Morphological Resources for the Study of Turkish Derived Nouns

Yağmur Öztürk
CRIT,
Université de
Franche-Comté
yagmur.ozturk@edu.univ-fcomte.fr

Izabella Thomas
CRIT,
Université de
Franche-Comté
izabella.thomas@univ-fcomte.fr

Snejana Gadjeva
CREE,
Institut National
des Langues et
Civilisations Orientales
snejana.gadjeva@inalco.fr

Abstract

In terms of morphological resources, Turkish turns out to be an underresourced language. In particular in the field of Natural Language Processing (NLP), there are not enough resources that sufficiently (and systematically) describe Turkish derivational morphology, especially concerning semantic aspects of the derivational process. The research aims to describe and use existing resources and studies to develop an NLP tool for Turkish nominal derivation. The first part of our study presents the current morphological analysers revealing a gap in derivational morphology of nominals. We then discuss how derivational morphemes, specifically nominal morphemes, are rendered in linguistic studies and the problems it poses for a systematic study. Finally, we introduce *Semantürk*, which is an ontology of semantic categories, and *DerivBaseTR*, which is a morpheme database with specific features, as the formalised resources we created for a systematic study of noun-to-noun morphemes.

1 Introduction

The study we propose came into existence following research in a morphosemantic project for the processing of Turkish derived nouns. The primary goal is to create a morphosemantic analyser that describes the internal structure of derived nouns as well as explicits the semantic role of each detected morpheme in the derived noun. A similar tool, *DériF* (Namer, 2002) exists in French with a "pseudo-definition" output as shown in (1).

(1) appauvrissement/NOM à [[a [pauvre ADJ] VERBE] ment NOM],
(appauvrissement/NOM, appauvrir/VERBE, pauvre ADJ),
"(Action – résultat de l'action) de appauvrir"
en. (Action – result of the action) of impoverish

Although Turkish is an agglutinative language, with a high degree of regularity and productivity in its derivational processes, as stated in Section 2, there is currently no morphosemantic analyser for Turkish. Additionally, there are no computerised resources, such as a morpheme database, that can be used for the development of a morphosemantic analyser. However, as discussed in Section 3.1, most of the existing analysers yield excellent results in inflectional morphology. Regarding the analysis of the derivatives, the analysers primarily focus on the derivation of verbs. In contrast, the analysis of nominal derivatives remains notably limited.

To enhance the formal and semantic analysis of nominal derivatives, it is necessary to build formalised resources. Our approach starts with the investigation of the representation of nominals in the descriptive linguistic studies and Turkish textbooks. Nonetheless, our analysis revealed another issue; a lack of formalised description of the nominal morphemes in Turkish derivational morphology. This matter is further discussed in Section 3.2.

The lack of a formalised description of derivational morphemes may also explain the lack of a morphosemantic analyser, especially for nominal derivatives. Therefore, we established a methodology to standardise the representation of nominal morphemes and their description, as presented in Section

Proceedings of the 4th Int. Workshop on Resources and Tools for Derivational Morphology (DeriMo 2023), pages 89–98,
Dubrovnik, Croatia, 5-6 October 2023.

4. This approach considers the formal, categorial and semantic aspects of the morphemes to enable their automatic processing. It then resulted in the development of two different resources, built in an Open Science perspective[1]. These resources are *Semantürk*, an ontology of semantic categories and *DerivBaseTR*, a database of Noun-to-Noun (N-to-N) morphemes and their corresponding descriptions.

2 Turkish Derivational Morphology

2.1 Formal Level

Turkish is an agglutinative language where suffixation is the predominant morphological process. In N-to-N derivation, derivational morphemes are all bound morphemes attached to a free morpheme, which may be either a simple word (zero derivation) or a complex word (having one or more derivational morphemes). Exceptions aside, the root, whether complex or not, does not change. Typically, the suffix is concatenated directly to the root word, as in (2).

