

# Establishing Morpho-semantic Relations in FarsNet (a focus on derived nouns)

**Nasim Fakoornia**  
Shahid Beheshti University  
Tehran, Iran  
nasim\_fakournia@yahoo.com

**Negar Davari Ardakani**  
Shahid Beheshti University  
Tehran, Iran  
n\_davari@sbu.ac.ir  
na34@soas.ac.uk

## Abstract

This paper aims at a morpho-semantic analysis of 2461 Persian derived nouns, documented in FarsNet addressing computational codification via formulating specific morpho-semantic relations between classes of derived nouns and their bases. Considering the ultimate aim of the study, FarsNet derived nouns included 12 most productive suffixes have been analysed and as a consequence 45 morpho-semantic patterns were distinguished leading to creation of 17 morpho-semantic relations. The approach includes a close examination of beginners, grammatical category and part of speech shifts of bases undergoing the derivation process. In this research the morpho-semantic relations are considered at the word level and not at the synset level which will represent a cross-lingual validity, even if the morphological aspect of the relation is not the same in the studied languages. The resulting morpho-semantic formulations notably increase linguistic and operative competence and performance of FarsNet while is considered an achievement in Persian descriptive morphology and its codification.

## 1 Introduction

A comprehensive and detailed description of the relevant linguistic levels is a prerequisite for achieving progress in natural language processing (NLP). Wordnets are very popular lexical ontologies, relying on morphological, semantic and morpho-semantic descriptions and formulations. FarsNet which is a Persian wordnet has been established in 2009 by NLP research lab of Shahid Beheshti University. It goes closely in lines and principles of Princeton WordNet, EuroWordNet and BalkaNet (shamsfard et al. 2010). The latest version of FarsNet (2.0) contains 22180 nouns (including

2756 derived nouns), 5691 verbs, 6560 adjectives and 2014 adverbs. Besides semantic relations (synonymy, hypernymy, hyponymy, meronymy and antonymy) and morphological relations (derivation), some additional conceptual relations such as domain and related to, have been devised in FarsNet. At present (2015), it consists of more than 36000 entries, organized in almost 2000 synsets. The present study which is aimed at formulating morpho-semantic relations of FarsNet's derived nouns provides the wordnet with the basic required information for automation of the relations.

According to Deléger et al. (2009), a morpho-semantic process decomposes derived, compound and complex words into their base and associates such process to their semantic interpretation. Through morpho-semantic analysis derived and compound words are analysed morphologically and relations between base and derivational form are interpreted semantically (Namer & Baud 2007). Raffaelli & Kerovec (2008) consider "morphosemantics" as the best expression describing studies which deal with links between form and meaning at the word level.

Derivation and compounding are the two main word formation processes. Persian derivational morphology consists of an affixal system in which the number of suffixes is more than prefixes. Persian derivational morphological processes include suffixation, prefixation, only a single case of circumfixation and no infixation (Davari and Arvin 2015). Affixation patterns in this language are generally regular however in some cases there are few exceptions (Megerdooonian 2000). According to Keshani (1992) Persian derivational processes are relying on almost 56 suffixes. The aim of the present study is to neatly explore, formulate and classify

the morpho-semantic patterns of derived nouns by analysing the relevant data in FarsNet. It is worth noting that the present article originates from a wider scope research by Fakoornia (2013), in which all FarsNet derived nouns (2756) were analysed in order to establish morpho-semantic relations between derived nouns and their bases. The derived nouns under study included 26 different suffixes. In this study the derivatives of 12 most productive noun marker suffixes (2461) have been focused. This study enriches FarsNet while improves morpho-semantic codification of Persian.

After a brief introduction to FarsNet word entries in general and noun entries in particular, the process of morpho-semantic pattern formulation will be elaborated for the selected suffixes.

