The Lexicon-Grammar Continuum: What Persian Complex Predicates Reveal

Abstract

We argue for the lexicon-grammar continuum by looking into the separability of Persian complex predicates, an open class of verbs that are thought to be either separable or inseparable. We contend that separability is best described as a scalar phenomenon rather than an all-or-nothing phenomenon. Analysis of *zadan* (hit) within the framework of Construction Grammar yields a semantic description of *zadan* complex predicates as a radial category and the frequency of separated complex predicates. It is argued that the more *zadan* is removed from its concrete prototype meaning, the less the resulting complex predicates are open to separation. Compositional complex predicates are also more commonly used in separated configurations, but only a weak negative correlation was found between frequency and separability. However, complex predicates at the extremes of the frequency continuum seem to show a reverse relationship, supporting the idea that high frequency linguistic chunks may gain autonomy with a less flexible internal structure.

Keywords: complex predicates, separability, frequency, compositionality, Construction Grammar, Persian

1. Introduction

Contrary to the modular view of the mind (Fodor, 1983; Chomsky, 2017), cognitive linguistic approaches posit a lexicon-grammar continuum with gradient boundaries between the traditional modules of language (Langacker, 2008: 15; Bybee, 2006; Goldberg: 1995: 23). In the cognitive linguistic view, discrete modules of language (e.g., morphology and semantics) exist but it is not always possible to ascribe given linguistic units to a single language module. An example is Persian Complex predicates (CPs) which are variously argued to be generated in the lexicon (Goldberg, 2003) or in the syntax (Folli et al, 2005). CPs are "composed of more

than one grammatical element (either morphemes or words), each of which contributes a non-trivial part of the information of the complex predicate" (Alsina et al, 1997:1). Persian has an immensely productive system of complex predicates that draws on a very small collection of simple verbs (Karimi, 2008) to readily create new verbs. Recent examples are *lāyk kard-an/zad-an* [like do-INF/hit-INF][1] 'like a social media post', *lāyk khord-an* [like collide-INF] 'be liked on social media', and *lāyk gereft-an* [like take-INF] 'get social media likes', which despite already existing Persian equivalents (e.g., *pasandid-an* [like-INF] 'like'), appear to be more commonly used (in a Google search for the phrase 's/he liked the post', 38 unique attestations appeared in the search results for *lāyk kardan* as opposed to three cases for *pasandidan*). The nonverbal element of a CP (*lāyk* in the above examples) is called a preverb since it always comes before the simple verb (although the default order can be reversed since Persian is a scrambling (Adli, 2010) language and has a flexible word order) and may include nouns, adjectives, adverbs, and prepositional phrases (Karimi, 1997).

A long-lasting question has been whether Persian CPs are generated as lexical units or as phrases (Goldberg, 1996, 2003; Folli et al, 2005; Müller, 2010); in other words, whether they are the outcome of morphological or syntactic rules. Choosing one argument over the other has raised even further questions. If CPs are single lexical units, it is argued that the internal structure of CPs should be inaccessible to syntactic rules (Steele, 1989; Booij, 2007: 67) so that other elements cannot separate the parts of a compound. This assumption, however, is in apparent contradiction with Persian CPs which can be easily separated by words belonging to various parts of speech (Megerdoomian, 2012). Some scholars have argued, therefore, that CPs are in fact phrases (Karimi-Doostan, 1997; Megerdoomian, 2002, 2012). An argument against the phrasal status of Persian CPs, however, has been that CPs resist separation by certain

---

[1] Persian verbs have a present and a past stem that are used for inflecting different tenses. These two stems are ***zan*** and ***zad*** for *zadan* and both will be seen in the example sentences. The stems are the host for various inflectional affixes.

elements such as adverbs (Goldberg, 2003). However, the interpretation of Persian CPs as lexical units fails to give consistent results since there are many instances where the preverb and the verb are separated by adverbs like in the sentence (1) below where *chune* [chin] *zadan* 'haggle/bargain' is separated by the adverb *khub* 'well':

(1)   *dust-am*   *ham*   **chune**   *khub*   **mi-zan-e**   *ham*   *pul*   *khub*   *mi-gir-e.*
       friend-1SG.POSS   too   chin   well   IPFV-hit-3SG.PRS   too   money   well   IPFV-take-3SG.PRS
       'My friend both haggles well and gets money well.'

It is no surprise that any attempt to distinguish separable CPs from inseparable CPs will fail if a lexicon-syntax dichotomy is assumed: some CPs tend to show more word-like features while others are more open to separation, and separability is scalar rather than discrete. On the contrary, this behavior is consistent with how the human mind handles categories as members along a continuum instead of all-or-nothing features (Bybee, 2013) as well as how "lexicon and grammar form a continuum of constructions" (Goldberg, 2006: 220; Langacker, 2008: 96). It is therefore more efficient to prefer continuity and gradience over discreteness or try, at least, to incorporate the two into an approach that could account for both discrete and scalar phenomena (Langacker, 2006).

Taking separability as a gradient phenomenon and adapting the framework of Construction Grammar, this article explores the separability of Persian CPs formed with the simple verb *zadan* 'hit' and investigates whether there are motivations behind how they behave in relation to preverb-verb separation. Networks of constructions are extended based on form/meaning similarities (Goldberg, 1995: 127), and it has been argued that highly frequent word strings may lose their internal structure and become autonomous (Bybee, 2006) during a process called chunking through which smaller units or chunks form bigger chunks and by repetition, the smaller chunks lose their identity and the whole chunk behaves as a lexical unit (Bybee, 2007). Karimi-Doostan claims that the morphosyntactic properties of the preverb are mainly responsible for CP separability and argues that only preverbs that are predicative nouns (e.g.,

komak 'help' vs. non-predicative nouns like *gush* 'ear') can be separated from the verb. However, this argument fails based on corpus data, and according to Fleischhauer and Neisani (2020) who claim that separation is related to compositionality. They, however, limit their arguments to adjectives while suggesting that compositional light verb constructions can be internally modified by adjectives. The present article, by contrast, argues that in addition to compositionality, which is also a matter of degree, frequency may play a role in separability as discussed in Bybee (2006). Therefore, the semantics of the CPs and their frequency will be assessed in relation to (in)separability. To this end, two hypotheses are put forward: 1) *the less compositional in meaning a CP is, the less likely it is to be used in a separated configuration*; and 2) *the higher the token frequency of a CPs is, the less likely it is to be used in a separated configuration*. Frequency analysis is a more practical measure to analyze separability while compositionality is a matter of degree and may vary from person to person depending on linguistic and encyclopedic knowledge as well as the context a linguistic item is used in (Butler, 1995). Compositionality should therefore be interpreted with caution.

The terms complex predicate and light verb construction are sometimes used interchangeably. However, complex predicate is preferred in this article because there is no consensus on how exactly one should define a light verb construction. For example, contrary to corpus data, Fleischhauer and Neisani (2020) claim that the preverb (the nonverbal element) in Persian constructions of this type cannot take the object marker *rā* or become the subject of a passive sentence. Arguments of this kind result from attempts to discretely distinguish LVCs from non-LVCs based on invented examples instead of investigating corpus data. The term CP avoids these confusions because it encompasses a wider range of constructions and is more consistent with cognitive approaches that consider category membership to be a gradient phenomenon (Rosch, 1978; Lakoff, 1987: 56). A form-function pairing is considered a construction as far as aspects of its meaning are not predictable from its component parts or if

it has sufficient frequency to be a learned pairing (Goldberg, 2019: 6-7; Langacker, 2008: 6).

Therefore, the data in this article vary from lower-frequency non-compositional to higher-frequency compositional CPs with different degrees of separation.

