Mapping semantic spaces

A constructionist account of the "light verb" eat in Persian

Neiloufar Family
EHESS, Paris

Within the framework of Construction Grammar, the author develops an analytical tool to map semantic regularities in semantic spaces. This is illustrated with the study of the semantic spaces of the Persian "light verb" *xordæn* "to eat". The light verb's semantic space is populated by "notional islands" where groups of light verb constructions, expressing similar notions, combine the light verb with a restricted, but large, class of preverbs. The analysis shows that each notional island possesses linguistic and cognitive properties that allow intuitive disambiguation. It provides new results about meaning construction, productivity, and compositionality, and the basis of crosslinguistic investigations for processes of verb formation as they have evolved in different languages.

Keywords: compositionality; Construction Grammar; eat; light verbs; meaning construction; notional islands; productivity; semantic space

1. Introduction

"Eat", or *xordæn*, is a common light verb (LV) in Persian, one of a dozen that form the core of the Persian verbal system. They can occur by themselves, but in far more instances they appear combined with an often nominal preverbal element to express verbal meanings that are usually far from their original meanings. Most such notions are expressed by simple verbs in other languages. In this study I will concentrate on the LV *xordæn* eat, in order to portray the type of structures that emerge in the semantic spaces of these verbs. These spaces consist of networks of what I have termed "notional islands" which include linguistic and cognitive parameters. Verb production is a semantic process, not syntactic, and a comprehensive analysis reveals semantic proximities between the verb *eat* in Persian and in languages which have different syntactic structure (See Bonvini; Boyeldieu; Hénault, this volume).

2. The Persian verbal system

One of the striking characteristics of the Persian verbal systems lies in its deceptively small number of simple verbs (for a comprehensive grammar of Persian see Lambton 1961; Lazard 1992; Mahootian 1997). There are less than two hundred simple verbs in this language, as opposed to several thousand in English or French. However, a limited set of around twenty of the simple verbs in Persian can occur as LVs in light verb constructions,[1] producing myriads of verbal notions.

Light verb constructions (LVCs) in Persian consist of a preverbal element (PV, usually a nominal, though it can also be an adjective, adverb, preposition, or prepositional phrase) followed by an LV[2] (Vahedi-Langrudi 1996; Karimi-Doostan 1997; Megerdoomian 2002). The resulting meaning often deviates from the simple semantic sum of the original meaning of each of the constituents (Karimi 1997; Goldberg 1996).

dæst ZÆDÆN	hand hit	applaud, touch
xæmiyaze KEʃIDÆN	yawn pull	yawn
dust DAʃTÆN	friend have	like/love
guʃ DADÆN	ear give	listen
hærf ZÆDÆN	sound hit	talk
negah DAʃTÆN	sight have	keep
email ZÆDÆN	email hit	send an email
tayp KÆRDÆN	type do	type
qose XORDÆN	grief eat	grieve

Due to the quasi-compositional structure of the LVC (since the forms are not fully idiomatic, but motivated from the meaning of the constituents), the system presents an ideal architecture for exploring theories that attempt to link natural language semantics and underlying cognitive representations. Furthermore, this system provides ample data to study fundamental linguistic properties, such as compositionality, productivity, and polysemy. These properties are general properties, common to all languages and possibly the human conceptual system, but especially accessible for study given the structure of this system.

Persian LVCs are not fully compositional. As can be seen in the examples above, there isn't always a clear algorithm that allows the meaning of the whole to be derived

1. The verbs that occur in these constructions include, but are not limited to, *zædæn* hit, *keʃidæn* pull, *xordæn* eat, *gereftæn* obtain, *dadæn* give, *amædæn* come, *daʃtæn* have.