(2) göz (en. eye)
gözlük (en. eyeglasses)
gözlükçü (en. optician)
gözlükçülük (en. opticianry)

However, the majority of bound morphemes conform to the vowel and consonant harmony rules, leading to allomorphy. These rules do not alter the semantic or grammatical features of the morphemes. Instead, they trigger formal and phonological adaptations of the morpheme. Particular conventional writing rules are used to represent all allomorphs in a single form. Firstly, if the morpheme starts with a capital consonant, it denotes consonant harmony. The uppercase consonant corresponds to a voiceless and voiced consonant pair as shown in Table 1. In cases where the root word ends with a voiceless consonant, such as the word *kitap*, the suffix begins with the voiceless consonant of the pair, as in the suffix -Cİ (3a). Otherwise, the suffix begins with the corresponding voiced consonant (3b).

Voiceless	**Voiced**	**Symbol**
ç [ʧ]	c [ʤ]	C
t [t]	d [d]	D
k [k]	g [g]	G

Table 1: Consonant pairs in consonant harmony

(3) a. kitap-çı
book-Cİ
“bookseller”
b. şarkı-cı
song-Cİ
“singer”

Morphemes can undergo either simple or complex vowel harmony rules. Simple vowel harmony is usually represented by the symbol A[2]. It applies to the two open vowels *a* and *e*. If the last syllable of the root word contains a front vowel, such as *e, i, ü or ö*, then the vowel in the suffix will be *e*, as in (4a). Otherwise, it will be the vowel *a* (4b).

(4) a. Türk-çe
Turk-CA
“Turkish language”

[1]Our resources are to be accessible and usable for future works.
[2]In some instances, the letter E represents simple vowel harmony. Here, we use the symbol A.

b. Fransız-ca
French-CA
"French language"

Complex vowel harmony is denoted by the letter İ[3] with four possible closed vowels: *i*, *ü*, *ı* or *u*. If the last syllable contains a closed vowel, then the vowel in the suffix will be identical (5a). Else, the vowel in the suffix will be its closed vowel counterpart (5b). Table 2 displays the possible combinations that arise due to the complex vowel harmony rule.

Last vowel	**Suffix vowel**	**Last vowel**	**Suffix vowel**
a [a]	ı [ɯ]	e [e]	i [i]
ı [ɯ]	ı [ɯ]	i [i]	i [i]
u [u]	u [u]	ü [y]	ü [y]
o [o]	u [u]	ö [œ]	ü [y]

Table 2: Complex vowel harmony

(5) a. ayakkabı-lık
shoe-lİk
"shoe cupboard"

b. göz-lük
eye-lİk
"eyeglasses"

2.2 Categorial Level

Derivational morphemes, unlike inflectional morphemes, allow for the creation of new lexemes, mainly characterised by a possible change in the word class. A significant number of morphemes come into play in Turkish nominalisation, such as N-to-N morphemes, Verb-to-Noun morphemes, Adjectives-to-Noun morphemes, and so on. However, our research focuses on N-to-N derivation which limits our scope to the semantics of nominals.

The distinction between word classes is very significant since the semantics of the morphemes closely correlates to the grammatical class of either the root or the derivative, as explained in Section 2.3. Turkish linguistic studies, particularly in morphology, offer a different perspective on word class distinction in comparison to the word class distinction put forth in Western linguistic studies. Derivational morphemes are classified into two separate categories, verbs (tr. *fiil*) and nouns (tr. *ad* or *isim*[4]). The latter includes numerals, adjectives, adverbs and pronouns (further discussed in Section 3.2 and Section 4).

Furthermore, this classification of nominal morphemes reflects their polycategorial nature. This is because many morphemes classified as nominal morphemes can result in derivatives of various word classes (nouns, adjectives, adverbs, or sometimes pronouns). (6) clearly shows the polycategoriality of the morpheme -CA as it can, attached to the noun *kadın* (en. woman), derive a new noun (6a), adjective (6b) or adverb (6c).

(6) a. kadın-ca → N-to-N
woman-CA
"the language of women"

b. kadın-ca → N-to-Adj.
woman-CA
"womanlike"

[3]In some instances, the letter I or H represents complex vowel harmony. Here, we use the symbol İ.

[4]These terms are synonymous and can be used interchangeably whithin the context of a nominal lexeme or a nominal class that covers different categories.

c. kadın-ca → N-to-Adv.
woman-CA
"womanly"

Moreover, a change in the meaning of the lexemes in (6) can be noticed, indicating a direct link between morpheme meaning and grammatical category. To minimize ambiguity in the analysis of nominal morphosemantics in Turkish derivational morphology, we restrict our analysis to N-to-N derivation.