2. Materials and Methods

The verb *zadan* is the second most frequent simple verb in Persian (Karimi-Doostan. 1997: 91) and is used in CPs with a wide variety of meanings. The term CP is used interchangeably in two senses in this article; one is the schematic abstract preverb-verb combination which underlies the generation of all CPs in the language, and the other is the token attestations of this schema realized as a verb in combination with various preverbs. The CP tokens cited in this article come from two sources.[2] TalkBank Persian[3] is a corpus of 33,143,061 sentences[4] with a word count of 474,773,547 collected from blog posts covering both formal and informal registers. Examples of older Persian texts are taken from the Persian Language Database (PLDB)[5], which includes texts ranging from the 10th century to the contemporary era. A concordance search was used to find examples of CPs with a 5-token interval to the right of the verb to make sure that the search would find as many CPs as possible. The results (a 10-thousand-line export limit is set by Sketch Engine) were then imported to an Excel spreadsheet to remove duplicate sentences and other possible noise in the data. For each CP construction that had 100 or more attestations in the data, a sample of 100 was generated randomly using the Excel RAND function. Sampling was repeated for a number of CPs with very high frequency (e.g., *harf zadan*[6] [speech/letter hit] 'talk') to make sure that different samples would turn

---

[2] Some CPs were excluded because it was impossible disambiguate them from other constructions in the corpus such as *sar* [head] *zadan* 'pay a short visit' (where zadan seems to indicate the suddenness of the action) because of the frequent usage of the preverb in other constructions. Manual disambiguation through thousands of lines was not practical either. Other CPs may have been simply left out unintentionally.

[3] Available at https://www.sketchengine.eu/talkbank-persian-corpus/.

[4] A number of duplicates in the corpus had to be accounted for when gathering the data.

[5] Available at https://pldb.ihcs.ac.ir/. No precise information is available about the size of this corpus.

[6] All the Persian equivalents of *zadan* in the CPs are in infinitive form with 'hit-INF' removed to save space and avoid repetition.

similar results in terms of separation. Each sample was then manually counted for separated cases. This excluded separation by "negation and inflectional affixes, auxiliaries, modals and emphatic elements" (Megerdoomian, 2001) because such units can unequivocally separate the components of all CPs. To see whether there was a correlation between frequency and (in)separability, the data for CPs with 100 or more attestations were analyzed in the R software package (V.2022.07.2).

While CPs are constructions on their own, they can be grouped into larger families based on similarities in their meaning and/or form (Bybee, 2010: 57; Goldberg, 2019: 63). The existence of such families highlights the role of analogy in the emergence of new constructions. There might be little semantic compatibility between a preverb and the verb on the surface, especially with highly abstract notions expressed by a verb with concrete physical meaning (section 3), but diachronic data reveal how a CP has emerged and become the basis for analogy. In order to see whether meaning plays a role in (in)separability, the CPs were categorized in different families based on the meaning they express and their association with the semantic structure of *zadan*. The general idea behind the analysis is that the families can be arranged as a radial category on the basis of their semantic approximation to the prototype meaning of the verb, while within each family, the CP constructions can also be on a continuum of more central to more peripheral examples of that family (Bybee, 2010: 79). The constructions and families are linked together through cognitive mechanisms including metaphor, metonymy, foregrounding, and analogy (Lakoff, 1987; Lakoff and Turner, 1989; Goldberg, 1995) and form a constructional network. Therefore, the bottom-up analysis includes the CPs as constructions, the families including semantically similar constructions, and ultimately the network that connects these families to the prototype *zadan*.

3. Analysis

3.1. *zadan*

More than 70 entries are found for *zadan* 'hit' in *Farhang-e Bozorg-e Sokhan* (Sokhan Comprehensive Dictionary), suggesting that not only is *zadan* a highly common verb, but it also has a variety of usages and meanings, many of which are realized as CPs[7]. To account for these CPs, a brief description of the semantic structure of the polysemy of *zadan* is provided first, followed by a discussion of various constructions that are motivated by this structure. We will demonstrate that the separability of CPs may be related to the position of the constructions in the constructional network of *zadan*.

Determining the prototype sense of *zadan* presents a challenge because the verb has two general senses: 'hitting two things against each other' and 'beating someone'. Since the latter entails the former (beating is a sort of hitting), it makes sense to move from general to more specific senses and consider *hitting* as the prototype sense. Moreover, concrete physical senses tend to be the basis for figurative extensions (Lakoff, 1993). Thus, one could argue that the prototype sense of *zadan* is 'hit an entity (NP$_2$) against another entity (NP$_3$) with force', and it can be realized syntactically as:

(NP$_1$) (NP$_2$)$_{DO}$ (NP$_3$)$_{CO}$ VP[8]

where NP$_1$ is the subject, NP$_2$ is the direct object (an instrument), and NP$_3$ is the item impacted by the blow. An example would be:

(2) *tup* (NP$_2$)   *ro*   **zad**   *tu surat-e   ronaldo* (NP$_3$). (an Instagram post)
     ball   OM   hit[3SG.PST] in face-EZ[9] Ronaldo
     'He kicked the ball into Ronaldo's face.'

---

[7] The CPs are not obtained from the dictionary.

[8] As a scrambling language, the default order of Persian sentences can be swapped in various ways depending on the context.

[9] Ezafe is a grammatical particle in Persian that is used to show possession and modification. It is generally realized as 'e' or 'ye' and is close to English 'of'.

The event expressed by *zadan* in (2) can be broken down into several semantic elements. First, it requires an agent (grammatical subject)[10], a theme (direct object, *tup* 'ball'), and a goal (complement, *surat-e ronaldo* 'Ronaldo's face') as the core elements. Many CPs, however, have a reduced argument structure (Megerdoomian, 2001) where even the subject is no longer an agent (e.g., the grammatical subject of *vāksan* [vaccine] *zadan* 'get vaccinated' is a patient). This is due the fact that *zadan* is used in constructions where it is not the argument structure, but other semantic components that are foregrounded. The verb *zadan* can be seen as a non-prototypical motion verb in which the goal is hit by the theme in a special manner: a single hit is a MOMENTARY and RAPID action over a SHORT period of time which also allows for quick repeated actions. The action usually requires a considerable amount of FORCE and requires that the theme and the goal come in CONTACT (McKoon and Love, 2011), otherwise it would be odd to use *zadan*. Example (3), although not ungrammatical, is marginal at best in Persian.

(3) *tup   ro   **zad**   tu surat-e   ronaldo ammā   na-khord   beh-esh*.
    ball   OM   hit[3SG.PST] in face-EZ   Ronaldo but   NEG-collide[3SG.PST] to-him/her
    !'He kicked the ball into Ronaldo's face, but it didn't hit him.'

In short, aside from the argument structure, the manner of action including momentariness, speed, brevity, and contact is also part of the meaning of *zadan*. In addition, the theme is usually but not necessarily smaller in size than both the agent and goal, and other elements like an instrument can be used to perform the action, but these are secondary to the meaning of *zadan* as they may or may not be present in a prototypical event.

3.2. *zadan* families

This section provides a description of the various usages of *zadan* along with the radial category showing the constructional network of these usages. In order to get a better idea of how the

---

[10] Persian is a pro-drop language. Subjects can be omitted because their reference is realized as a suffix on the verb.