2. The syntactic behavior of these constructions is also particular, though not the focus of this study. Suffice it to say that the two elements tend to be inseparable in a sentence, though it is not uncommon for them to be separated by grammatical elements in some contexts (e.g., adverbs, verbal affixes, prepositional phrases, etc).

from the meaning of the parts. For example, using *xordæn* as our prime example, the term *ʃæmʃir xordæn* (lit. sword eat) means "to be stabbed and wounded by a sword." In this case, the idea of being penetrated by the weapon and being wounded does not come directly from either of the lexical items. It would be superfluous to say that part of the meaning of *xordæn* is "to stab." First, the general meaning of this and similar LVCs is much more nuanced, and can be described as follows: "to be penetrated and stabbed by a sharp hand held or projectile weapon like [preverb]." No natural language permits such nuanced meanings for particular morphemes. Second, such a claim would overlook the strict restrictions on the types of preverbal elements allowed in these constructions. The meaning of the LVC must be elsewhere.

One proposition might be that each multi-word LVC is stored independently in the lexicon. But this would cause massive redundancy (e.g., the morpheme *xordæn* would be repeated hundreds of times, once for each LVC it occurs in), since each LV can form many LVCs. But, more importantly, such a proposition ignores the fact that LVCs are *productive* in Persian. Each time a new verbal notion needs to be expressed, a new LVC is constructed. In French or English, a morpheme is frequently used with verbal affixes to produce a verb (e.g., bottle- to bottle, email- to email). In Persian, new lexical items occur with an LV (e.g., *quti kærdæn-* can do- "can", *emejl zædæn-* email hit- "email"). Therefore, there must be an established mechanism linking form and meaning that allows the production and the unambiguous comprehension of new verbs.

The LVs cannot be *polysemous*, because this would require highly idiosyncratic meanings like the ones mentioned above, and this wouldn't explain productivity. On the other hand, proposing that the LVs are semantically bleached does not resolve the problem either, since the PV would end up carrying the burden of a highly idiosyncratic and context dependent meaning. As we will see later, the different meanings expressed in LVCs based on a single LV are not totally unrelated, which indicates that the LVs contribute some semantic content. The meaning of each construction is motivated by both of its elements, but the specific nuances arise at a different level, namely, that of the construction.

The approach taken in this study provides a new perspective from which this type of linguistic structure can be analyzed. Before presenting the methods of this analysis, an outline of different meanings and uses of the verb *xordæn* will be presented.

3. The verb *xordæn*

Historically, the meaning of this verb has been *to ingest*, or *to eat*, and sometimes *to waste*. According to two dictionaries of Modern Persian, *xordæn* has a dozen different meanings: eat (usually after chewing), drink, gnaw, devour, waste or spend, corrode,

cause itching, make appear as used, being in the line of damage, receive, be beaten, take and never give back, hit, strike, touch, fit, match, be synchronized, and ending up somewhere (Haim 1995; Afshar et al. 2002).

Persian speakers can usually differentiate the full verb definitions in the above list of dictionary entries. The other, seemingly unrelated definitions are simply vague post-extractions of the meanings from a number of LVC uses of *xordæn*. Persian lexicographers and dictionary writers have great difficulty in clearly separating LV and full usage, and this produces inadequate entries for the verbs in question. This generates a poor rendition of the richness of the verbal system in Persian, given that there aren't many simple verbs and the entries for LV uses are incomplete. The contexts in which these meanings emerge are not given. We will see below that categorizing the meanings produced by these verbs requires a different method than listing some vague meanings that emerge in LVC contexts without specifying the restrictions on their interpretation. However, we will first examine uses of *xordæn* in non-LVC contexts, without trying to provide a comprehensive description of the vast amount of polysemy that occurs. The full verb polysemy of *xordæn* is beyond the scope of this article.

When asked for the meaning of *xordæn*, Persian speakers unanimously and immediately answer *to eat*. When expressing the action of *ingesting*, *xordæn* is transitive and takes a volitional subject, or in Dowty's terms[3] (Dowty 1991), a proto-agent argument. It also takes a second nominal phrase (NP) as the object that is being *ingested*:

XORDÆN (EAT)
Syntax: NP_1 NP_2 XORDÆN

(1) *Ali nan-ra*[4] *xord*
Ali bread-ACC ate
'Ali ate the bread'

3. In his article, Dowty suggests that traditional thematic roles (agent, patient, etc) are too rigid to be applicable to certain empirical data. He proposes proto-roles (*proto-agent* and *proto-patient*). Most proto-agents have the property of having a volitional involvement in an action, having sentience or perception, causing an event or change of state in another participant, or moving (relative to the position of something external). Proto-patients, on the other hand, undergo changes of state, are causally affected by another participant, or are stationary relative to the movement of another participant. Arguments might have traits that correspond to both these role types, but are assigned the proto-role from which they take most of their traits. I adopt this type of role assignment when discussing the arguments of the verbal elements under study.

