations one by one.

In TR-wordnet of BalkaNet, only one-to-one mappings between Turkish synets and the English synsets in PWN were included due to the use of ILI (Bilgin et al., 2004). In KeNet, on the other hand, although one-to-one matchings

made up the majority of the interlingual relations between Turkish and English, one-to-many mappings between the two languages were also included (Bakay et al., 2019b). One-to-many mappings between KeNet and PWN were mainly used when selecting a single synset in one of the languages was not possible. That is, the two most common cases where one-to-many matchings were needed were (i) when a sense distinction in English is not reflected in Turkish, e.g., the English synsets "inequitable unjust", "unfair unjust", "unrighteous", "undue unjustified unwarranted" and "unlawful wrongful" all correspond to *haksız nahak* in Turkish or (ii) vice versa; when a sense distinction that is present in Turkish is not reflected in English, e.g., the Turkish synsets *hafiflemek*, *hafiflemek azalmak* and *kırılmak yatışmak* all correspond to the English synset "abate let up slack off slack die away".

Overall, the inclusion of one-to-many matchings in KeNet allowed for a better matching between the two languages.At the end of the manual interlingual matching between KeNet and PWN, 19,398 synsets in Turkish were associated with 19,208 synsets in English. 3,500 were one-to-one and 1,250 of them were one-to-many matchings. Furthermore, out of 5,000 most frequent senses in English, 4,417 (88%) were matched with their Turkish equivalents in order to have the matchings of the synsets that are most commonly used.

6 Challenges

We have faced many resource-related and language-related challenges in creating KeNet. We faced an important resource-related problem in the creation of synsets due to not having a Turkish dictionary that marks the synonymy relation in a systematic fashion. We also encountered language-related problems in constructing synsets. For example, some synsets in KeNet included synset members with different POS categories. The reason for this was that in Turkish some words can be used in different POS categories with similar meanings. This resulted in wrong groupings of synonyms, which we had to deal with in the split process. In building the hypernymy relations, one of our biggest challenges was that some synsets in PWN did not have corresponding synsets in KeNet. When this was the case in the upper parts of the tree, we came up with new synsets that would connect the lower parts of the tree with the upper parts as the tree would otherwise be missing some transitionary nodes. Moreover, in building interlingual relations, we realized that having only one-to-one mappings would not be a correct matching for Turkish and English and hence, we decided to include one-to-many mappings, which was a time-consuming process to conduct. Lastly, we had to devote a huge amount of time and labor in all the stages in creating KeNet as in most of the stages, we could not refer to earlier work and conducting the stages automatically would be misleading, thus human annotators had to work manually. This was mainly due to the structural and lexical differences between Turkish and other highly-investigated languages such as English.

7 Conclusion

In this paper, we presented the process for creating the Turkish wordnet KeNet and discussed the challenges that we encountered. Our biggest challenge was to work on a low-resource language since most of the studies in the field focus on highly-investigated languages like English. The differing morphological and syntactic properties of Turkish also prevent us from adopting the exact approaches used in these studies. Although structural differences have created problems mostly for morphological analyses, we also encountered cases where semantic mappings and/or relations in English could not be directly copied to Turkish. This discrepancy was observed in both intra- and inter-lingual semantic relations. Overall, unavailability of sources which can be easily linked to Turkish forced us to include a huge amount of manual annotations.

KeNet has been used as a source in other NLP studies on Turkish such as Turkish PropBank TRopBank (Kara et al., 2020), Turkish SentiNet HisNet (Ozcelik et al., 2020), Turkish FrameNet (Marsan et al., 2020) and domain wordnets for Estate (Parlar et al., 2019) and Tourism (Arican et al., 2020). Having a common wordnet source for different NLP studies in a given language can be a great advantage for the linking of these sources. That is, when the sense categorization between two different sources of a given language do not match well, as in the case of English PropBank and English WordNet PWN (Bakay et al., 2019a), the linking of the available sources becomes more challenging.

References

B. N. Arican, M. Ozcelik, D. B. Aslan, E. Sarmis, S. Parlar, and O. T. Yıldız. 2020. Creating Domain Dependent Turkish WordNet and SentiNet. In *Proceedings of GWC 2020*.

O. Bakay, B. Avar, and O. T. Yıldız. 2019a. Comparing Sense Categorization between English PropBank and English WordNet. In *Proceedings of GWC 2019*, pages 307–314.

O. Bakay, O. Ergelen, and O. T. Yildiz. 2019b. Integrating Turkish WordNet KeNet to Princeton WordNet: The Case of One-to-Many Correspondances. In *ASYU*.

O. Bakay, O. Ergelen, and O. T. Yildiz. 2019c. Problems caused by semantic drift in wordnet synset construction. In *UBMK*.

O. Bilgin, O. Cetinoglu, and K. Oflazer. 2004. Building a wordnet for Turkish. *Romanian Journal of Information Science*, 7:163–172.

M. Derwojedowa, M. Piasecki, S. Szpakowicz, M. Zawisławska, and B. Broda. 2008. Words, Concepts and Relations in the Construction of Polish WordNet. In *Proceedings of GWC 2008*, pages 162–177.

R. Ehsani, E. Solak, and O.T. Yildiz. 2018. Constructing a WordNet for Turkish Using Manual and Automatic Annotation. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 17(3).

R. V. Fjeld and L. Nygaard. 2009. Nornet - a monolingual wordnet of modern Norwegian. In *NODALIDA 2009 workshop: WordNets and other Lexical Semantic Resources - between Lexical Semantics, Lexicography, Terminology and Formal Ontologies*, pages 13–16.

N. Kara, D. B. Aslan, B. Marsan, K. Ak, and O. T. Yıldız. 2020. TRopBank: Turkish PropBank V2.0. In *Proceedings of LREC 2020*, pages 2763–2772.

K. Linden and L. Carlson. 2010. Construction of a FinnWordNet. *Nordic Journal of Lexicography*, 17:119 – 140.

B. Marsan, N. Kara, M. Ozcelik, B. N. Arican, N. Cesur, A. Kuzgun, E. Saniyar, O. Kuyrukcu, and O. T. Yıldız. 2020. Building the Turkish FrameNet. In *Proceedings of GWC 2020*.

G.A. Miller. 1995. WordNet: a lexical database for English. *ACM Communications*, 38:39–41.

M. Ozcelik, B. N. Arican, O. Bakay, E. Sarmis, N. B. Bayazit, O. Ergelen, and O. T. Yıldız. 2020. HisNet: A Polarity Lexicon based on WordNet for Emotion Analysis. In *Proceedings of GWC 2020*.

S. Parlar, B. N. Arican, M. Erkek, K. Cayirli, and O. T. Yildiz. 2019. Emlak Alanına Özgü Kelime Ağı. In *Proceedings of the Signal Processing and Communication Applications Conference*.

B. S. Pedersen, S. Nimb, J. Asmussen, N. H. Sørensen, L. Trap-Jensen, and H. Lorentzen. 2009. DanNet: the challenge of compiling a wordnet for Danish by reusing a monolingual dictionary. *Language resources and evaluation*, 43:269–299.

B. Sagot. 2008. Building a free French wordnet from multilingual resources. page 24. ACM.

D. Tufis, D. Cristea, and S. Stamou. 2004. BalkaNet: Aims, methods, results and perspectives. a general overview. *Romanian Journal of Information Science*, 7:9–43.

P. Vossen. 2007. EuroWordNet: A multilingual database for information retrieval. In *DELOS workshop on Cross-Language Information Retrieval*.