families of the constructions are related, the final network of these families is presented in Figure (1) below, but its details will be discussed in section 3.3.
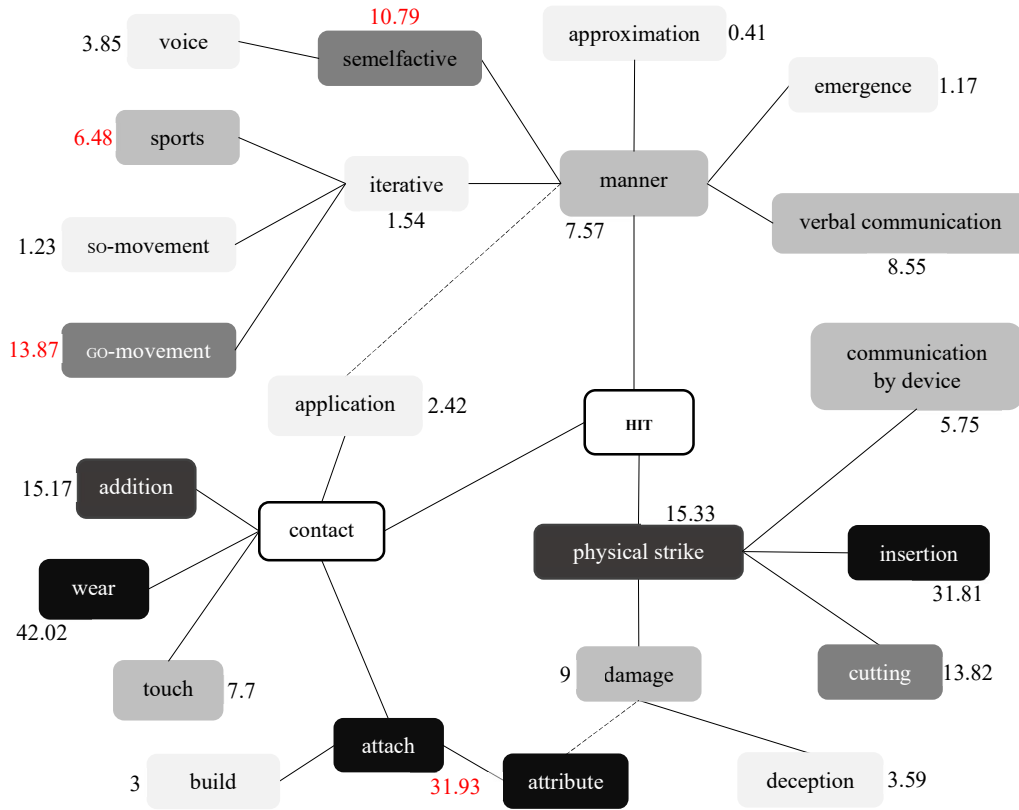

Figure 1. Constructional network of *zadan.* Darker shades show higher separability.

3.2.1. Physical strike

This family of constructions is closely related to the prototypical meaning in that the agent hits the goal using a body part (theme). The preverb incorporates the theme by identifying various types of hitting like a punch or slap or more general terms like beating and hitting. The action is mainly performed to commit a forcible assault but can include cases where the force is marginal or figurative interpretations are intended. For example, the concrete CP *mosht* [fist/punch] *zadan* 'punch' has a figurative implication in sentence (4) and does not include any kind of physical activity. The CP is also separated by four words including an adjective that modifies the preverb *mosht* and a prepositional phrase that shows the target:

(4) *mellat* **mosht**-*e mohkam-i be tablighāt-e doshman **zad**.*
     nation fist-EZ firm-INDF to propaganda-EZ enemy hit[3SG.PST]
     'The nation punched the enemy's propaganda hard (counteracted the propaganda).'

Preverb	Preverb equivalent	CP equivalent	Separations (n)	Total
nok	beak	peck	1	182
kotak	stick, beating[11]	beat	3	1919
zānu	knee	kneel, accept defeat	3	1137
nish	sting	sting	4	641
chang	claw	claw, grasp	4	1222
tane	body	jostle	8	285
zarbe	hit/blow	hit	17	182
lagad	kick	kick	24	601
talangor	flick	flick, stimulate	25	168
mosht	fist	punch	35	440
sili	slap	slap	45	1023

Table 1. The family of PHYSICAL STRIKE CPs.

The CPs are ordered in Table (1) by the number of separations attested in a sample of 100 sentences. Note, however, that the numbers are mostly an indication of how flexibly the CPs behave in terms of separation and therefore, zero attestation in subsequent tables does not necessarily imply inseparability.

3.2.2. Damage

A physical strike is carried out to impose physical harm or damage to the goal, as the CPs in the previous section show. By metonymic extension, *zadan* is also used in CPs where the event may or may not be physical, but the resulting state would be similar to physical harm and impose a negative condition on the goal. The relation is evident in *zarbe zadan* 'hit, damage', which bridges the transition between PHYSICAL STRIKE (5a) and DAMAGE (5b) as it is usable in both contexts:

(5) a. *yek* **zarbe**-*ye ārām be gorbe **mi-zan-ad**.*
        one hit-EZ gentle to cat IPFV-hit-3SG.PRS
        'S/he kicks the cat gently.'

     b. *ānhā ghasd-e **zarbe zad**-an be keshvar rā dār-and.*
        they intention-EZ hit hit-INF to country OM have-3PL.PRS
        'They intend to damage the country.'

Preverb	Preverb equivalent	CP equivalent	Separations (n)	Total
sadame	injury	damage, hurt	3	1181

---

[11] *kotak* has gone through a metonymic shift from 'a stick used for beating' to just 'beating'.

gand	stench	ruin, screw up	5	701
āsib	damage	damage, hurt	7	1362
latme	hurt	damage, hurt	7	1751
khesārat	detriment	cause loss	14	245
zarar	loss	damage	19	257

Table 2. CPs in the family DAMAGE.

3.2.3. Deception

There is a third family type that shares the property NEGATIVE IMPACT with *zadan* but differs in that it includes CPs that do not denote force (unlike DAMAGE where force and direct hitting could be present) and the preverbs do not designate entities that could be imposed through physical motion or force. Here, *zadan* suggests a negative state as the result of the action of an agent.

Preverb	Preverb equivalent	CP equivalent	Separations (n)	Total
nāro	betrayal	betray	0	63
rudast	deceit	outsmart	0	30
gul	deception	deceive	0	1363
rakab	knee/deceit	mislead	1	17
hoghe	trick	trick	6	52
kalak	stick, cheating	mislead	6	158

Table 3. CPs in the family DECEPTION.

3.2.4. Cutting

CUTTING seems to be a metonymic extension of usages like *gardan* [neck] *zadan* 'behead' which requires hitting hard with a sharp instrument against someone's neck. What is backgrounded in the process though is the actual hitting. Thus, *zadan*, both as a simple verb (like in sentence (6) below) and in CPs, is used in contexts where something is cut with a sharp object regardless of the force like in sentence (6) where hair can be cut with an electric hair trimmer:

(6) *diruz   raft-am   ārāyeshgāh   sar-e   mu-hām-o   ye   kam   **zad-am**.*
  yesterday go-1SG.PST barbershop   head-EZ   hair-1SG.PL.POSS-OM one   little hit-1SG.PST
  'I went to the barbershop yesterday [and] had my hair cut a little.'

Preverb	Preverb equivalent	CP equivalent	Separations (n)	Total
tigh	blade	shave	0	241
ghalame	scion	perform stem cutting	0	95
boresh	incision	incise, cut	6	401
gardan	neck	behead	17	502
rish	beard	shave	28	62
rag	blood vessel	cut one's wrist	28	162

Table 4. CPs in the family CUTTING.

12

3.2.5. Insertion

Collocation of prototypical *zadan* with nouns that denote sharp pointy objects has motivated the use of this verb in the sense of INSERTION. This usually includes a single hitting-like action as in the example (7) where a plug is inserted into a socket and can also be seen in frequent highly separable CPs in Table (5). Through metonymic extension, INSERTION is also extended to 'pumping fuel' which involves putting the pointy nozzle into the tank (Table 5).

(7)   *avval doshākhe   rā be   piriz-e   bargh   be-zan-id*.
      first   plug   OM to   socket-EZ   electricity   IMP-hit-2PL.PRS
      'First, insert the plug into the socket.'

Preverb	Preverb equivalent	CP equivalent	Separations (n)	Total
gāz	(natural) gas	pump gas	*12	*
nitrozhen	nitrogen	pump nitrogen	-	-
gāzo'il	diesel fuel	pump diesel fuel	1	8
benzin	gasoline	pump gas	13	184
serom	serum	get an IV	*	*
āmpul	ampoule	inject	40	223
vāksan	vaccine	vaccinate/get vaccinated	44	161

Table 5. CPs in the family INSERTION.