## 2 FarsNet Word Entries

Entries include phonological transcription, part of speech, synonyms and their classifications in to a synset, word meaning and an example. A beginner will be selected for each lexeme. According to Miller et al. (1990) a beginner is a primitive semantic component of any word in its hierarchically structured semantic field. Beginners could be used in the recognition of domains synsets. Different syntactic types can be related to each other in FarsNet; mapping each entry to its corresponding concept in Princeton WordNet 3.0 is also possible (Shamsfard et al., 2010). Using this information is essential in establishing morpho-semantic relations. Table 1 shows the prevailing noun beginners in FarsNet.

Noun beginners				
1. act	6. cognition	11. location	16. plant	21. shape
2. animal	7. communication	12. motive	17. possession	22. state
3. artifact	8. event	13. object	18. processes	23. substance
4. attribute	9. feeling	14. person	19. quantity	24. time
5. body	10. group	15. phenomenon	20. relation	25. food

Table 1: list of noun beginners in FarsNet

The synsets which do not fall into any of the above categories will be tagged by the label *nothing*. The semantic relations are also established among the synsets with the same POS. Synsets with different POS will be tagged by labels such as “related to”. There are 3 choices for mapping a synset to the correspondent one in Princeton WordNet 3.0: equivalence mapping, near-equivalence mapping and no-mapping. Finally, the morphological relations among senses, such as derivational relations are marked.

Besides specifying a noun type (such as common, proper, countable, uncountable, pronoun, number or infinitive), a classification on the basis of some more general semantic features (such as belonging to human, animal, location or time) is provided.

## 3 Data Analysis

For the purpose of this study the noun corpus of FarsNet (22180) were thoroughly explored. First of all, the list of derived nouns (2756) was prepared. Then they were broken into their roots and affixes. From among 26 suffixes, in this paper, the 12 most frequent were selected (2461 derivatives), described and analysed. They are listed in table 2, the morphological descriptions are compatible with Keshani's (1992) description of Persian suffixes.

	Suffix	POS	Semantic load
1	“-i”	n-n a-n n-d d-d	Any type of noun, adjective & adverb
2	“-e”	n-n v-n a-a	Any type of noun & adjective
3	“-æk”	n-n a-n v-n	Any type of noun
4	“-fje”	n-n	Diminution similarity
5	“-gah”	n-n v-n d-d	Location Body part time
6	“-dan”	n-n	Location Body part dish

7	“-gær”	v-n n-n n-a	Profession object
8	“-ban”	n-n	Similarity
9	“- ænde”	v-a v-n d-a a-a	Any type of noun & adjective
10	“-ar”	v-n v-a n-n n-a	Any type of noun & adjective
11	“-ej”	v-n a-n	Any type of noun
12	“-ane”	n/ a- a/ d n/ d- n	Any type of adjective & adverb food

Table 2: A list of selected suffixes

The following information is required to link each noun to its base:

- Morphological information of nouns; including POS of the base and derivative as well as other noun types such as; proper, common, number, etc.
- Semantic category; including human, animal, location, time or nothing.
- Beginner; such as act, person, feeling, event, etc., (table 1).
- The derivational relation between derived noun and its base.

#### 4 Morpho-semantic Analysis of Selected Suffixes

In this part we will scrutinize our 12 most productive selected noun marker suffixes from a morpho-semantic point of view. More information about the other Persian suffixes could be found in Fakoornia (2013).

##### 4.1 “-i”

In FarsNet, 4125 nouns ends in letter /i/ among which 1880 nouns are considered to be derivatives of suffix “-i”.