2.3 Semantic Level

A morpheme is traditionally defined as the smallest meaningful unit of a language. This approach is especially appropriate for the description of agglutinative languages. As mentioned earlier, derivational morphemes enable the formation of new lexemes. This leads to a change in the word class, but it can also lead to a change in the meaning, as shown in (6). Meaning can change significantly, which is the case between the nominal form (6a) which refers to an abstract entity and the adjectival form that denotes a more qualitative concept in (6b). However, it can also be more ambiguous as in (6b) and (6c), with both examples showcasing the qualitative aspect.

It is important to note that morpheme polysemy is not necessarily related to polycategoriality. In fact, a morpheme that creates N-to-N derivatives can produce entirely different meanings. In (7), the morpheme -lİk first produces a concrete material object designated by the noun (7a). However, it also creates an abstract noun (7b). The combination of the morphemes -Cİ-lİk results in the abstraction of the lexeme, noted as a recurrent distributive pattern. Therefore, the meaning of the morpheme in question can also be context-dependent. This can be observed with various morphemes, cf. example (4) given previously, where the addition of the suffix -CA to a noun denoting nationality results in a noun denoting the language or the dialect spoken in that nation.

(7) a. göz-lük
eye-lİk
"eyeglasses"

b. gözlükçü-lük
optician-lİk
"opticianry or the occupation of an optician"

A semantic category can also be conveyed by different morphemes, resulting in synonymous or quasi-synonymous morphemes. Typically, the diminutive morphemes -Cİk and -cAğİz both convey a sense of a pity felt by the speaker towards the referred entity, as shown in (8).

(8) a. kedi-cik
cat-Cİk
"the poor little cat"

b. adam-cağız
man-cAğİz
"the poor little man"

Therefore, the correlation between form and meaning can be qualified as a many-to-many relationship, that is a morpheme can be associated to one or more semantic categories, just as a semantic category can be associated with one or more morphemes. It can be either dependent on the category or its distribution.

Lexicalisation is also a phenomenon present in the Turkish language. Some derivatives can show a high degree of lexicalisation. Some morphemes can be synchronically difficult to detect and more root dependent where many others are completely distinct and are independent from the root word. Lexicalised derivatives are not taken into consideration in this research as the morpheme in these cases loses its semantic component and requires an etymological analysis.

3 Resources and Studies in Turkish Nominal Morphology

3.1 Nominal Morphemes in NLP Tools and Resources

A lot of research on Turkish language is currently being conducted in the fields of NLP (Oflazer and Saraçlar, 2018; Çöltekin et al., 2023). One of the issues we met concerns the availability of existing resources as was highlighted in Çöltekin et al. (2023): "The locations of published resources are not always stable and/or permanent. The URLs indicating the location of the resources in papers or on the webpages of the authors or institutions are not always maintained and the resources often disappear after publication. Although our efforts to reach out to the authors/creators of the resources often yielded positive results, it is desirable to diminish these barriers to keep up with the fast-paced research community."

While there are numerous studies available for the French language, e.g. Missud et al. (2020); Mailhot et al. (2020); Varvara et al. (2022); Hathout and Namer (2022), to our knowledge, very few focus on the derivation of Turkish nouns, and even less to the particular subject of N-to-N derivation. Among the most well-known NLP tools in Turkish, there is *Zemberek*[5] (Akın and Akın, 2007), an open-source Java library (no longer updated). The morphology processing section offers various analyses, i.e. single word morphological analysis, stemming and lemmatisation, contextual ambiguity resolution, and word generation. However, the processing mainly results in inflectional analyses, with word generation producing an output of inflected forms of the entry word, as shown in the examples of outputs for the entry *ev* (en. house) in (9).