4. RA = accusative or object marker. This morpheme usually marks a definite object of the verb, though its detailed definition has been investigated in several studies (Lazard 1970; Dabir-Moghaddam 1990; Ghomeshi 2003; Roberts 2005). Only in certain specific LVCs can the PV accept this morpheme. In the large majority of cases, the PV is a non-definite entity.

In Modern Persian, the verb for *drink*, *nuʃidæn*, is progressively being replaced by the verb *xordæn*, especially in less formal registers of the language. As with languages such as French, the verb for *drink*, *xordæn*, can also mean *to drink alchohol in excess* (*xejli xorde* – much eaten/drank- he has drank in excess) (see Boyeldieu this volume). *Xordæn* can also mean *to cause irritation*, or more precisely *to eat away at* (e.g., a material). In French, the word for itch is also derived from the verb *eat*, *manger* (*démangeaison*) (see Hénault this volume):

(2) *in boluz-e*[5] *pæʃmi tænæm-ra mi*[6]*-xor-æd*
this blouse-GEN wool body.1SG-ACC PROG-eat-3SG
'This wool blouse irritates my skin'

After *ingesting*, the second most popular and general definition of the verb *xordæn* is *to collide* or *hit* (French: *entrer en collision, être heurté*). In these cases it is an intransitive verb taking a proto-patient argument that collides with another entity (expressed in the propositional phrase headed by [*be*] *"to"*):

XORDÆN (HIT)
Syntax: NP_1 to NP_2 XORDÆN

(3) *dæst-æm be miz xord*
hand-1SG to table hit
'My hand hit the table'

This same syntactic structure is used when *xordæn* expresses *to match* or *fit and complement*:

(4) *ræng-e in kerevat be pirahænet mi-xor-æd*
color-GEN this tie to shirt-2SG PROG-hit-3SG
'The color of this tie complements your shirt'

This structure is also used in the more colloquial use of *xordæn*, expressing *reaching a position in space*:

(5) *in xijaban be otoban mi-xor-æd*
this street to highway PROG-hit-3SG
'This street hits the highway'

Though the two different argument structures don't seem to be related, the meanings each expresses is polysemous. For the purpose of this study, it is important to note

5. This morpheme, -e, is called the *ezafe* marker, expressing the genitive case, or possession. The use of this nominal marker has also been investigated at length (see Samiian 1994; Ghomeshi 1997).

6. PROG = progressive.

that the full verb *xordæn* is highly polysemous, like its counterparts in most other languages. Hence, the somewhat vague translation of *xordæn* with the word *eat* in English is for convenience only, and does not represent any global equivalence between the two lexemes in Persian and English. As the reader has been familiarized with the lexeme *xordæn*, for clarity we will note the LV use of *xordæn* as XORDÆN, since the exact translation is not possible.

The problem of classification goes beyond one of classical polysemy in LV uses of *xordæn*, where it occurs with a particular PV that contributes to the meaning of the whole construction. *Xordæn* adopts and conveys new meanings when it combines with particular types of PVs. These meanings are not directly related to its full verb meanings and are often difficult to isolate from the construction itself.

Initially, it is daunting to discern how speakers construct and decipher such varied constructions, such as:

qæza xordæn	food XORDÆN	eat
afsus xordæn	sorrow XORDÆN	be sorrowful
tækan xordæn	movement XORDÆN	jerk, shake, wag
tʃaqu xordæn	knife XORDÆN	get stabbed

Except for a handful of prepositions and prepositional phrases that form more idiomatic LVCs, only nominal PVs combine with *xordæn*. Our compiled corpus indicates *xordæn* as the basis of over 200 verbal notions, expressing meanings from *eating* to *being embarrassed* and *shaking*.