3.2.6. Communication

The verb *zadan* is used in two major constructions that convey COMMUNICATION, namely VERBAL COMMUNICATION and COMMUNICATION VIA DEVICE, and both seem to have originally emerged in relation to different aspects of this verb. It is very likely that VERBAL COMMUNICATION is rooted in the quick movements of the jaw as one of the main prerequisites of talking as seen in the archaic CP *zanakh* [chin] *zadan* 'talk' dating back as far as the 12th century (Dehkhoda dictionary) or more recent CPs *chāne* [chin] *zadan* (Karimi, 1997) 'haggle' and *fak* [jaw] *zadan* 'ramble on'. On the other hand, *harf* [speech] *zadan* 'talk' dates back approximately to the 14th century[13]. The CP *telegrāf* [telegraph] *zadan* 'send telegraph' in COMMUNICATION VIA DEVICE dates back to

---

[12] The asterisk mark indicates that it was not possible to disambiguate and isolate instances of a CP in the corpus because of homonymy, while a dash line means that a particular CP was not found in the corpus, but it could be found in other sources. Such CPs were only added to show different usages of *zadan*.
[13] No instances of *harf zadan* were found in earlier texts in the PLDB corpus.

the 19th century with the introduction of telegraph possibly also motivated by how telegraph messages were produced (hitting using a handle).

VERBAL COMMUNICATION				
Preverb	**Preverb equivalent**	**CP equivalent**	**Separations (n)**	**Total**
chāne	chin	haggle	0	514
negh	mutter	mutter	0	450
fak	jaw	talk nonstop	1	88
topogh	mistake	bumble	1	63
ghor	grumble	grumble	3	1077
tashar	snap	snap at someone	3	148
gap	chat	chat	4	823
kenāye	irony	talk sarcastically	7	226
bolof	bluff	bluff	8	88
ta'neh	sarcasm	talk sarcastically	9	368
lāf	boast, claim	brag, claim	21	266
zer	nonsense	talk nonsense	21[14]	180
harf	letter, speech	talk	28	9029
COMMUNICATION VIA DEVICE				
zang	bell	call	0	8125
telefon	telephone	call	5	1149
imeyl	email	send email	9	411
payāmak	short message	text	9	195

Table 6. CPs in the families VERBAL COMMUNICATION and COMMUNICATION VIA DEVICE.

Two interesting CPs are *lāf/harf zadan* which, when separated, may have different meanings from nonseparated counterparts ('claim' versus 'brag'; 'say' versus 'talk'). For example, *harf zadan*, a very common CP in Persian, means 'talk' in sentence (8a) while it means 'say' in (8b) where the preverb is modified by the number two and separated from *zadan* by the adjective *hesābi* 'sensible'.

(8) a. *kamtar* **harf be-zan-im** *va* *bishtar* *gush* *kon-im*.
      less speech SBJV-hit-1PL.PRS and more ear [SBJV]do-1PL.PRS
      'We should talk less and listen more.'

     b. *kollan* *do* *tā* **harf**-*e* *hesābi* *tu omr-esh* **zad-eh**.
      totally two CLF speech-EZ square in lifetime-3SG.POSS hit-PRF
      'S/he has only said two sensible things his/her whole life.'

---

[14] Although the number is high, separated cases were almost exclusively separated by two adjectives *moft* 'free' and *ziadi* 'redundant'. This suggests that CPs also have different degrees of separability in terms of the type of elements that separate the preverb and verb. The CP *zer-e moft/ziadi zadan* actually means 'talk nonsense, lie, or make false accusations' depending on the context and an analogy to *harf-e moft/ziadi*, with the same meaning.

It is also worth noting that some of the CPs in table (6) have a negative connotation (e.g., *ghor zadan* 'grumble'). These CPs highlight the negative impact of *zadan* on the target, in this case the audience, and further support the interrelation of different usages in a semantic network.

3.2.7. Contact

It was stated in section 3.1 that a prototypical event of *zadan* requires the CONTACT of the theme and goal (McKoon and Love, 2011). The family of CPs that follow can be grouped under an umbrella term called CONTACT because they include in one way or another the element of contact in addition to other semantic elements of *zadan* like speed and shortness of the action.

3.2.7.1. Touching

The CPs here are mainly used to describe events of TOUCHING which involve a body part (theme) coming in contact with an entity (goal or location) in a hitting-like manner. The intensity of the force applied is minimal or marginal while the focus has been shifted to the endpoint of the path schema (Johnson, 1987: 113-117; Lakoff, 1987: 278), namely the theme reaching and touching the goal.

Preverb	Preverb equivalent	CP equivalent	Separations (n)	Total
dast	hand	touch	*	*
angosht	finger	leave fingerprint as a signature	*	*
lis	lick	lick	0	167
nākhonak	small fingernail	sneak a taste	1	61
emzā	signature	sign	12	152
buse	kiss	kiss	15	960

Table 7. CPs in the family TOUCHING.

The less transparent CP *emzā zadan* 'sign' is included here because it is directly related to and probably formed in analogy with *angosht zadan* as a way of signing documents (with the index finger) which is itself a metonymic extension of *hitting* the finger on a surface. Example (9) shows these two CPs joined and used in a phrase.

(9) *majbur-am kard   bargeh-hā-ye   sefid rā **angosht** va   **emzā   be-zan-am***.
forced-1SG do[3SG.PST] paper-PL.EZ   white OM finger   and signature SBJV.hit-1SG.PRS
'S/he forced me to sign and put my fingerprint on unfilled papers.'

3.2.7.2. Addition, Application

There are a number of CPs that in addition to the notion of CONTACT, foreground the use of an instrument or the theme that is going to 'reach' a particular location. What further motivates the use of *zadan* is the manner in which these instruments are used in that they mostly represent repetitive quick actions with constant brief hits to the surface. The CPs are shown in Table (8) under the family APPLICATION because they convey the application of a device or instrument against a particular location. It is no surprise that some CPs have marginal similarity to other CPs in a construction since they are the result of analogy as a productive source of new constructions (Bybee, 2010: 57). Examples are *seshuār* [hairdryer] *zadan* 'blow-dry' and *nakh* [thread] *zadan* 'floss' which seem to be analogous to *shāne* [comb] *zadan* 'comb' and *mesvāk* [toothbrush] *zadan* 'brush one's teeth' respectively.

APPLICATION				
Preverb	**Preverb equivalent**	**CP equivalent**	**Separations (n)**	**Total**
sombāde	sandpaper	sandpaper	-	-
nakh	thread	floss one's teeth	0	1
seshuār	hairdryer	blow-dry	0	1
lif	washcloth	use washcloth to wash the body	0	13
bores	brush	brush one's hair	0	45
mesvāk	toothbrush	brush one's teeth	1	638
otu	iron	iron	3	70
shāne	comb	comb	4	340
ADDITION				
vāks	(shoe) polish	polish	4	200
shāmpu	shampoo	shampoo	8	69
roghan	oil	add oil, lubricate	11	122
namak	salt	add salt	11	327
lāk	nail polish	put on nail polish	12	159
odkolon	cologne	put on cologne	17	56
rang	paint	paint	19	608
atr	perfume	put on perfume	28	254

Table 8. CPs in the families APPLICATION and ADDITION.

ADDITION is one of the most productive constructions of *zadan* to express the addition of specific materials (the preverbs) to another entity as long as the target entity is larger than those materials. The preverbs mostly refer, but not limited to liquids and edibles as well as other powder-like or fine-grained materials that can be easily carried and added to a desired target (Table 8 shows a sample of high frequency CPs). Although it is hard to find an association

between adding spices and *zadan* such as in (10a), the example in (10b) shows that splashing

water '*āb*' on face is literally similar to the prototypical *zadan* where the theme is hit against

the goal, thus turning *zadan* into a felicitous verb for splashing water and similar activities.

(10) a. *beh-esh* **advie** *va* **na'nā** **zad-am**.
        to-3SG spice and mint hit-1SG.PST
        'I added spices and mint to it.'

  b. *shir-e āb* *rā bāz* *kard-am va* *chand* *mosht* **āb** *be* *surat-am* **zad-am**.
     tap-EZ water OM open do-1SG.PST and several fist water to face-1SG.POSS hit-1SG.PST
     'I turned the tap on and splashed a few handfuls of water on my face.'

