

An Account for Compound Prepositions in Farsi

Zahra Abolhassani Chime

Research Center of Samt, Tehran, 14636

Ph.D in Linguistics

zabolhassani@hotmail.com

Abstract

There are some sorts of 'Preposition + Noun' combinations in Farsi that apparently a Prepositional Phrase almost behaves as Compound Prepositions. As they are not completely behaving as compounds, it is doubtful that the process of word formation is a morphological one.

The analysis put forward by this paper proposes "incorporation" by which an N^0 is incorporated to a P^0 constructing a compound preposition. In this way tagging prepositions and parsing texts in Natural Language Processing is defined in a proper manner.

1 Introduction

Prepositions have very versatile functions in Farsi and at the same time very important roles in linguistics especially in computational linguistics. Most of the linguists consider them as members of a closed set in which nothing can be added and behavior of which is completely static. However this paper tries to touch some aspects of the fact that not only this set is not a closed one but also the behaviors of its members are so dynamic that we can call the set a productive one. Having considered this fact about very frequent Farsi prepositions, we can come up with a useful model for language recognition.

There is a large discrepancy among linguists in classifying Farsi prepositions that whether or not there are compound prepositions and if there are how the process of their word formation should be accounted for as their characteristics are not as straight forward as it is expected from other compound categories.

Some Iranian Linguists have ignored this class altogether (Khānlari (1351), Shafāii (1363), Bāteni (1356), Seyed vafāii (1353)). Some

believe they are not compound without putting forward any explanation but some sort of description. (Homā'yanfārox (1337), Sādeghi (1357), Kalbāsi (1371)). Some believe they are compounds without analyzing them (Mashkur (1346), Khatib Rahbar (1367), Gharib (1371), Meshkatodini (1366)) and still some have defined them as prepositional phrases in one way or another (Gholam Alizade (1371), Samiian (1983)). However we can not find a comprehensive account for this class of prepositions. This paper tries to tackle the problem from a different generative view as well as a familiar way in LA-morph (Hausser: 2001) in parsing through which we can account for the diversity of their behavior and present them in tree configuration.

For reasons of computational efficiency and linguistic concreteness (surface compositionality) the morphological component of the SLIM theory of language take great care to assign no more than one category (syntactic reading) per word form surface whenever possible (Hausser, 2001: 244). As Farsi does not enjoy the benefit of "space" in word recognition we have to resort to other clues to find out exact way of parsing and tagging. This paper helps to make sure about the category of one construction of prepositions.

2 Constructions of 'Preposition + Noun' in Farsi

From among all constructions in Farsi in which a preposition and a complement -generally NPs - occurs, there are 4 classes which seem to have different behaviors of that usual PPs (prepositional phrases) although they have exactly similar structure to that of PPs; These classes are as follows from which we just turn our attention to the first one:

1. preposition + noun

- e.g. /bar/ + /asās-e/
on + basis
/e/ an obligatory genitive ending,
2. noun + preposition
e.g. /banā/ + /bar/
based + on
3. preposition + time / location item
e.g. /az/ + /pase/
from + behind
4. time / location item + preposition
e.g. /pošt/ + /be/
back + to

From the form point of view, we can simply consider preposition such as /bar/ ‘on’, /az/ ‘from/of’, /dar/ ‘in’, /bā/ ‘with’, /be/ ‘to’ as (real) prepositions and what comes immediately after, as complement.

However, a close observation reveals that not in all constructions consisting of a preposition and a noun the immediate noun can be considered as the noun head of the NP complement. That is in some phrases the head preposition is the compound preposition (a preposition and a noun) and then the noun after this construction is the complement:

5. /bar/ + /asās-e/ + /motāle’āt/
p complement (n)
“on + bases” (of) researches

The first question we try to answer is: Does the immediate noun after the preposition in (5), behave like other nouns as complements in PPs?

To answer this question we should make sure whether the noun (complement) is as independent as the other nouns in ‘preposition + nouns’ making prepositional phrases, or it is somehow merged with the preposition producing compound preposition.

There are some structural tests to reveal this. If the noun here expands as other nouns in other prepositional phrases we can conclude that the related structure is a phrase, otherwise it is better to think about them as compound prepositions.