(9) a. evime
ev-im-e
house-1SG.POSS-DAT
"to my house"

b. evimde
ev-im-de
house-1SG.POSS-LOC
"in my house"

Another well-known tool is *TRmorph*[6] (Çöltekin, 2010), an open-source morphological analyser, written using a Foma Finite State Transducer (FST) compiler, which produces a list of possible analyses for an out-of-context lexeme. In addition to a complete inflectional analysis, it accurately identifies verbal derivational morphemes. However, it only identifies a short list of the most productive nominal morphemes. 17 derivational suffixes with nominal roots are described in the resource. Only seven of these (four of which have been regrouped) are annotated as N-to-N suffixes: -lİk$_{\langle lik \rangle}$, -Cİk$_{\langle dim \rangle}$, -cAk$_{\langle dim \rangle}$, -(İ)cAk$_{\langle dim \rangle}$, -cAğız$_{\langle dim \rangle}$, -Cİ$_{\langle ci \rangle}$, -gil$_{\langle gil \rangle}$, which is a rather small sample of nominal suffixes. However, the part-of-speech categorisation of the morphemes by *TRmorph* does not exactly match ours. For instance, unlike in our classification, -CA is not categorized as an N-to-N morpheme in this analyser.

A new open source Java library, *Turkish Morphological Analyzer*[7] (Yıldız et al., 2019), was released in 2019. Again, only four N-to-N suffixes are identified: -Cİ, -Cİk, -(İ)ncİ, lİk. However, they added specific tags, `AGT`, `DIM`, `ORD` and `NESS` respectively, representing a possible semantic role of these derivational suffixes.

Trnlp[8] (Bayol, 2018) is an ongoing project, an open source Python API. It has several components including lemmatisation, stemming, spellchecking and tokenisation. It identifies a more diverse set of nominal derivational suffixes. Although it gives good results, it still needs improvement: 1. the suffixes listed in the N-to-N section are not all correct (e.g. -m is included but actually corresponds to the first person possessive suffix); 2. among the 27 suffixes listed as N-to-N suffixes, several do not result in nominal derivatives (e.g. -sİ results in adjectival derivatives); 3. the output of the analysis is not disambiguated. Nevertheless, it produces an analysis on 15 nominal suffixes, which is one of the best

[5] https://github.com/ahmetaa/zemberek-nlp
[6] https://github.com/coltekin/TRmorph
[7] https://github.com/olcaytaner/TurkishMorphologicalAnalysis
[8] https://github.com/brolin59/trnlp

results we have observed so far.

As our project is carried out in an Open Science perspective, we did not analyse publicly unavailable resources. Some examples are PC-KIMMO-based analyser (Oflazer, 1994), SakMP (Sak et al., 2008), ITU Turkish NLP Web Service (Eryiğit, 2014).

Not only are there very few tools available for Turkish derivation in nominal morphology, but there are also no available computerised morphological resources. To our knowledge, there are no accounts of:

- dictionaries with morphological descriptions,
- exhaustive inventories of morphemes, whether formalised or not.

For instance, while the French Wiktionary has 1,935,402 entries, the Turkish Wiktionary has only 3,958 entries[9] and therefore does not provide a usable dataset for any morphological analysis. Moreover, it does not contain any information on derivatives. As shown in Figure 1[10], there is only the "definition" (or a synonym of the word as given in this example) of the word whereas *fakirlik* (en. "poverty") is a noun derived from *fakir* (en. "poor") with a very productive suffix -lİk.

Figure 1: Example from Turkish Wiktionary

To overcome the scarcity of easily accessible and available resources in derivational morphology from an NLP perspective, we collected data from various linguistic studies in order to design and then implement new computerised resources. However, this is not a trivial task as we faced several difficulties originating from the descriptions proposed in these studies, as discussed in the following subsection.

3.2 Nominal Morphemes in Linguistic Studies

The linguistic books we examined were Turkish (Adalı, 2004; Korkmaz, 2014; Boz, 2015), French (Bazin, 1994) and English grammar books (Göksel and Kerslake, 2005) as well as a few Turkish textbooks for learners (Bozdémir, 1991; Erikan et al., 2008). We have also looked at the two other sources, an article by Akçataş and Taşdemir (2020), and a master's thesis by Ozturk (2016), focusing on the morphosemantics of Turkish morphemes. However, new difficulties arose during the data collection. These difficulties were more or less common to all of the above-mentioned studies as listed below.

1. Lack of descriptions in alphabetically ordered lists

Descriptive linguistic studies of the Turkish language mainly consist of a set of morphemes listed alphabetically with instances of derived words without any explanation on the morphotactics or the semantic value of the morpheme, e.g. Adalı (2004).