3.2.7.3. Attachment, Attribute

Similar to ADDITION, *zadan* is used for a similar productive semantic structure, the contact of a

smaller theme with a larger goal/location, to convey the sense ATTACHMENT. The difference, of

course lies in the nature of the theme and how the whole event is interpreted. In this sense,

zadan is used as a simple heavy verb (11a) where the verb designates a completely physical

activity. By metonymic extension, however, it can be used in CPs that convey the sense

ATTRIBUTE as in (11b) where *barchasb* [sticker] *zadan* 'label, accuse' is a CP.

(11) a. *farmān-e* *shāh rā be* *divār-hā* **zad-and**.
        order-EZ shah OM to wall-PL hit-PL.PST
        'They posted the Shah's order on the walls.'

  b. *be* *dowlat* **barchasb**-*e* *nezhādparast* **zad-and**.
     to government sticker-EZ racist hit-3PL.PST
     'They called the government racist/accused the government of racism.'

  c. *to* *rā bā* *safvān* **tohmat** **zad-e** *kard-e* *ast*.
     you OM with Safvan accusation hit-PRF do-PRF be-1SG.PRS
     'He has accused you [of having an affair] with Safvan.'

Preverb	Preverb equivalent	CP equivalent	Separations (n)	Total
bohtān	libel	calumniate	4	54
efterā	slander	slander	7	161
tohmat	accusation	accuse	22	1781
barchasb	sticker	label, accuse	48	505
ang	mark	label, accuse	64	350

Table 9. CPs in the family ATTRIBUTE.

Nevertheless, it seems that the CPs *barchasb/ang zadan* are used figuratively as the result of

both metonymic extension as well as analogy to existing CPs such as *tohmat zadan* which has

been in use since at least the 9th century where *zadan* appears to be related to the sense PHYSCAL STRIKE and DAMAGE as seen in example (11c).

As a further step, the CPs have developed into schematic constructions (Bybee, 2010: 25) in the form of '*preverb __ zadan*' where the slot can be filled with adjectives and NPs (e.g., *nezhādparast* 'racist' in 11b) that describe the property attributed to someone. In fact, the high number of separations is the direct outcome of this schematic construction.

3.2.7.4. *Wearing*

Closely associated with ATTACHMENT and ADDITION, some pieces of clothing or accessories can be used with *zadan* to convey the sense WEARING. Similar to prototypical *zadan* which is usually used with themes smaller than goals, the preverbs (theme) generally denote small accessories (e.g., a headphone or eyeglasses) that can be attached to the body or other pieces of clothing (Table 10). It is very unlikely, however, to use *zadan* for wearing jackets or pants. This argument is supported by the sentence in (12) where *kerāvāt* 'tie' and *kot shalvār* [jacket pants] 'suit' are used with the verbs *zadan* and *pushid-an* [wear-INF] 'wear' respectively.

Preverb	Preverb equivalent	CP equivalent	Separations (n)	Total
hedband	headband	wear headband	1	2
hedfon	headphone	wear headphone	3	4
cheshmband	blindfold	wear blindfold	7	31
pāpion	bow tie	wear bow tie	8	27
kerāvāt	tie	wear tie	25	193
māsk	mask	wear mask	37	139
tip	style	wear a particular type of	41	134
neghāb	mask	wear mask	50	170
eynak	glasses	wear glasses	65	319

Table 10. CPs in the family WEARING.

(12) *na-tunest   ghāne'-am   kon-e   kot-shalvār   be-push-am   o   kerāvāt   be-**zan**-am*.
NEG-can[3SG.PST] content-1SG   do-3SG.PST suit   SBJV-wear-1SG.PRS and   tie   SBJV-hit-1SG.PST
'She couldn't persuade me to wear a suit and a tie.'

3.2.7.5. *Building, Establishment*

Contrary to the previous CP families such as ADDITION and WEARING where productivity was limited to a subset of nouns, *zadan* is used frequently to convey the sense ESTABLISHMENT of an

entity as in (13a) where starting or establishing a business (*forushgāh* 'store, hypermarket') is expressed by means of *zadan*.

(13) a. *man   khod-am   dār-am   **forushgāh   mi-zan-am***.
         I   self-1SG   have-1SG.PRS   store   IPFV-hit-1SG.PRS
         'I'm opening my own store.'

   b. *ājor-hā   rā chid-and   hālā dār-and   saghf   rā   **mi-zan-and***.
      brick-PL   OM arrange-3PL.PST now have-3PL.PRSc ceiling OM   IPFV-hit-3PL.PRS
      'They finished the walls, now they are building the ceiling.'

The usage is of course a metonymic extension of the contexts where *zadan* denotes physically BUILDING a construction (theme) as in (13b), and the example (13a) may also be interpreted as building a store. The latter sense probably is associated with how something is built: bringing together and *attaching* the materials needed for setting up a construction. A similar situation is found in English where 'knock together' means assemble *hastily* as in 'A temporary stage has been roughly knocked together' (Riemer, 2005: 280) with more focus on the rapid manner of action.

Preverb	Preverb equivalent	CP equivalent	Separations (n)	Total
ordu	camp	camp (military)	0	271
halghe	ring	encircle, gather	0	1381
chādor	veil, tent	set up a tent, camp	3	371
kheyme	tent	set up a tent, dominate	5	291
pol	bridge	build a bridge, connect	10	314

Table 11. CPs in the family ESTABLISHMENT.

Table (11) only shows examples of CPs that, beside their literal meaning, can be used figuratively, such as *pol zadan,* which may refer to both the building a bridge and creating a metaphoric connection. However, as stated above, *zadan* is used frequently and mostly in colloquial speech in the sense ESTABLISHMENT as the construction '__ *zadan'* with the slot filled with NPs referring to buildings and businesses, including virtual ones (14):

(14) *ye   **peyj** barā-ye ānlāyn   shāp   **zad-am***.
      one   page   for-EZ   online   shop   hit-1SG.PST
      'I've launched a website for online shopping.'

3.2.8. Manner

It was mentioned in section 3.1 that prototypical *zadan* is generally a fast and short activity that can "happen in an instant in time" (McKoon and Love, 2011). The constructions in this section include CPs that are associated with *zadan* in terms of the manner in which a situation is conceptualized. For example, Table (12) shows two sets of CPs related to laughing and moving that convey repetitive, short, or sudden actions (verbs formed by analogy may not show these properties). Other more specific CPs are all categorized under the macro-family MANNER as EMERGENCE, APPROXIMATION, SEMELFACTIVE, and ITERATIVE.

Preverb	Preverb equivalent	CP equivalent	Separations (n)	Total
vul	wiggle	wiggle	0	12
mallagh	somersault	somersault	0	15
ghalt	roll	roll	2	244
ghahghahe	guffaw	guffaw	4	192
poshtak	somersault	somersault	6	49
khandeh	laughter	laugh	8	395
nishkhand	sneer	sneer	8	76
labkhand	smile	smile	12	6988
puzkhand	smirk	smirk	17	326

Table 12. CPs in the family MANNER.

3.2.8.1. Emergence

The family of EMERGENCE CPs (or more specifically *producing*) refers to situations where some sort of entity or material is seen to have appeared on a particular location. Although the process is gradual, the appearance is conceived to be rather sudden like what is seen in *tāval* [blister] *zadan* 'blister' or *zang* [corrosion] *zadan* 'rust'. This family also seems to be associated with the CP *birun* [out] *zadan* 'go outside, leave' where going out is a sudden and instantaneous action with no prior plans. Although there is a sudden movement involved in *birun zadan*, the CPs in EMERGENCE no longer involve movement (Table 13).

Preverb	Preverb equivalent	CP equivalent	Separations (n)	Total
zang	corrosion, rust	rust	*	*
shekarak	sugar grain	crystalize	0	46
shure	dandruff	get dandruff	0	17
shokufeh	blossom	blossom	0	101
javāne	sprout	sprout	0	968
yakh	ice	freeze	0	1998
kapak	mold	get mold	0	258
tabkhāl	fever blister	get fever blisters	1	18
tāval	blister	blister	3	240

jush	acne	get acne	4	285

Table 13. CPs in the family EMERGENCE.