3 Extending the structure under discussion

3.1 Premodifiers

The noun in prepositional phrases, can be extended in different ways while as the examples below show, the related structures cannot:

3.1.1 Demonstratives

6. bar (*in) asās-e motāle’āte dānešmandān
on (this) bases-of researches-of scientists
havā-ye zamin garmtaršode ’ast
climate-of earth increased has

“Based of scientists’ researches the climate of earth has increased”.

- 6’) bar (in) bām-e xāne kasi rāh miraft.
on (this) roof-of house someone (was) walking

3.1.2 Superlatives

- 7) bar (*jadid-tarin) asās-e motāle’āt-e ...
on the newest basis-of researches-of

- 7’) bar (zibā-tarin) bām-e xāne ...
on the most beautiful roof-of house

3.1.3 Exclamatories

- 8) bar (*che!) asās-e motāle’āt-e ...
on what! a basis-of researches-of

- 8’) bar (che!) bām-e xāne ...
on (what!) a roof of house

3.1.4 Quantifiers

- 9) bar (*har) asās-e motāle’āt-e ...
on (every) basis-of researches-of

- 9’) bar (har) bām-e xāne ...
on (every) roof-of house

3.1.5 Question words

- 10) bar (*che) asās-e motāle’āt-e ...?
on what basis-of researches

- 10’) bar (che) bām-e xāne-i ...?
on what roof-of house

3.1.6 Indefinite /yek/ ‘one’

- 11) bar (*yek) asās-e motāle’āt-e ...
on one basis-of researches

- 11’) bar (yek) bām-e xāne ...
on (one) roof-of house

3.2 Post Modifiers

Nouns in prepositional phrases can expand with post modifiers while nouns in our structure cannot.

3.2.1 Plural Markers

12) az Jāneb (*haye) dowlat va mardom
from side (s)-of government and nation
masā'eli matrah šod.
affairs raised was

“Some affairs were raised by government and nation.”

12') as ketāb (ha-ye) Ali estefāde kardam.
from book (s)-of Ali used I did.

“I used Ali’s books.”

3.2.2 Adjectives

13) be elate (*puš-e) bārandegi madāres ta'til
šod.
to cause-of (vain-of) raining schools closed
were.

“schools were closed because of the vain reason of raining.”

13') bar bām-e (ziba-ye) xāne qadam bogzar.
on roof-of (beautiful-of) house step put.

“step on the beautiful roof of the house.”

3.2.3 Appositives

14) bar asās-e (*pāye-ye) motāle'āt-e
dānešmandān
on basis-of (base-of) researches-of
scientists

14') Ali az xāne (mahale zendegi)-ash dur šode
ast.
Ali from house (place-of living)-his far made
is.

“Ali has left his house-his place of living.”

3.3 Conclusion

The conclusion we extract out of these observations imposes some hypotheses:

1) The noun in these kinds of structures has lost its independent status and the whole structure has turned into a morphological compound preposition.

2) The intended construction, is a special kind of “compound” probably a syntactic compound, in which not all characteristics of morphological compounds can be observed.

To evaluate the first hypothesis, we should first identify the criteria of compound words in these apparent phrases.

4 Compound Words in Farsi

Farshid vard (1351) believes it’s very difficult to identify and define the compound words in Farsi, because to gain the criteria of compound words, we should recognize compound forms from some other related and close structures, such as derived words and phrases.

In a phrase, grammatical roles of the parts are devoted as one to the head and the whole group rather than the parts contributes to the role of the phrase. Different ways of argumentation that can be established for distinction between phrases and compound words can be classified into 4 classes: phonological, morphological, syntactic and semantic

4.1 Phonological Argumentation

It is assumed that prepositions in Farsi do not bear any accent. This assumption comes from the fact that accent pattern in Farsi is in a any that the last or the farthest member of the group (phrase) takes the accent, except in marked structures; and as prepositions do not occur at the end of the phrase (PPs are head-first, as the other phrases in Farsi), they never take the accent. Eslami (1379: 28) states this fact as the “Head-escape Principle”:

“In all cases, with expanding the head of a syntactic phrase, the accent of the phrase falls on the farthest member.”