Xordæn produces mostly intransitive LVCs where the subject is a proto-patient undergoing a change of state or experiencing a state. The meanings of the actions expressed by this LV are generally ones of being effected, and usually have a negative connotation. *Xordæn* is an inchoative verb, that can be either telic or atelic,[7] depending on the type of action it expresses (undergoing an action or experiencing a state). The LV doesn't seem to have retained any of the lexical content of the full verb *xordæn*, but rather a figurative or metaphorical extension of the original meaning. The *eat* meaning of *xordæn* serves as a basis for an extension expressing the idea of *undergoing* an atelic action. In other cases, the notion of *collide* gives rise to undergoing telic or sudden actions.

Some of these constructions are close to pure idioms, where the meaning is strictly non-compositional and its productivity limited. The term *idiom* can be defined so that it covers only totally frozen expressions. The meaning of these expressions can't be remotely discerned from the meaning of its parts and are stored individually in the lexicon.

7. A complete action, or one that expresses an action that occurs "in X" time, is a telic action. On the other hand, an incomplete action, or one that occurs "for X" time, is atelic.

ja xordæn	place eat	be surprised, shocked
pærse xordæn	poor eat	be forced to beg
juʃ xordæn	boil eat	be anxious
kafur xordæn	campher eat	become impotent

There is also a set of truly transparent LVCs, where the meaning is compositional. The LV expresses its full verb meaning, *eat*, and the PV expresses a food. For example, any type of food occurring with *xordæn* will mean to ingest that type of food.

nan xordæn	bread eat	eat bread
ab xordæn	water eat	drink water
qæza xordæn	food eat	eat food

These two types of LVCs, idiomatic (semantically opaque) and compositional (semantically transparent), are marginal cases. The majority, and most interesting LVCs occur between these two extremes of semantic transparency, where most often, the contribution of the two parts cannot be fully discerned. The meaning of the forms is constrained by syntactic and sometimes idiosyncratic semantics, but the forms display more freedom (syntactic and semantic) than conventional idioms, allowing for productivity. These constructions have a status between idioms and grammatical phenomena, and are semi-transparent.

pa xordæn	foot eat	be stepped on
xis xordæn	wet eat	be soaked
otu xordæn	iron eat	be ironed
liz xordæn	slide eat	to slip

Karimi (1997) concludes that the majority of LVCs are *idiomatic combining expressions* whose idiomatic meanings are composed on the *basis* of the meaning of their parts (Karimi 1997: 23). She discusses certain non-productive LVCs with opaque meanings (the term CV refers to compound verb, or what we call LVC):

> Let us examine *châne zadan* (chin hitting=to negotiate) and *xar kardan* (donkey doing=to fool someone) with regard to their compositionality. Once we learn the meaning of these idiomatic CVs, we can see the relation between their parts. That is, *châne zadan* involves *figuration* indicating that *negotiation* requires *moving the chin*. *Xar kardan*, on the other hand, becomes transparent on the basis of the idiomatic meaning of *xar* "donkey" which implies *foolishness* and *stupidity*: there is an apparent relation between the concrete and the abstract meaning of *xar* that provides the idiomatic meaning of the CV *xar kardan*. (Karimi 1997: 24)

For *xordæn*, LVCs often portray images of swallowing, or being penetrated or pierced by something, or undergoing some process. More often than not, this process is to the detriment, and out of the control of the subject.

ʃæmʃir xordæn	sword XORDÆN	be wounded by a sword
dæst xordæn	hand XORDÆN	be touched, altered
zæxm xordæn	wound XORDÆN	be wounded

As mentioned in the previous section, it would be counter-intuitive to analyze each of the constructions as a separate lexical entry. By examining the large data[8] set closely, we can see that certain patterns emerge. Namely, groups of LVCs with the same LV and a particular type of PV cluster together to express highly correlated notions. In other words, clusters form "notional islands" in the semantic spaces of each LV. In the next section, we will discuss the development of the methodology used to explore these semantic spaces.

