# Using Corpus Pattern Analysis (CPA) To Recognize Meanings of Zadan

## Gholamhossein Karimi Doostan[1], Arezoo Razani[2]

## 1. INTRODUCTION

With the advent of huge corpora in recent decades, there has appeared a tendency among researchers to utilize them in lexicography. According to Miller (2018), a major issue involved in Persian lexicography is the presence of light verbs. The present study sought to address the issue using corpora and methods of computational lexicography.

A light verb is in fact a complex predicate composed of two or more predicate elements: a semantic light verb and a second component (including a noun, an adjective, or a prepositional phrase). In Persian, there are 150 to 200 full verbs, and other verb-related ideas are expressed by light verbs, which explains why they are of such great significance in the language. Light verbs have been investigated from different aspects including compoundability, separability, and syntactic structure. The present function-oriented research, however, was based exclusively on data extracted from a corpus.

## 2. MATERIALS AND METHODS

The methodology of this study was adopted from Hanks' (2013) technique of corpus pattern analysis. The method has been employed for Italian and Spanish, and the project entitled Pattern Dictionary of English Verbs is currently in progress using the same technique, which associates a word use to a syntactic-pragmatic pattern using data extracted from a corpus. The major characteristic of the corpus pattern analysis technique is its inductive examinations as well as dependency on corpus data, as Hanks (2013) argues that a word in isolation has only a potential meaning, which is activated differently in every context.

The present study investigated the uses of the Persian verb zædæn using the 500-million-word Bijankhan corpus. zædæn is one of the most frequently used Persian light verbs, which is utilized also as a full verb. In the first step, 2000 matches were examined to obtain a list of non-verbal elements that co-occurred with the light verb zædæn, which amounted to more than 20000 in the next step, examined using the corpus patterns analysis technique.

---

[1] Professor of Linguistics, University of Tehran, Iran, Tehran; Gh5karimi@ut.ac.ir
[2] PhD candidate in Linguistics, University of Tehran, Iran, Tehran (corresponding author); arezoo_razani@ut.ac.ir

## 3. RESULTS AND DISCUSSION

For application of the technique to Persian, co-occurrences of the verb zædæn with each of the listed non-verbal elements were searched for in the corpus, and the obtained matches were then semantically classified based on the question of whether all the utterances had the same meaning. Next, the semantic classes were examined based on the question of whether all the utterances had the same thematic structure, and the associated structural and semantic patterns were specified in each case. Hanks (2013) introduces the notion of semantic type for semantic classification. A group of words with a common meaning can co-occur with a verb as a theme, where each word is called a semantic class, and the commonality is referred to as a semantic type. Furthermore, a regular grammar is used for structural examination of matches, on which basis any coherent discourse can be regarded as a corpus consisting of paragraphs and containing items such as subjects, predicates, objects, complements, and adverbs.

## 4. CONCLUSION

The examination of the verb zædæn in the corpus obtained 600 different structural uses and 400 particular meanings. The light verb construction has been defined in previous research as one where a non-verbal element, including a noun, an adjective, or a prepositional phrase, constitutes a single structure in combination with a light verb. The results obtained from the analysis of the patterns extracted from the corpus indicated that the verb zædæn is used in Persian in four ways, where the verbal structure can have a compositional or an idiomatic meaning:

1. full verb zædæn
2. non-verbal element plus light verb zædæn
3. [semantic type] plus light verb zædæn
4. verb phrase containing light verb zædæn that constitutes a semantic frame, acquiring a full meaning through addition of a variable semantic component, usually an adjective.

The type-1 and type-2 uses concern cases that have been referred to in a range of previous studies and examined as full verbs. The type-3 and type-4 uses, however, pertain to new cases that call for investigation in regard to other Persian light verbs.

**Keywords:** Corpus; Corpus Pattern Analysis (CPA); Light Verbs; Meaning; Persian