15. [[az] [xāne]]
“from the house”

16. [[az] [xāne-ye] [rezā]]
“from the house-of Reza”

The above observations, i.e.: 1. Accent on the last modifier and 2. Accent on the last syllabus of the word we conclude that the pattern of accent of the compound prepositions and prepositional phrases are absolutely the same.

In fact phonological reasons and criteria do not help of any kind.

4.2 Morphological Argumentation

All what was mentioned in previous section as expanding possibility of PPs can also be considered as morphological criteria.

4.3 Syntactic Argumentation

4.3.1 Topicalization

In topicalization “one word” can be topicalized out of a phrase but not out of a compound word.

17. Tamiz kardan-e ketāb-xāne bā Ali-st.
cleaning-of book-case with Ali is.

“cleaning book-case is with Ali”

17'. *ketāb tamiz kardan-e xāne-ash bā Ali-st.
book cleaning-of case-its with Ali is.

“book, cleaning of its case is with Ali.”

In (17) (ketāb) is a part of a compound word from which no part can be topicalized.

Now let’s see what happens if we topicalize a word in our construction.

18. bā Ali dar mored-e dānešgāh sohbat kardam.
with Ali in case-of university talk I made.

“I talked with Ali about the university.”

18'. *mored-e dānešgāh, bā Ali daresh sohbat kardam.
case-of university, with Ali in-it talk I made.

“About university, I talk about it with Ali.”

4.3.2 Coordination

Two similar constituents can be coordinated but not parts of compound words:

Noun out of PPs:

19. Hasan bā [dust va došman] modārā mikonad.
Hassan with [friend and enemy] bears

“Hassan bears every one.”

Parts of prepositions:

19'. *be [dalil-e va ellat-e] sarmā madrese-ha ta‘til šod.
to [reason-of and cause-of] cold schools closed became.

“Because of cold schools were closed.”

4.4 Semantic Argumentation

Close semantic observation of these constructions reveal that the nouns in the above mentioned combinations are special kind of nouns with particular semantic features.

All the nouns are “noun-referential” and “abstract”.

/dar mored-e/, /dar zamine-ye/, /bar asās-e/
in case-of in field-of on basis-of
“about” “about” “on”

/bar hasb-e/, /az heis-e/, /az lahāz-e/
on according from aspect from aspect
“according” “according” “point of view”

/bar asar-e/
on cause-of
“because of”

Another point to be mentioned is a delicate semantic difference between the meaning of these nouns in other constructions and in combination with prepositions. For example “dalil” in following two sentences does not bear the same semantic features.

20. man dalil-e harf-haye šomā rā nemifahmam.
I reason-of talks your don’t understand.

“I do not understand the reason of your talks”.

20'. man be dalil-e harf-haye šomā jalase rā tark kardam.

I to cause-of talks your meeting left.

“I left the meeting because of your talks”.

“dalil” in (20) has the semantic components of “argumentation, base, reason”, but in (20’) “because, for”.

Still another point worth mentioning is that most of the class members are synonymous in one way or another:

- dar mored-e, dar zamine-ye, dar xosus-e, dar bāre-ye, dar bāb-e, dar atrāfe,
“about”
- bar asās, bar paye-ye, bar hasb-e
“on, on the basis”
- az nazar-e, az heis-e, az lahāz-e, az jahat-e
“according to”
- be mojarad-e, be mahze
“once”
- be mojob-e, be ellat-e, be dalil-e
“because of”

5 Concluding the Discussion

Through same constituency tests, we showed that these constituents do not obey the phrasal characteristics. On the other hand, criteria of distinguishing compound words from syntactic phrases demonstrate that these forms are not so merged into each other in a way that they can be called fixed morphological compounds. It seems that they are in a transition phase from PPs to compound Ps. So although they are compounds we should look for the process of word formation

to take place in some other places rather than the morphology, i.e. in syntax.

The argumentation proposed by the author is “incorporation”, which can account for the behavior of such constructions in Farsi.

6 Incorporation

Incorporation brings out two changes in sentence representation: 1. It produces a compound category of word level (X^0). 2. It establishes a syntactic relationship between two places: the original position of the moved category (situ) and the target position. The former is a morphological and the latter is a syntactic change.

Baker (1988) considers X^0 movements similar to those of XP, with all constraints and conditions applicable to both. He also proposes “Government Transparency Corollary” to account for the grammatical changes. Movement automatically changes the governance features of a structure and the reason is that it creates a grammatical dependency between two distinct phrases.