“-i” is an extremely productive Persian suffix. It has the potential for connecting to bases with different grammatical category, to compound words and even to syntactic phrases.

a. “-i” connects to nouns and adjectives and makes abstract noun, expressing an attribute or a state. The process is highly productive in Persian. Thus if “-i” connects to a noun or an adjective with different types of beginners, the resulting derivative beginner will be *attribute* or *state*. Considering the mentioned regularity the relation could be expressed as follows: “derivative attribute of base”, for example “bideGati attribute of bideGat”, (carelessness attribute of careless). FarsNet includes 802 tokens of such nouns.

b. “-i” connects to agent nouns and present participles, describing a job or an act and makes noun infinitive referring to a field, a job or an act. In Persian the beginner of agent noun is *person* and the beginner of gerund is *act* or *cognition*. So if “-i” connects to a noun belonging to *person* or to present participle, the beginner of derivative will be *act* or *cognition*. Following this the relation “base agent of derivative” is predictable. For example; “mohændes agent of mohændesi” (engineer agent of engineering). FarsNet includes 890 tokens of such nouns.

c. “-i” connects to agent noun and makes nouns referring to location or territory. So if “-i” connects to a base which is *person*, and makes a derivative referring to *location*, we will have the relation “derivative location of base” for example “tælaforuġi location of tælaforuġ” (jewelry location of jeweler). FarsNet includes 15 tokens of such nouns.

d. Other structures include the use of “-i” to refer to colors. Colors inherited from *property*. Thus if the base beginner is anything and the derivative beginner is *property*, the relation “base the same color as derivative” will be established. For example: “porteGal the same color as porteGali” (orange the same color as orange). FarsNet includes 15 tokens of such nouns.

e. “-i” connects to some other nouns, verbs and adjectives (excluding the above mentioned ones) and makes derivatives, referring to *feeling*, *process*, *event*, *act*, *person*, *object*, *nothing* etc. So if the base POS is verb, noun, adjective (other than present participles) and the derivative beginner could be anything, we will have the relation “derivative related to base”. For

example; “barani related to baran” (raincoat related to rain). FarsNet includes 144 tokens of such nouns.

“bædæxlagi attribute of bædæxlag” (irritability attribute of irritable) and also “bædæxlag agent of bædæxlagi”.

A summary of what has been explicated is listed in the table 3:

f. There are also 14 derivatives of “-i” in FarsNet which can be classified in both (a) and (b). In this case relations of “derivative attribute of base” and “base agent of derivative” can be established. For example:

input				output			
	Base POS	suffix	Base beginner	derivative POS	derivative beginner	morpho-semantic relation	number
a	n/adj	“-i”	anything	n	attribute/state	derivative attribute of base	802
b	n/pres. part.	“-i”	person	n	act/ cognition	base agent of derivative	890
c	n	“-i”	person	n	location	derivative location of base	15
d	n	“-i”	anything	n	Property	derivative the same color as base	15
e	v/n/adj	“-i”	anything except above	n	anything	derivative related to base	144
f	n/ pres. part.	“-i”	Person	n	attribute/state/act/cognition	“derivative attribute of base” and “base agent of derivative”	14
Total				1880			

Table 3: morpho-semantic patterns of suffix “-i” derivatives

According to the above patterns, “- i”’s word formation processes are formulated. The beginners are given in parenthesis and the frequency of each pattern is given in bracket.

a. Noun (person)/ present participle + “-i” = noun (act/ cognition) → base agent of derivative <890>.

b. Noun (anything)/ adjective + “-i” = noun (attribute/ state) → derivative attribute of base <802>.

c. Verb/ noun (other) / adjective + “-i” = noun (anything) → derivative related to base <144>.

d. Noun (person) + “-i” = noun (location) → derivative location of base <15>.

e. Noun (anything) + “-i” = noun (property) → derivative the same color as base <15>.

f. Noun (person) + “-i” = noun (attribute/ state/ act/ cognition) → derivative attribute of base/ base agent of derivative <14>.

As can be seen, “-i” is frequently involved in forming derivatives with beginners such as act, cognition and attribute. Few numbers of its derivatives are categorized under location and property. Formula (3) shows those patterns not covered in other structures.

#### 4.2 “-e”

7 morpho-semantic patterns have been distinguished for suffix “-e”:

a. Verb + “-e” = noun (anything except act) → derivative related to base verb form: “sorude related to sorudan” (song related to sing) <47>.