2. Difference in morpheme categorisation

As introduced in Section 2.2, Turkish linguistic studies introduce a different word class categorisation, describing morphemes of different word classes in the section dedicated to nominal morphology. For example, (10), extracted from the section "Suffixes that attach to nominals to form nominals" in Göksel and Kerslake (2005), is a morpheme that produces an adjective. We can also find suffixes attaching to or deriving adverbs and pronouns in addition to nouns and adjectives.

(10) *-(A)C* Attaches to nouns to form adjectives: *anaç* 'motherly', *kıraç* 'infertile'

This is a traditional categorisation of word classes in the literature of Turkish linguistics (and other Turkic linguistics in general). The inclusion of adjectives, adverbs, pronouns and numerals in a single nominal class reflects the close interaction of these classes and their ability to function as nominal

[9] https://fr.wiktionary.org/wiki/Wiktionnaire:Statistiques (last accessed: June 21st, 2023)
[10] https://tr.wiktionary.org/wiki/fakirlik (last accessed: June 21st, 2023)

elements, whether the lexeme is polycategorial or not. Syntax, in Turkish, has a relatively flexible lexeme order so that nouns, adjectives, adverbs and pronouns can occur in different positions, including a nominal position, i.e. adjectives can be used as nouns in a sentence, without any formal indication on the functional change apart from the syntactic position. In addition, they can easily function as nouns and take nominal inflectional suffixes.

3. Non-exhaustiveness

The number of morphemes described varies from study to study, as illustrated by a few examples in Figure 2. Introductory studies (Bazin, 1994) or pedagogical textbooks (Bozdémir, 1991; Erikan et al., 2008) for language learning do not have complete descriptions of derivational morphemes. They tend to focus on a few of the most productive ones. Among the remaining linguistic studies, Göksel and Kerslake (2005) and Korkmaz (2014) have the highest number of morphemes described[11]. This variation is due to different approaches to morpheme description. Indeed, Korkmaz (2014) also describes dead affixes in lexicalised forms, which is on a borderline with a diachronic approach to morphemes. Göksel and Kerslake (2005) include many loaned morphemes mainly of Arabic or Persian origin. Incoherence in morpheme description between different sources also explains the difference in the number of morphemes described.

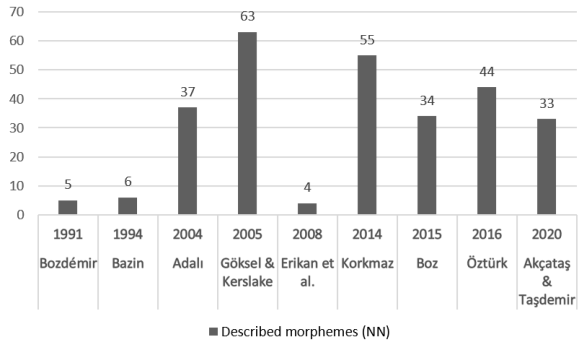

Figure 2: Number of morphemes per source

4. Incoherence in morpheme description

Different sets of references show discrepancies in different descriptive aspects. For example, there is a difference in the description of the suffix -sAl in Göksel and Kerslake (2005) and Korkmaz (2014). On one hand, in Göksel and Kerslake (2005), this suffix is described as mainly a Noun-to-Adj. suffix which, in rare cases, also forms nouns: *kumsal* 'sandy beach'. On the other hand, Korkmaz (2014) clearly states that the suffix is not related to the form *sal* in *kumsal*. We can also see incongruent morpheme representations across various sources, as for the morpheme -cAğIz. In Korkmaz (2014), we have -CağIz, whereas in Adalı (2004), the morpheme -IZ (ız, iz, uz, üz) is a separate morpheme entry attached to stems ending with the morpheme -CAK (-cak, -cek, çak, çek), including non-grammatical stems such as **çocukcak*, **kızcak*, etc. Another difference we noted in most of the sources, is that each morpheme is described with a different set of information throughout the same source. The semantic function of a suffix is explained for some of the morphemes, as in (11) (Göksel and Kerslake, 2005). However, some suffixes are described only from a grammatical point of view (10). The description is therefore unsystematic and may be incoherent.