3.2.8.2. *Approximation*

While there seems to be no relationships between the preverbs and *zadan* in Table (14) in contemporary Persian, the relationship could be found in the original meaning of the word *hads* 'guess', which is of Arabic origin. It is defined in eastern philosophy as reaching a conclusion quickly without deliberate thinking (Jorjani, 1991: 37) and through simple observations. The definition is apparently compatible with the semantic structure of *zadan* as denoting a quickly performed action. APPROXIMATION is not a productive construction, yet novel examples can be found with a similar meaning, such as example (15) (a blog post) where a quick comparison is expressed using *zadan*.

Preverb	Preverb equivalent	CP equivalent	Separations (n)	Total
andāze	size	measure	0	40
takhmin	estimation	estimate	0	2077
hads	guess	guess	1	4504

Table 14. CPs in the family APPROXIMATION.

(15) *ye* **_moghāyese_** **_be-zan_** *tu gugel mi-bin-i ke kodum behtar-e.*
one comparison IMP-hit-2SG.PRS in Google IPFV-see-2SG.PRS that which better-be[3SG.PRS]
'Compare [them] on Google [and] you'll see which one is better.'

3.2.8.3. *Semelfactives*

SEMELFACTIVES are defined as punctual events with no result state so that an entity is in the same state as it was before the event (Van Valin, 2005: 32-33; Smith, 1991: 55-58) like the English verb 'flash'. This line of argument is compatible with a number of CPs (Table 15) that have *zadan* as their verbal head to convey brief short-lived events which may include actually moving and hitting such as *pelk* [eyelid] *zadan* 'blink' or simply indicate the instantaneous nature of an event like *jaraghghe* [spark] *zadan* 'spark'. The latter is also used metaphorically as a schematic construction *jaragheh-ye ___ zadan* with the slot filled with NPs and the whole construction means to stimulate, trigger, or start the NP ('cold war' in 16).

(16) *āyā rusie* **_jaraghe_**-*ye jang-e sard rā* **_zad-e_** *ast?*

whether	Russia	spark-EZ	war-EZ	cold	OM	hit-PRF	be[3SG.PRS]

'Has Russia sparked the cold war?'

Preverb	Preverb equivalent	CP equivalent	Separations (n)	Total
bargh	lightning	gleam, flash	0	641
pelk	eyelid	blink	0	740
ra'd-o-bargh	thunderbolt	strike a thunderbolt	2	26
cheshmak	wink	wink, twinkle	4	663
jaraghe	spark	spark, activate	40	381

Table 15. CPs in the construction *semelfactive.*

Another group of CPs included in the SEMELFACTIVE family are referred to here as SOUND constructions because they convey the production or emission of different types of sounds (Table 15). They may not portray prototypical semelfactives, but examples from old texts suggest that *zadan* might have come to use to describe semelfactive events like the following excerpt from 12th century:

(17) **bāng** bar lashkar **zad** va goft tā sāken shav-and.
call on army hit[3SG.PST] and say[3SG.PST] so inert become-3PL.PST
'He shouted at the army to make them stop.'

Here, bāng 'shout' refers to a *hey* sound directed towards the army, which is another reason for using *zadan* as a goal-oriented verb. Having a negative impact on the addressee, the sentence also implies STRIKE or DAMAGE caused by *zadan*, which is also implicit in some but not all the examples in Table (16). The manner of the event is also pointed out by other researchers including Bashiri (1981) and Family (2006) who claim that CPs such as *dād zadan* [shout hit] and *dād keshidan* [shout drag], both meaning 'to shout' only differ in the length of the event they denote with the latter describing a longer event.

Preverb	Preverb equivalent	CP equivalent	Separations (n)	Total
dād	shout	shout	-	-
sedā	sound	call (a name)	0	3538
jār	shout	announce, tell everyone	0	363
zār	desperation	cry loudly	1	559
na're	roar	shout, roar	4	215
faryād	outcry	shout	6	6956
jigh	scream	scream	8	840
sut	whistle	whistle	8	1046

Table 16. CPs in the family SOUND.

3.2.8.4. Iterative

Contrary to the SEMELFACTIVES that are very brief single events, ITERATIVE events are those that could happen repeatedly without any interruptions. In other words, they are repetitions of a semelfactive event in a single occasion (Dik, 1997: 236). The CPs in Table (17) represents different types of ITERATIVE events grouped together by similarity.

ITERATIVE				
Preverb	**Preverb equivalent**	**CP equivalent**	**Separations (n)**	**Total**
mak	suck	suck	0	57
mowj	wave	surge	0	1349
dāman	skirt	provoke, escalate[15]	0	6150
bāl	wing	flap	0	456
par	feather	flap, fly	0	1169
susu	shimmer	shimmer	0	104
nafas nafas	pant (heavy breathing)	pant	0	300
lah lah	pant (tongue sticking	crave, act greedy	4	121
ghol	boiling [with bubbles]	boil	8	31
SPORTS				
dambel	dumbbell	dumbbell lift	0	16
tanāb	rope	jump rope	0	103
vazne	weight	weightlift	14	205
MOVEMENT GOAL-ORIENTED				
pā	foot	pedal	-	-
hendel	handle	crank	0	10
rekāb	pedal	pedal	0	301
kelid	key	start	2	1267
estārt	start	turn the car's ignition, start	41	473
MOVEMENT SELF-ORIENTED				
dowr	turn	wander, turn	-	-
ghadam	step	stroll	0	4460
darjā	in place	march in place, thread	0	418
parse	rambling	ramble	0	978
charkh	wheel, roll	wander	0	377
sag dow	[dog run] rat race	work nonstop with no result	0	68
gasht	excursion	roam, patrol	7	253

Table 17. Families of different CPs with ITERATIVE meaning.

A prime example of an iterative event is *susu* [shimmer] *zadan* 'shimmer' which is a repeated instance of the semelfactive verb *cheshmak zadan* 'twinkle'. However, not all the CPs are used in this way especially those that seem to be formed through analogy. For instance, *estārt* [start] *zadan* 'turn the car's ignition' is an analogy to *hendel* [handle] *zadan* 'crank' which is itself an analogy to *pā* [foot] *zadan* 'pedal] (the analogy argument makes sense considering the chronology of the invention of bicycles and using hand cranks before electric starters were invented). It is also used frequently (26 of the 41 separated instances in Table 17) as a non-

---

[15] A metonymic extension of moving the skirt to inflame a fire to escalating an adverse situation.

iterative schematic construction *estārt-e ___ rā zadan* with slot filled with event-denoting NPs like the sentence (18) below where the slot is filled with *ghahremāni* 'championship':

(18)   *ghadam-e   avval rā mohkam bardār-im   va   **estārt**-e ghahremāni ra **be-zan-im**.*
        step-EZ   first   OM firm   pick-1PL.PRS   and   start-EZ   championship OM IMP-hit-1PL.PRS
        'Let's take the first step firmly and go for the championship [title].'

3.3. Separability

The main purpose of this article is to assess (in)separability in relation to both meaning and frequency of *zadan* CPs using data from the TalkBank Persian corpus. As concerns semantics, Figure (1) depicts the constructional network or radial category of the CPs discussed in this article. It is, however, by no means a complete description of the wide network of *zadan*. There are other CPs that may have been overlooked unintentionally or not found in the corpus. Each family of constructions is linked to *zadan* through various mechanisms as described in the previous sections, and conveys a special meaning based on the contributions made by the preverb and the verb. The shading in the boxes in Figure (1) represents degrees of separability with darker shades indicating more separability. This is based on the relative frequency (the total number of separations divided by the total number of the samples in each family) of separated CPs (written next to each construction) instead of the total number of separated cases because each family has different numbers of CPs. It could be argued that the more a CP is distant from the prototypical meaning of *zadan*, the less likely it is to be used in separated configurations. This is most evident with the constructions that are linked to *zadan* through its semantic component MANNER. On the other hand, those constructions that are closer to the prototype meaning and include physical CONTACT are more likely to be separated since prototype *zadan* requires at least the contact of two entities (McKoon and Love, 2011).