- b. Noun (anything) + “-e” → noun (other) → derivative related to base: “ruze related to ruz” (fast<sup>1</sup> related to day) <42>.
- c. Adjective + “-e” = noun (anything) → base attribute of derivative: “jævan attribute of jævane” (young attribute of sprout) <23>.
- d. Noun (object/ body) + “-e” = noun (anything) → derivative similar to base: “dæhane similar to dæhan” (opening similar to mouth) <16>.
- e. Noun (quantity) + “-e” = noun (time) → base quantity of derivative: “dæh quantity of dæhe” (ten quantity of decade) <4>.
- f. Verb + “-e” = noun (act) → derivative act of base verb form: “xænde act of xændidæn” (laughter (n.) act of laugh (v.)) <3>.
- g. Diminutive noun (person) + “-e” = noun (person) → derivative pejorative sense of base: “doxtæræke pejorative sense of doxtæræk” (bad girl pejorative sense of little girl) <1>.
- d. Verb + “-æk” = noun (anything) → derivative related to base verb form: “gæltæk related to gæltidæn”, (roller related to roll) <4>.
- e. Noun (anything) + “-æk” = noun (food) → derivative similar to base “pæfmæk similar to pæfm”, (cotton candy similar to wool) <3>\*
- f. Noun (person/ animal) + “-æk” = noun (person/ animal) → derivative diminutive of base: “doxtæræk diminutive of doxtær”, (little girl diminutive of girl) <2>.
- g. Noun (body) + “-æk” = noun (act) → base agent of derivative: “naxon agent of naxonæk”, (nail agent of pick) <1>.
- h. Noun (body) + “-æk” = noun (body) → derivative related to base: “guʃæk\*\* related to guʃ”, (eardrum related to ear) <1>.

\* Formula (a) and (e), however similar cannot be merged into a single category as in (a) although the beginner of both derivative and base can be anything, the tokens of each category are exclusive. It should be mentioned that in pattern (e) the beginner of derivative can be the same as the base.

\*\* As the POS and the beginner of the word “guʃæk” (eardrum), do not change in the derivation process, during computational codification it is classified in second formula, however, according to its meaning it cannot be entered in that group, thus it should be manually excluded and entered in a general relation (derivative related to base) formulated for it.

As can be seen, “-e” often links to verbs and creates derivatives with different types of beginners; it seldom results in pejorative nouns.

### 4.3 “-æk”

Suffix “-æk”\* shows 8 morpho-semantic patterns in Persian:

- a. Noun (anything) + “-æk” = noun (anything except food) → derivative similar to base: “surætæk similar to suræt”, (mask similar to face) <22>.
- b. Noun (anything except person/ animal/ food) + “-æk” = noun (anything except person, animal and food) → derivative similar to base and derivative diminutive of base: “ʃæhræk similar to ʃæhr” and “ʃæhræk diminutive of ʃæhr”, (town similar to city) and (town diminutive of city) <11>.
- c. Adjective + “-æk” = noun (anything) → base attribute of derivative: “sorx attribute of sorxæk”, (red attribute of measles) <6>.

### 4.4 “-ʃe”

Suffix “-ʃe” shows 2 morpho-semantic patterns:

- a. Noun (anything) + “-ʃe” = noun (anything) → derivative diminutive of base and derivative similar to base: “dæryaʃe diminutive of dærya” and “dæryaʃe similar to dærya”, (lake diminutive of sea) and (lake similar to sea) <28>.

“-ʃe” in some nouns does not refer to similarity or diminution but it merely indicates a vague relatedness, an example is “ʔænbærtʃe”, (sachet). In such situations the relation “derivative related to base” is formulated, but during computational codification derivatives belonging to this

<sup>1</sup> abstain from certain foods, as for religious or medical reasons (especially during the day)

structure, automatically classified in the previous structure which should be manually removed from it. In FarsNet there was only one derivative of this type. Thus the formula would be:

- b. Noun (anything) + “-fje” = noun (anything) → derivative related to base: “ʔænbærtfje related to ʔænbær”, (sachet related to ambergris) <1>.