(11) *-Daṣ/Deṣ* Added to nouns to form nouns denoting possessors of a shared attribute: *yandaṣ* 'supporter', *kardeṣ* 'sibling' (from *karın* 'abdomen'), *meslektaṣ* 'colleague (i.e. person of the same profession)'.

[11] A few loaned prefixes are mentioned in several of the sources, but are not further studied or described.

4 Formalised Morpheme Description in Machine-readable Resources

As discussed earlier, we assume that the semantics of N-to-N morphemes can be identified using existing linguistic sources. In this section, we present the processing steps of our methodology for the development of two formalised resources, *DerivBaseTR* and *Semantürk*. Figure 3 illustrates the workflow for the creation of these two independent resources.

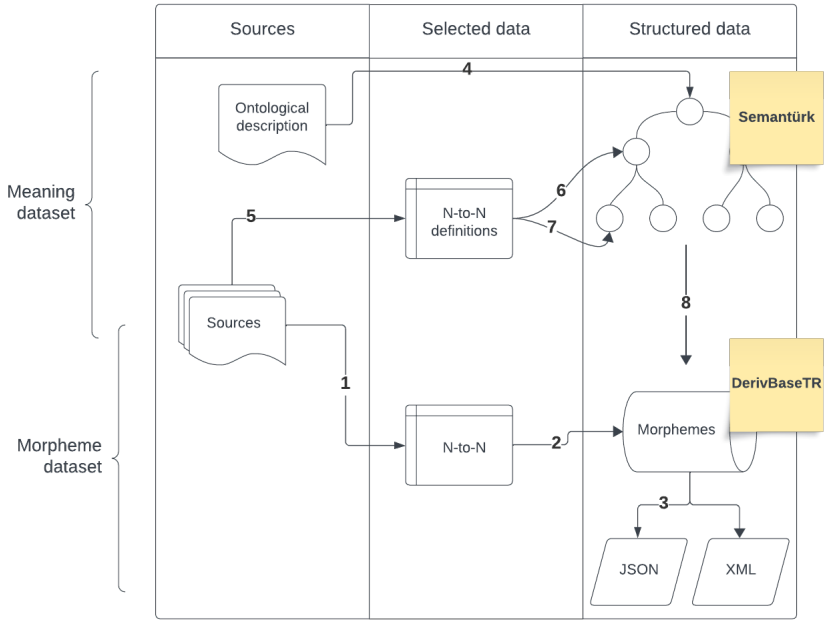

Figure 3: Processing workflow

1/ Following the examination of the existing sources in Turkish linguistics, we extracted the morphemes producing N-to-N derivation. This task was complicated by the different morpheme categorisation in Turkish linguistics[12]. Some linguists claim that the categorial flexibility of the lexemes is a proof of a functional variation rather than a categorial variation. That is, the categorial function of a lexeme is syntax-dependent. However, it can also be argued that lexemes inherently carry categorial information, so that their category can be identified in the lexicon[13]. In fact, any given word in a dictionary, such as Türk Dil Kurumu Sözlükleri[14] (TDK sözlükleri, *the dictionaries of the Turkish Language Association*), is assigned a grammatical category per meaning. After excluding all dead suffixes which result in lexicalised forms, and selecting suffixes from sources where the grammatical category (or categories) of the root and the derivative were already given as nouns, we studied the examples of derivatives for the unclassified ones. We proceeded to the selection by identifying the "primary function" (Göksel and Kerslake, 2005) of the examples of derivatives given in the morphemes description with the help of the TDK dictionaries. 2/ We then formalised and stored all the information given on the selected morphemes in an Excel file. In this way, we collected the morpheme representation[15], its allomorphs, its origin, and examples of derivatives. We also added *Base category* and *Derived category* entries in the morphemes' descriptive properties to ensure the possibility of adding other grammatical categories. We developed a first version of *DerivBaseTR* with a formalised description of the morphemes at both the formal and categorial levels, offering the possibility of filtering or ordering the morphemes by features. 3/ We plan to add the possibility to generate a json and/or xml file of the stored data. This would facilitate and enable its use in any NLP project. We have chosen two formats in order to make it accessible to a wider public.