Leiber (1992: 14) says that there are some facts that show to some extent there should be same interaction between syntax and morphology. Thus X parameters and related systems are not merely applicable to syntax, but morphology too.

However incorporation of this kind in Farsi is abstract, i.e. there is no overt movement.

During incorporation process head X^0 (here N^0) moves from its place towards P node and attaches to the P (dar) as it is shown in figure 1 and 2.

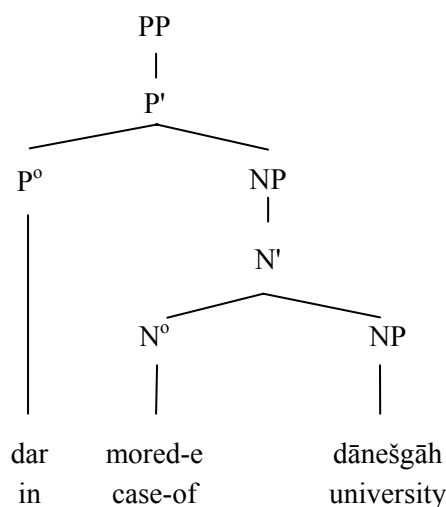


Figure 1

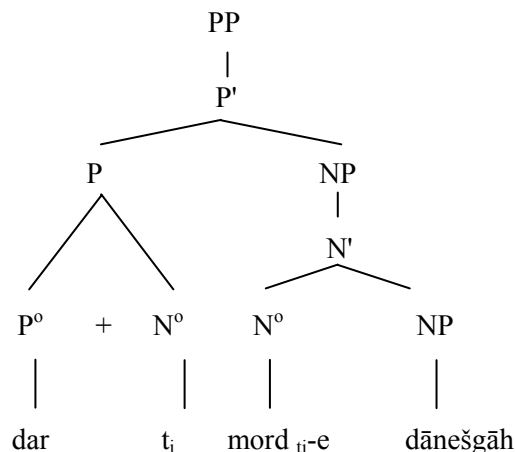


Figure 2

“dar+mored-e” dominated by a P node has the features of preposition and in this way θ -role change of “mored” is realized as preposition in combination with an original preposition. This syntactic process gives the following results:

1. A noun head (N^0) dominated by NP as a complement of a pp, α -moves and incorporates to the preposition head (P^0).
2. Moved N^0 is governed and dominated by a preposition node.
3. The output of the combination of the N^0 and P^0 is a compound P^0 .
4. The preposition (dar) “in” which before incorporation assigned θ -r to NP, after incorporation together with the noun (mored-e) assigns the θ -r to the NP (dānešgāh).
5. The resulted compound is a “syntactic compound”.

The needed conditions for incorporation of N^0 to P^0 can be summarized as follows:

1. P^0 should be morphologically simple and among the members of this group: dar “in”, be “to”, bā “with”, az “of, from”, bar “on”. They do not take genitive ending /-e/ (kasre-ezāfe) and having the [-V, -N] features are considered as “true” prepositions (Samiian, 1992)
2. N^0 should be morphologically simple and having all the features of [non-referential, abstract, complement-taking, indefinite].

Hereby it becomes clear why not every combination of “preposition + noun” lead to “compound prepositions” through incorporation, even if their occurrence bears a high frequency. The algorithm-like of this process is shown in figure 3.

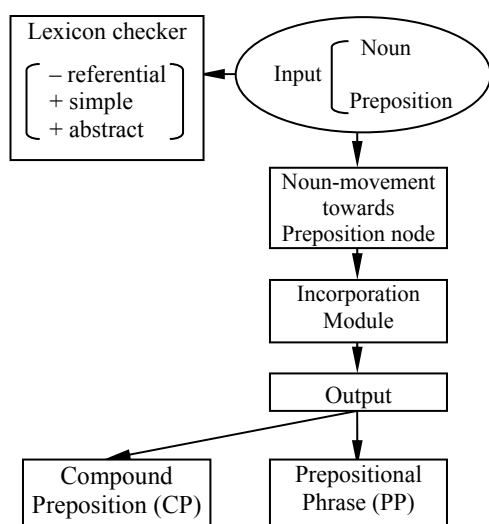


Figure 3

Prepositions are functional and so syntactic categories rather than lexical ones. I believe word formation of this category is motivated by syntax, in different ways one of which was argued here. This account contributes to the discipline of computational linguistics in labeling prepositions in Farsi, as this area of preposition labeling has been very challenging.