The presence of schematic constructions with a free slot to be filled by different linguistic elements results in higher numbers and darker shades for SEMELFACTIVE and GOAL-ORIENTED MOVEMENT that would be expected to have less relative frequency because of lower

compositionality.[16] This is not to imply that dark-shaded families do not include CPs with very low separation rate or the other way round, but that schematic CPs can already be separated to a large extent and affect the relative separability of a construction family.

In fact, since construction membership is gradient, it is no surprise that different CPs in a single family also show different degrees of separability. However, it is rather complicated to generalize over the CPs in each family because the construction families are generally small in size, and usage patterns of the CPs and even the preverbs vary significantly. For instance, the count of *bohtān* 'libel' was less than 500, while it was more than 8000 for *tohmat* 'accusation' in the TalkBank Persian corpus (Table 9). More data could provide more reliable results, but a general observation is that separation of a CP depends on whether it is possible to arrive at its meaning compositionally, if the CP has been developed into a schematic construction, or if it is part of a larger productive construction (type productivity). The latter means that the verb has acquired a new conventional meaning (e.g., ADDITION) and can be combined with different nouns as far as there is a similarity between new constructions and those already attested in a family of constructions (Goldberg, 2019: 63) or in other words, the meaning is compositional. This is more evident with WEAR and ADDITION where the constructions are partially productive in that new forms are accepted as long as the preverb is similar in both meaning and physical properties to attested examples.

In addition to the semantic properties of the constructions, another hypothesis was that there is a negative correlation between frequency and separability in that more frequent forms tend to be entrenched and therefore less exposed to internal modification (Bybee, 2007). In order to test this hypothesis, first the CPs with token frequencies of less than 100 were removed from

---

[16]If *estārt zadan* 'start' and *jaraghe zadan* 'spark' are removed as schematic CPs from SEMELFACTIVE and GOAL-ORIENTED MOVEMENT families, the number will drop from 13.87 and 10.79 to 0.95 and 1.84 percent respectively, making them congruent with the low separability argument (the same is also true for ATTRIBUTE CPs as schematic constructions).

the data since lower frequencies do not give a reliable enough indication of separability. Pearson's correlation coefficient test was run and at a p-value = 0.02, the correlation coefficient was calculated at -0.21. The r-value suggests that although there is a negative correlation between token frequency and separability, the correlation is weak. However, the raw count (n=159) of separations for CPs with 1000 token frequency and above (near 20 percent of the data) shows that high frequency CPs tend to be less separated. On the other hand, the bottom 20 percent of the CPs (frequency between 100 and 225) have a raw count of 432 which is 2.7 times higher than the top 20 percent (Table 18).

	Token frequency	Total separations in the samples	Total attested tokens
lower 20 percent	100-225	432	4586
upper 20 percent	1000 and above	159	78184

Table 18. Separation and frequency at the two extremes of the continuum.

The middle cases, however, do not follow a consistent pattern and it is therefore not possible to make a generalization based on frequency except that at the two extremes of the frequency continuum, CPs tend to show a reverse correlation with separation. Yet, the observation may also support the gradient nature of separability of CPs with extreme cases showing two different behaviors while the middle cases incorporate a combination of the two. In addition, this behavior further supports the claim that constructions are the result of chunking (Bybee, 2010: 36; Dąbrowska, 2004: 223). Linguistic items are fused into each other as the result of frequent co-occurrence and therefore create larger chunks. On the other hand, chunks that are used frequently may undergo entrenchment and become established, conventional units (Langacker, 2008: 16-17) in speakers' minds. This process, for example, could be the reason *zadan* has acquired the meanings 'attach', 'add' or 'wear': frequent usage with specific nouns resulting in a larger chunk as *X zadan*. Another effect of frequency, according to Bybee (2013, 2007) is that highly frequent chunks may further become autonomous and obtain a lexical status. Therefore, the fact that CPs at the higher frequency spectrum show less separability as opposed to the

lower frequency spectrum could be an indication of the effects of chunking and frequency on Persian CPs or at least on *zadan* CPs. However, it should be noted that at the end, any conclusion is open to argument because as Schmid (2017) argues "the outcome of chunking differs depending on whether the language has fixed or flexible word order". Persian word order is remarkably flexible[17] and this possibly is a factor in competition with chunking and frequency effects.

5. Discussion

This article explored the separability of the CPs formed with *zadan*, the second most frequent verb in Persian (Karimi-Doostan, 1997: 91) with a complex semantic structure that expresses a wide range of concepts from concrete, physical to completely abstract notions. To this end, a semantic analysis of the CPs was proposed based on the semantic structure of *zadan* and the CPs were categorized in different constructions presented as a radial category (Figure 1). We argued that the constructions that are closer to the physical meanings of *zadan* are more open to preverb-verb separation than those that are extended usages of *zadan* based on the manner properties of the prototype usage such as speed, repetition, and momentariness. In addition, CPs in more productive families are more open to separation since the meaning of the construction can be accounted for compositionally. Therefore, the first hypothesis, stating that CP (in)separability is correlated with compositionality is supported although within each family CP separability is also a matter of degree since all CPs do not show the same behavior. This is in line with Fleischhauer and Neisani (2020) who argue for the role of compositionality in adjectival modification of the preverb, but the present article shows that compositionality and separability are matters of degree and best described in terms of families of constructions instead of individual CPs. The second hypothesis about token frequency and (in)separability

---

[17] A simple ditransitive sentence like [*sārā*] [*ketāb ro*] [*be omid*] [*dād*] 'Sara gave the book to Omid' may be uttered in 14 different ways in Persian depending on the context.

cannot be supported affirmatively because of the weak correlation, but the interesting point is that in comparison, CPs on the higher end of the frequency range are less used in a separated configuration than those on the lower end, supporting the argument that highly frequent word strings may act as chunks and their internal structure may become opaque (Bybee, 2006; Matsumoto, 1999). It could be argued, therefore, that frequency at both extremes has a reverse correlation with separability.

The upshot of these arguments is that separability is not an all-or-nothing phenomenon that could be analyzed solely on the basis of meaning or frequency. This could call for further exploration of other factors that might be involved in how chunking and separation work in relation to CPs. Further research may investigate the correlation of compositionality and frequency to see whether CPs with higher token frequency are also semantically opaque. In addition, one should also consider information structure and the question of how much information is needed when referring to an event (Grice's maxims of quality and quantity). An accusation (Table 9), for instance, always implies a sort of wrongdoing that is expected to be provided as required information, but it is less likely to expect the same amount of information when talking about walking (Table 17). Further research, however, is required to answer the question, but a general idea regarding CPs, at least in English language, is that ease of modification might be a reason for frequent uses of CPs (Bonial, 2021) and it seems plausible to argue the same for Persian CPs. Even the CPs that most resist separation like *anjām dādan* [finish give-INF] 'do, perform' (Karimi-Doostan, 2011) can also be found in separated configurations since a general function of CPs is giving information in an efficient way. Another question is whether separability is related to the meaning of the preverb, the verb, or the whole construction. Karimi-Doostan (2011) argues that separability is solely determined by the semantics of the preverb, while ignoring the fact that even those CPs which he assumes to be inseparable can also be separated. However, his arguments fail to explain why it is possible,

for example, to modify the noun *ghofl* 'lock' with the adjective *bozorg* 'big' in the CP *ghofl zadan* 'lock' while the same does not hold true in *ghofl kardan* 'lock'. This behavior shows that separability cannot be attributed to only one aspect of a CP and that it is a function of the meaning of the whole construction. Another line of possible future research is to explore whether there are differences between CPs in the number of words that can come between preverbs and verbs. This is especially of consequence for natural language processing, word recognition and machine translation tools that need to be able to identify CPs even when a long string of words separates the preverbs from their verbs.