[12] Mentioned in Section 2.2 and Section 3.2.
[13] Gorgülü (2012, Ch. 1) gives an insight of the different theories on the subject matter.
[14] https://sozluk.gov.tr/
[15] As aforementioned, we sometimes encountered discrepancies in morpheme representation. We chose the morpheme that best represented the actual allomorphs found in derivatives.

Semantürk, the second resource is an ontology of semantic categories encoding meanings. Therefore the semantic category refers to the meaning of the morpheme and is representative of it. We have built this resource, written in Web Ontology Language (OWL), using a hybrid methodology applying both a top-down and a bottom-up method. 4/ Firstly, the main structure of the ontology is adapted from an existing tagset for the description of nominal semantics in French (Huguin et al., 2022). This tagset is based on WordNet's[16] top concepts called Unique Beginners (Fellbaum, 1998). Initially not defined for morpheme description, it proved adaptable as we applied the set to define the N-to-N morphemes at the semantic level as explained later. 5/ We then collected all the definitions and meanings found in the various sources and stored them in a single file, aligning them by morphemes and source. 6/ Once we had collected all the morpheme definitions, we matched them to the main structure of our ontology. 7/ If no match was found, or if the existing category was too broad to reflect the meaning of the morpheme, we created a new semantic category. As the semantic categories are hierarchically ordered, we could adapt the set and add new semantic categories specific to Turkish derivational morphemes, with the possibility of having different levels of granularity.

8/ In addition, we added a new *Semantic category* entry to *DerivBaseTR* and annotated each morpheme with the semantic categories of *Semantürk*, so that the morphemes are now described at the formal, categorial and semantic levels. Some morphemes present semantic transparency and are annotated with only one semantic category. Others are more ambiguous and have multiple semantic categories.

5 Conclusion

Prior to the construction of the morphosemantic analyser, the establishment of a formalised descriptive resource of derivational morphemes is necessary. The development of formalised resources requires the establishment of a specific framework for the description of Turkish morphemes. Therefore, we have created two different sets of resources: an ontology of semantic categories for the description of morphemes called *Semantürk* and *DerivBaseTR*, a database that formalises the description of morphemes at the formal, categorial and semantic levels. The resources are built with the perspective of possibly being used as additional components in various linguistic or NLP projects, and extended with other types of morphemes or new features. As the majority of published computerised resources are either not available or not easily accessible, this project is conducted in an Open Science perspective. We hope to provide extendible and interoperable resources to help improve the progress of the research in processing of the Turkish derivational morphology.

References

Oya Adalı. 2004. *Türkiye Türkçesinde Biçimbirimler*. Papatya, Istanbul, Türkiye.

Ahmet Akçataş and Serpil Taşdemir. 2020. Türkiye türkçesinde kök ya da gövdeye gelen ekler üzerine bir anlambilim incelemesi. *Avrasya Dil Eğitimi ve Araştırmaları Dergisi* 4(1):129–149.

Ahmet Afşın Akın and Mehmet Dündar Akın. 2007. Zemberek, an open source NLP framework for Turkic languages. *Structure* 10:1–5.

Esat Mahmut Bayol. 2018. Türkçe Doğal Dil İşleme Macerası. https://turkceddi.blogspot.com/2018/08/turkce-dogal-dil-isleme-maceras-her.html.

Louis Bazin. 1994. *Introduction à l'étude pratique de la langue turque*. Librairie d'Amérique et d'Orient, Paris, 3rd edition.

Erdoğan Boz. 2015. *Türkiye Türkçesi, Biçimbilimsel ve Anlamsal İşlevli Biçimbilgisi*. Gazi Kitabevi Tic. Ltd. Şti., Ankara, Türkiye, 4th edition.

Michel Bozdémir. 1991. *Méthode de turc*, volume 1. L'Asiathèque, maison des langues du monde, Paris.

Catherine Erikan, Ayhan Erdal, and Marie Koçoğlu. 2008. *Apprenons le turc ensemble / Beraber Türkçe Öğrenelim*, volume 1. Ataturquie, Paris, 2nd edition.

[16]Wordnet is a lexical database, used for more than 200 languages, including Turkish.

Gülşen Eryiğit. 2014. ITU Turkish NLP Web Service. In *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, pages 1–4. https://doi.org/10.3115/v1/E14-2001.