4. Theoretical methods

Faced with the massive data set from the Persian verbal system, with some verbs producing over 500 LVCs, an appropriate framework is needed to capture the emerging patterns. Exploring the mechanisms underlying meaning construction requires an approach that can accommodate idiosyncrasy, compositionality, productivity, and polysemy.

Cognitive linguistic approaches, especially Construction Grammar (CG) (Fillmore & Kay 1996; Goldberg 1995), provide an adequate framework for such studies. Formal theories only take into account the abstract structure of linguistic units and do not necessarily include all the semantic and pragmatic information associated with the structure.

Cognitive linguistic theories focus on the cognitive processes involved in language processing and understanding. Rather than focusing only on abstract and formal structures in language, cognitive linguists incorporate other areas of cognition into the study of language. These perceptual and conceptual parameters include space, time, and force-dynamics. In this view, language is a means through which humans, confined to physical bodies in particular environments, are able to express a range of phenomena, both perceived and conceived. In other words, language must be explained through other capacities other than just purely linguistic ones.

The basic tenets of CG include the following (Goldberg & Jackendoff 2004):

1. There is a cline of grammatical phenomena from the totally general to the totally idiosyncratic.

8. The data set used for this study was compiled using several dictionaries and texts. No set is ever comprehensive, since, as we will see, new forms are always possible.

2. Everything on this cline is to be stated in a common format, from the most particular, such as individual words to the most general, such as principles for verb position, with many subregularities in between. That is, there is no principled divide between "lexicon" and "rules."
3. At the level of phrasal syntax, pieces of syntax connected to meaning in a conventionalized and partially idiosyncratic way are captured by *constructions*.

The pairings between form and meaning in CG are often called *constructions*. These can be as small as words or affixes, or they can be whole sentence structures. Crucially, constructions include both universal and general knowledge as well as idiosyncratic and language-specific information. Accordingly, a construction exists if one or more of its properties cannot be strictly predicted from the inherent properties of the lexical elements within it. The description of a construction specifies the types of elements that can occur within it, as well as how they will interact with the construction. Constructions are similar to idioms. Idioms are listed in the lexicon with a syntactic structure, a meaning, and often phonological information.

Importantly for the present study, in verbal constructions, the verb alone does not determine the argument structure of the sentence. Rather, the argument structure results from the composite effects of the verb and the construction. This property of constructions allows for great reduction of verbal polysemy in the lexicon.

Most properties of constructions are based on semantic and syntactic information. However, CG also leaves room for experiential knowledge and its effect on the construction meaning. Real knowledge includes information a human has of how the world functions and what types of actions are pragmatically plausible.

Constructions can occur in "families," where a group of constructions share some syntactic and/or semantic properties. These families can be partially productive, based on semantic or pragmatic parameters. No claim is made that constructions or the parameters used to construct them are innate or universal, although there are presumably strong universal constraints.

This approach avoids redundant entries in the lexicon, as well as multiple and idiosyncratic senses for certain verbs:

> the reason for postulating constructions is analogous to the reason why other researchers have wanted to postulate a lexical rule: in order to capture generalizations across instances. Moreover, it is claimed here that what is stored is the knowledge that a particular verb *with its inherent meaning* can be used in a particular construction. This is equivalent to saying that the composite fused structure involving both verb and construction is stored in memory. By recognizing the stored entity to be a *composite* structure, we gain the benefits […] over a lexical rule account. For example, we avoid implausible senses such as "to cause to receive by kicking." It is the *composite* structure of the verb and construction that has this meaning. We also allow other syntactic processes to refer to the inherent lexical semantics of

> the verb. Thus we do not lose the information conveyed by the verb, because the verb is not changed into a new verb with a different sense. (Goldberg 1995: 140)

An analytical tool was developed, inspired by such approaches, to map semantic regularities in the semantic spaces of each LV in Persian. Our analysis implies that an LV's semantic space is populated by "notional islands" where groups of LVCs expressing similar notions appear by combining the LV with a constrained, but large, class of PVs. Each island possesses linguistic and cognitive properties that allow intuitive disambiguation. Similar studies have been done for English.