6. Conclusion

This article showed that the separability of Persian CPs is not an all-or-nothing phenomenon, and neither can it be accounted for based on a single factor or in relation to individual CPs. The notion of separability is best described in terms of families of constructions that are semantically related to the core meaning of the simple verb. The more these constructions are removed from the prototype meaning of the simple verb, the less likely it is that they will be used in separated configurations. The gradient nature of CP separability and compositionality is yet another reason that language corresponds to a lexicon-grammar continuum instead of discrete modules. Token frequency also shows that constructions at the two frequency extremes are in a reverse correlation with separability, suggesting that the internal structure of CPs of very high frequency becomes less accessible to syntactic modification through the process of chunking in which linguistic units develop stronger bonds through repetition.

Abbreviations

CLF	Classifer	OM	Object Marker
CO	Complement	PL	Plural
DO	Direct Object	POSS	Possessive
EZ	Ezafe	PRF	Perfect
IMP	Imperative	PRS	Present tense

INDF	Indefinite	PST	Past Tense
INF	Infinitive	REL	Relative
IPFV	Imperfective	SBJV	Subjunctive

References

Adli, A. (2010). Constraint cumulativity and gradience: Wh-scrambling in Persian. *Lingua, 120*(9), 2259-2294. doi:10.1016/j.lingua.2010.01.003

Alsina, A., Bresnan, J., & Sells, P. (1997). Complex predicates: Structure and theory. In A. Alsina, J. Bresnan, & P. Sells (Eds.), *Complex predicates.* Stanford: CSLI Publications.

Anvari, H. (2003). *Farhang-e Bozorg-e Sokhan (Sokhan Comprehensive Dictionary).* Tehran: Sokhan.

Bashiri, I. (1981). *Persian syntax.* Minneapolis: Burgess Publishing Company.

Bonial, C. (2021). Précis of take a look at this! Form, function, and productivity of English light verb constructions. *Colorado Research in Linguistics, 25*. doi:https://doi.org/10.33011/cril.v25i.1341

Booij, G. (2007). *The grammar of words.* New York: Oxford University Press.

Butler, K. (1995). Compositionality in cognitive models: The real issue. *Philosophical studies: An international journal for philosophy in the analytic tradition, 78*(2), 125-151.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language, 8*(4), 711-733. Retrieved from https://www.jstor.org/stable/4490266

Bybee, J. (2007). Diachrnoic linguistics. In D. Geeraerts, & H. Cuyckens (Eds.), *The Oxford handbook of cognitive linguistics* (pp. 945-87). New York: Oxford University Press.

Bybee, J. (2010). *Language, usage and cognition.* Cambridge: Cambridge University Press.

Bybee, J. (2013). Usage-based theory and exemplar representations of constructions. In T. Hoffmann, & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 49-69). New York: Oxford University Press.

Chomsky, N. (2018). Two notions of modularity. In R. G. de Almeida, & L. R. Gleitman (Eds.), *On concepts, modules, and language: Cognitive science at Its core* (pp. 25-40). New York: Oxford University Press.

Dąbrowska, E. (2004). *Language, mind and Brain: Some psychological and neurological constraints on theories of grammar.* Edinburgh: Edinburgh University Press.

Dehkhoda, A. (1998). *Loghatnāmeh (Dictionary).* Tehran: University of Tehran Press.

Dik, S. C. (1997). *The theory of functional grammar: The structure of the clause.* Berlin: De Gruyter Mouton.

Family, N. (2006). *Explorations of semantic space: The case of light verb constructions in Persian.* Paris: Ecole des Hautes Etude en Sciences Sociales.

Fleischhauer, J., & Neisani, M. (2020). Adverbial and attributive modification of Persian separable light verb constructions. *Journal of Linguistics, 56*(1), 45-85. doi:https://doi.org/10.1017/S0022226718000646

Fodor, J. A. (1983). *The modularity of mind: An essay on.* Cambridge, MA: MIT Press.

Folli, R., Harley, H., & Karimi, S. (2005). Determinants of event type in Persian complex predicates. *Lingua, 115*(10), 1365-1401. doi:10.1016/j.lingua.2004.06.002

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure.* Chicago: The University of Chicago Press.

Goldberg, A. (1996). Words by default: Optimizing constraints and the Persian complex predicate. *Proceedings of the 22th Annual Meeting of the Berkeley Linguistics Society* (pp. 132-146). Berkeley, CA: Berkeley Linguistics Society.

Goldberg, A. (2003). Words by default: The Persian complex predicate construction. In E. Francis, & L. Michaelis (Eds.), *Mismatch: Form-Function Incongruity and the Architecture of Grammar* (pp. 83-112). CSLI Publications.

Goldberg, A. (2006). *Constructions at work: The nature of generalization in language.* New York: Oxford University Press.

Goldberg, A. (2019). *Explain me this: Creativity, competition, and the partial productivity of constructions.* Princeton: Princeton University Press.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason.* Chicago: University of Chicago Press.

Jorjani, A. (1991). *Alta'rifāt (definitions).* Tehran: Naser Khosro.

Karimi, S. (1997). Persian complex verbs: Idiomatic or compositional. *Lexicology, 3*, 273-318.

Karimi, S. (2008). *Opening remarks: Scholarship on complex predicates*. International Conference on Complex Predicates. Paris: Universite Sorbonne Nouvelle.

Karimi-Doostan, G. (1997). *Light verb constructions in Persian.* Essex: University of Essex.

Karimi-Doostan, G. (2011). Separability of light verb constructions in Persian. *Studia Linguistica, 65*(1), 70-95. doi:https://doi.org/10.1111/j.1467-9582.2011.01178.x

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: The University of Chicago Press.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 202-251). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139173865.013

Langacker, R. W. (2008). *Cognitive grammar: A basic introduction.* New York: Oxford University Press.

Langacker, R. W. (2006). On the continuous debate about discreteness. *Cognitive Linguistics, 17*(1), 107-151.

Matsumoto, M. (1999). Composite predicates in Middle English. In L. J. Brinton, & M. Akimoto (Eds.), *Collocational and idiomatic aspects of composite predicates in the History of English* (pp. 59-96). Philadelphia: John Benjamins Publishing Company. doi:10.1075/slcs.47.24mat

McKoon, G., & Love, J. (2011). Verbs in the lexicon: Why is hitting easier than breaking? *Language and Cognition, 3*(2), 313-330. doi:10.1515/LANGCOG.2011.011

Megerdoomian, K. (2001). Event structure and complex predicates in Persian. *Canadian Journal of Linguistics, 46*(1-2), 97-125. doi:10.1017/S0008413100017953

Megerdoomian, K. (2002). *Beyond words and phrases: A unified theory of predicate composition.* Los Angeles: University of Southern California.

Megerdoomian, K. (2012). The status of the nominal in Persian complex predicates. *Natural Language and Linguistic Theory, 30*(1), 179-216. doi:10.1007/s11049-011-9146-0

Müller, S. (2010). Persian complex predicates and the limits of inheritance-based analyses. *Journal of Linguistics, 46*(3), 601-655. doi:10.1017/S0022226709990284

R Core Team (2022). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Riemer, N. (2005). *The semantics of polysemy: Reading meaning in English and Walpiri.* Berlin: Mouton de Gruyter.

Rosch, E. (1978). Principles of categorization. In E. Rosch, & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.

Schmid, H. (2017). A framework for understanding linguistic entrenchment and its psychological foundations. In H. Schmid (Ed.), *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge* (pp. 9-35). Berlin: De Gruyter Mouton.

Smith, C. S. (1991). *The parameter of aspect.* Dordrecht: Kluwer Academic Publishers.

Steele, S. (1989). The relation of morphology to syntax. *Annual Review of Anthropology, 18*(1), 157-175. doi:10.1146/annurev.an.18.100189.001105

Van Valin, R. D. (2005). *Exploring the syntax-semantics interface.* Cambridge: Cambridge University Press.