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. Language, Speech, and Communication. The MIT Press, Cambridge, London.

Emrah Gorgülü. 2012. *Semantics of nouns and the specification of number in Turkish*. Simon Fraser University.

Aslı Göksel and Celia Kerslake. 2005. *Turkish: a comprehensive grammar*. Routledge comprehensive grammars. Routledge, London.

Nabil Hathout and Fiammetta Namer. 2022. ParaDis: a family and paradigm model. *Morphology* 32(2):153–195. https://doi.org/10.1007/s11525-021-09390-w.

Mathilde Huguin, Lucie Barque, Pauline Haas, Fiammetta Namer, and Delphine Tribout. 2022. *Guide d'annotation Demonext: typage lexical des noms du français*.

Zeynep Korkmaz. 2014. *Türkiye Türkçesi Grameri, Şekil Bilgisi*. Türk Dil Kurumu Yayınları, Ankara, Türkiye, 4th edition.

Hugo Mailhot, Maximiliano A. Wilson, Joël Macoir, S. Hélène Deacon, and Claudia Sánchez-Gutiérrez. 2020. MorphoLex-FR: A derivational morphological database for 38,840 French words. *Behavior Research Methods* 52(3):1008–1025. https://doi.org/10.3758/s13428-019-01297-z.

Alice Missud, Pascal Amsili, and Florence Villoing. 2020. VerNom : une base de paires morphologiques acquise sur très gros corpus (VerNom : a French derivational database acquired on a massive corpus). In Christophe Benzitoun, Chloé Braud, Laurine Huber, David Langloi, Slim Ouni, Sylvain Pogodalla, and Stéphane Schneider, editors, *Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 2 : Traitement Automatique des Langues Naturelles, Nancy, France, June 8-19, 2020*. ATALA et AFCP, pages 305–313. https://aclanthology.org/2020.jeptalnrecital-taln.30/.

Fiammetta Namer. 2002. Acquisition automatique de sens à partir d'opérations morphologiques en français : études de cas. In *Actes de la 9ème conférence sur le Traitement Automatique des Langues Naturelles. Articles longs, TALN 2002, Nancy, France, June 2002*. ATALA, pages 237–246. https://aclanthology.org/2002.jeptalnrecital-long.21/.

Kemal Oflazer. 1994. Two-level description of turkish morphology. *Literary and Linguistic Computing* 9(2):137–148. https://doi.org/10.1093/llc/9.2.137.

Kemal Oflazer and Murat Saraçlar. 2018. *Turkish Natural Language Processing*, volume 1 of *Theory and Applications of Natural Language Processing*. Springer, Cham, Switzerland.

Seda Ozturk. 2016. *Création et reconnaissance de néologismes par méthode de suffixation*. Université de Franche-Comté, Besançon, France.

Haşim Sak, Tunga Güngör, and Murat Saraçlar. 2008. Turkish language resource: Morphological parser, morphological disambiguator and web corpus. In Bengt Nordström and Aarne Ranta, editors, *Advances in Natural Language Processing*. Springer, Lecture Notes in Computer Science. https://doi.org/10.1007/978-3-540-85287-2_40.

Rossella Varvara, Justine Salvadori, and Richard Huyghe. 2022. Annotating complex words to investigate the semantics of derivational processes. In *Proceedings of the 18th Joint ACL - ISO Workshop on Interoperable Semantic Annotation within LREC2022*. European Language Resources Association, pages 133–141. https://aclanthology.org/2022.isa-1.18.

Olcay Taner Yıldız, Begüm Avar, and Gökhan Ercan. 2019. An open, extendible, and fast turkish morphological analyzer. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*. INCOMA Ltd., pages 1364–1372. https://doi.org/10.26615/978-954-452-056-4_156.

Çağrı Çöltekin. 2010. A freely available morphological analyzer for turkish. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Ressources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2010/pdf/$109_{Paper.pdf}$.

Çağrı Çöltekin, A. Seza Doğruöz, and Özlem Çetinoğlu. 2023. Resources for Turkish natural language processing: A critical survey. *Language Resources and Evaluation* 57(1):449–488. https://doi.org/10.1007/s10579-022-09605-4.