Wierzbicka (1982) provides a detailed sketch of the different *have a V* constructions and then compares them with each other. If the complement of the verbal complement belongs to one of the categories defined in a construction, the meaning of the whole can be predicted from the meaning of the construction. The existence of these types of constructions allows for productivity, since the type of complement that occurs with *have* in each of the constructions is highly specific and defined within the construction. One can imagine making new forms based on this structure.

Another study related to this issue deals with the productivity and acceptability of certain complements with a given LV in English (Stevenson et al. 2004). Stevenson et al. find that the complements that occur in particular constructions belong to a particular semantic class. In other words, the constructions are again found to partially depend on and partially determine the type of complement that combine with the LV.

In analyzing the list of Persian LVCs of a given LV in our corpus, the first task was to discover semantic resemblances and sets of constructions, as with the different *have* constructions in English in the above study. Variation of the type of PV, as defined by their common properties, results in a variation of the meaning of the LVC. Conversely, variation of the LV also results in the variation of the meaning of the LVC. It is essential to look both for PVs that combine with multiple LVs on one hand, and single LVs and their possible PVs on the other, to obtain clues as to how the meanings change in different environment. Once the content of the construction is unraveled, productivity is explained (new verbs form based on the structure of the construction).

Focusing on individual LVs, we can more easily isolate groups of LVCs in what we call *islands* (Family 2006). These *islands* are clusters of LVCs which express similar verbal notions based on the same LV, and a specific type of PVs. The islands seem to form based on certain, *but not all*, inherent attributes of the PV. In other words each PV has attributes which activate certain meanings of the LV and the LV in turn contributes relevant features inherent to it, creating a meaning different from the meaning of either component. Each island of LVCs has an underlying construction that encodes this information. The result is an LVC with a meaning that's not necessarily predictable from the meaning of its parts. The LVC will belong to an island with other LVCs that share PVs with similar common attributes and that serve to express similar verbal concepts. Constructions encode semantic as well as syntactic structure.

We will illustrate the patterns that allow for island formation through concrete examples from *xordæn*. *Xordæn* has at least fifteen identifiable islands, each described in detail in the next section.

Xordæn has some islands that express more abstract or emotional notions, or notions that can't be defined solely through physical or perceptual attributes. For example, one of *xordæn*'s islands takes a PV expressing a continuous, irrepressible, negative feeling that has to be suffered as the result of ones personal actions or state of mind. The subject is affected negatively by this feeling.

æfsus xordæn	regret XORDÆN	regret
ænduh xordæn	sorrow XORDÆN	grieve a lost chance/opportunity
æzab xordæn	torture XORDÆN	suffer heavily

(6) *sohrab æz dæst dadæn-e rostæm xejli æfsus xord*
Sohrab from hand give-GEN Rostæm much regret ate
'Sohrab heavily regretted the loss of Rostæm'

While the meaning extensions from the full verb are difficult to define, we can't conclude that *xordæn* must have multiple meanings completely independent of its full verb meaning. For instance, if it actually meant *to suffer*, we would also assign an abundant number of other meanings to account for the data. Further, we would expect the verb *xordæn* to be utilized in every instance of the expression of *suffering*, which is not the case. An example of a context in which suffering is expressed and where *xordæn* does not appear is an island formed with some LVCs of *keʃidæn* "to pull." This island expresses continuous sufferance without necessarily being the result of an action, but rather of injustice.

æzab keʃidæn	torture KEʃIDÆN	grieve, be tortured
rænj keʃidæn	rage KEʃIDÆN	suffer
entezaar keʃidæn	waiting KEʃIDÆN	long for

(7) *æz duri-e dust-æm æzab keʃid-æm*
from distance-GEN friend-1SG torture pulled-1SG
'I suffered from being far from my friend'

Another piece of evidence, supporting the fact that the LVs don't have numerous unrelated and different meanings that surface in each LVC, is that Persian speakers will only utter the core meaning of these words when asked for them out of context. Thus, if asked for the word for *suffer*, *xordæn* would not be uttered, but rather different LVCs probably based on *xordæn* (or possibly *keʃidæn*, as seen above).