Although Voutilainen (2003) believes that data-driven taggers seem to be better suited for the analysis of fixed-word-order poor-morphology languages like English, but the finding of this paper is applicable to Farsi parts of speech recognition at least in the area of compound prepositions.

Prepositions are one sort of parts of speech, the recognition of which can be helpful in stemming for information retrieval (IR), since knowing a word's POS can help tell us which morphological affixes it can take. It can also help an IR application by helping select out nouns or other important words from a document. Automatic POS taggers can help in building automatic word-sense disambiguating algorithms, and POS taggers are also used in advanced ASR language models such as class-based n -grams (Jurafsky and Martin, 2000: 288)

Acknowledgement

My special thanks go to Masood Ghayoomi at the Institute for Humanities and Cultural Studies for his supports and encouragements in my research.

References

- Baker, M. C. (1988) *Incorporation, A Theory of grammatical function changing*. The University of Chicago Press, Chicago.
- Bateni, Mohammadreza (1356) *Tosife Sāxtemane Dasturie Zabāne Farsi*, Tehran, Amirkabir Publication.
- Eslami, Moharam (1379) *Šenaxte Navāye Goftāre Zabāne Farsi va Karborde ān dar Bāzsazi va Bāzšenāsie Rayaneie Goftar*, Ph.D diss., Tehran University, Linguistic department.
- Farshidvard, Khosrow (1351) “Kalameye morakab va meyāre tašxise ān”, *Proceedings of 2nd Iranian Researches Seminar*, Vol. 1, Mašhad University.
- Gharib, Abdolazim et al (1371) *Dastare Panj Ostaād*, Ašrafī Publication, 10th ed.
- Gholām Ali Zade, Khosrow (1374) *Sāxte Zabāne Farsi*, Ehyāye Ketāb Publication.
- Hausser, Roland (2001) *Foundations of Computational Linguistics*, Springer.
- Homayoun Farokh, Abdorahim (1337) *Dasture Jāme Zabāne Fārsi*, Tehran, Elmi Publication.
- Jurafski, D. and J. H. Martin (2000) *Speech and Language Processing: An Introduction to Natural Language Processing, Computational linguistics and Speech Recognition*. Prentice Hall, Pearson Higher Education.
- Kalbasi, Iran (1371) *Sāxte Ešteqāqie Vāje dar Fārsie Emruz*. The Institute of Studies and Cultural Researches.
- Khanlari, Parviz (1351) *Dasture Zabāne Fārsi*, Tehran Bonyad Farhangy Iran.
- Khatibrahbar, Khalil (1367) *Dasture Zabāne Farsi: Ketabe Harfe ezāfe va Rabt*. Sadi Publication.
- Lieber, R. (1992) *Deconstructing Morphology*, The University of Chicago Press.
- Mashkur, M. Javad (1346) *Dasturnāme dar Sarf va Nahve Zabāne Fārsi*, Shargh Publication Institute.

Meshkatodini, Mehdi (1366) *Dasture Zabāne Fārsi bar Payeye Nazariye Gaštāri*, Ferdowsi University

Sadegi, Aliashraf (1349) “Horufe ezāfe dar Farsie moaser”, *Journal of literature and Humanities*, Tehran University, pp (441-470).

Samiiian, Vida (1983) *Structure of Phrasal Categories in Persian: An X-bar Analysis*. Ph.D diss. University of California, Los Angeles.

Samiiian, V. (1991) *Prepositions in Persian and the Neutralization Hypothesis*. California State University, Fresno.

Seyed Vafai (1353) “Horufe ezāfe dar zabāne Farsie moaser”, *Journal of Literature and Humanities*, Tehran University, pp (49-86).

Shafaii, Ahmad (1363) *Mabanie Elmie Dasture Zabāne Farsi*, Novin Publication.

Voutilainen, Aro (2003) in Mitkov, Ruslan(ed), *The Oxford Handbook of Computational Linguistics*, Oxford University Press.