#### 4.5 “-gah”

Suffix “-gah” shows 3 morpho-semantic patterns:

- a. Noun (anything)/ verb + “-gah” = noun (location) → derivative location of base: “dærmangah location of dærman”, (health centre location of treatment) <83>.
- b. Noun (anything) + “-gah” = noun (body) → derivative related to base: “gijgah related to gij”, (temple related to dizzy) <6>.
- c. Verb + “-gah” = noun (anything) → derivative related to base verb form: “didgah related to didæn”, (viewpoint related to view) <1>.

The above shows that the number of derivatives, having location as their beginner is more than the other beginners. Moreover the suffix rarely connects to a verb.

#### 4.6 “-dan”

Suffix “-dan” shows a single morpho-semantic pattern in Persian:

- a. Noun (anything) + “-dan” = noun (anything) → derivative location of base: “goldan location of gol”, (vase location of flower) <11>.

#### 4.7 “-gær”

Suffix “-gær” shows 4 morpho-semantic patterns:

- a. Noun (act) + “-gær” = noun (person) → derivative agent of base: “arayeʃgær agent of arayeʃ”, (stylist agent of makeup) <28>.
- b. Noun (anything except act) + “-gær” = noun (person) → derivative related to

base: “ahængær related to ahæn”, (blacksmith related to iron) <13>.

- c. Noun (anything) + “-gær” = noun (object) → derivative instrument of the base: “næmayeʃgær instrument of næmayeʃ”, (monitor instrument of display) <3>.
- d. Verb + “-gær” = noun (object) → derivative agent of base verb form: “roftægær agent of roftæn”, (dustman agent of sweep) <1>.

#### 4.8 “-ban”

Suffix “-ban” shows 3 morpho-semantic patterns:

- a. Noun (anything) + “-ban” = noun (person) → derivative protector of base: “jængælban protector of jængæl”, (woodsman protector of wood) <17>.
- b. Noun (anything) + “-ban” = noun (object) → derivative related to base: “sayeban related to saye”, (sunshade related to shade) <3>.
- c. Verb + “-ban” = noun (person) → derivative agent of base verb form: “dideban agent of didæn”, (sentinel agent of guard) <2>.

#### 4.9 “-ænde”

Suffix “-ænde” shows a single morpho-semantic pattern:

- a. Verb + “-ænde” = noun (anything) → derivative agent of base verb form: “ʔafarinænde agent of ʔafæridæn”, (creator agent of create) <76>.

#### 4.10 “-ar”

Suffix “-ar” shows 4 morpho-semantic patterns:

- a. Noun (anything) + “-ar” = noun (anything) → derivative related to base: “dadar related to dad”, (God related to justice) <5>.
- b. Verb + “-ar” = noun (act) → derivative act of base verb form: “goftar act of goftæn”, (speech act of say) <2>.
- c. Verb + “-ar” = noun (person) → derivative agent of base: “xæridar agent of xæridæn”, (buyer agent of buy) <2>.
- d. Verb + “-ar” = noun (anything except act and person) → derivative related to base

verb form: “saxtar related to saxtæn”,  
(structure related to construct) <2>.

#### 4.11 “-eʃ”

Suffix “-eʃ” shows 3 morpho-semantic patterns:

- a. Verb + “-eʃ” = noun (act) → base act of derivative verb form: “Gorridæn act of Gorrefʃ”, (roar (v.) act of roar (n.)) <68>.
- b. Verb + “-eʃ” = noun (anything except act) → derivative related to base verb form: “deræxʃeʃ act of deræxʃidæn”, (shine act of shine) <15>.
- c. Noun (anything) + “-eʃ” = noun (anything) → derivative related to base: “yoneʃ related to yon”, (ionization related to ion) <8>.

#### 4.12 “-ane”

Suffix “-ane” shows 3 morpho-semantic patterns:

- a. Noun (anything)/ adverb + “-ane” = noun (food) → derivative food of the base: “sobhane food of sobh”, (breakfast food of morning) <7>.
- b. Verb + “-ane” = noun (object) → derivative instrument of base verb form: “resane instrument of resandæn”, (media instrument of broadcast) <6>.
- c. Noun (anything) + “-ane” = noun (anything except food) → derivative related to base: “ʔængoʃtane related to ʔængoʃt”, (thimble related to finger) <5>.

The 2 represented exceptions; “guʃæk” (eardrum) and “ʔænbærtʃe” (sachet) will naturally and respectively fall in the formulated relations “derivative similar to base” or “derivative diminutive of base” and “derivative diminutive of base” or “derivative similar to base”, however considering the meaning of their bases and the resulting derivatives, they do not belong to the mentioned relations, thus some other relations should be formulated to include them.

## 5 Conclusion

Morpho-semantic analysis of a selection of 2461 derived nouns in FarsNet showed 45 morpho-semantic patterns and 17 morpho-semantic relations (such as “derivative agent of base”, “derivative location of base”, etc.) for 12 most

productive suffixes. Considering that only 2 words out of 2461 (0.08%) did not fall into the patterns, it could be concluded that the patterns have successfully provided the foundations for establishing automatic relations between derived or complex nouns and their bases in FarsNet. The coincident consideration of the words’ morphological features such as their POS, their semantic and grammatical category (e.g. agent noun, participle noun, present participle, etc.) as well as recognizing the beginners of the bases (e.g. act, person, food, etc.) and their change after the affixation process have been the key criteria in formulating the relations which were especially crucial for the majority of studied suffixes that were polysemous. Defining and codifying these morpho-semantic patterns leads us to coherent establishment of morpho-semantic relations in FarsNet and hence has a remarkable developing impact on the applicability of the data base in machine translation, question answering systems, etc. Although In this research the morpho-semantic relations are considered at the word level and not at the synset level, mapping the results to the relations formulated in other languages wordnets will provide a cross-lingual validity, even if the morphological aspect of the relation is not the same in the mapped languages.

## References

- Davari Ardakani, Negar and Mahdiyeh Arvin. 2015. Persian. In N. Grandi and L. Kortvelyessy, editors. *Edinburgh Handbook of Evaluative Morphology*. Edinburgh University Press, Edinburgh, pages 287-295.
- Deléger, Louise, Fiammetta Namer and Pierre Zweigenbaum. 2009. Morphosemantic Parsing of Medical Compound Words: Transferring a French Analyzer to English. *International Journal of Medical Informatics*, 78 (1): 48-55.
- Fakoornia, Nasim. 2013. Morphosemantic Analysis of Nouns in Persian and English Aiming at Computational Codification. Master’s thesis, Shahid Beheshti University, June.
- Farshidvard, Khosrow. 2007. *Derivation and Compounding in Persian*. Zavar press, Tehran.
- Keshani, khosrow. 1992. *Suffix Derivation in Contemporary Persian*. Iran University Press, Tehran.
- Megerdoomian, Karine. 2000. *Persian Computational Morphology: A unification-based Approach*. NMSU,

CRL, Memoranda in Computer and Cognitive Science (MCCS-00-320).

Miller, George A. et al. 1990. Introduction to Wordnet: An Online Lexical Database. *Journal of Lexicography*, 3 (4): 235-244. doi:10.1093/ijl/3.4.235

Namer, Fiammetta and Robert Baud. 2007. Defining and Relating Biomedical Terms: Towards a Cross-language Morphosemantics-based System. *International Journal of Medical Informatics*, 76(2-3): 226-233.

Raffaelli, Ida and Barbara Kerovec. 2008. Morphosemantic fields in the Analysis of Croatian Vocabulary. *Jezikoslovlje*, 9 (1-2): 141-169.

Shamsfard, Mehrnoush et al. 2010. *Semi-Automatic Development of FarsNet; The Persian WordNet*. 5<sup>th</sup> Global WordNet Conference (GWA8020), Mumbai, India.