It is clear that *xordæn* has a special meaning that only shows up in specific constructions. Meanings embedded in *xordæn* are triggered by certain properties inherent in the PVs with which it combines and the construction in which they occur. Also, as can be seen from the description above, the meaning of the construction is more nuanced than *to suffer*.

Islands are groups of LVCs where one type of PV combines with a particular LV to produce different LVCs with highly related meanings. Each island is assigned a construction that includes a specific LV, a type of PV (defined by common attributes, including physical, perceptual, semantic, and experiential knowledge), and the meaning contributed by the construction. This meaning portrays general aspectual and syntactic information, as well as idiosyncratic semantic information associated with it (not linearly predictable from the meaning of its constituents).

It is important to note that not all the LVCs constructed with a specific LV are valid members of one of the islands of the LV. There are many LVCs with opaque meanings that can be considered idiomatic, as well as some transparent LVCs that don't seem to fall into any of the islands. However, in most cases a motivation for the use of a particular LV in the LVC can be sought, even if isolated in the semantic space.

As a visual aid, diagrams like the one below of the LV *xordæn* depict the semantic space of each LV in this analysis. The periphery of such a diagram represents the islands described above. The proximity of the islands (belonging to the same branch) express closer similarity of notions than islands further away. This configuration is one of several possible configurations, there is no strict metric on this space.

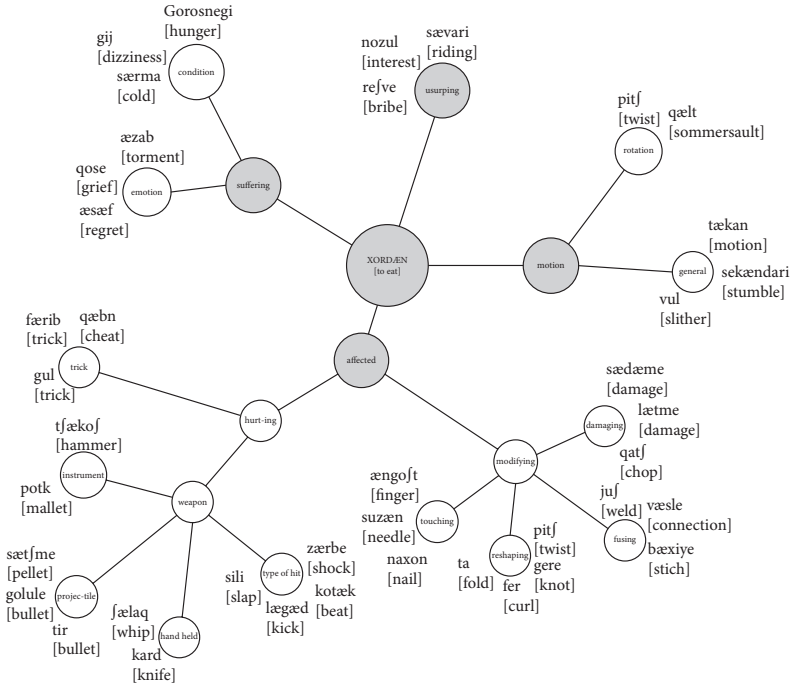

Diagram 1. XORDÆN'S complete semantic space.

Furthermore, it is important to remember that there are no simple verbs in Persian with which to express the ideas expressed by the LVC islands. For example, the only way to express *to fuse* (bottom right in the diagram below) is to combine *xordæn* with the appropriate PV.

5. *XORDÆN'S* islands

In this section, we will present the LVCs and islands of *xordæn* in detail. For each island, we give the general meanings, the type of PV implicated, and elaborate details about the constructions (context and semantic nuances). The LVC islands of *xordæn* fall into four broad categories. These encompass meanings related to being affected, suffering, exploiting, and being agitated. These "meanings" branch off further into finer-grained classifications whose end nodes are the islands.

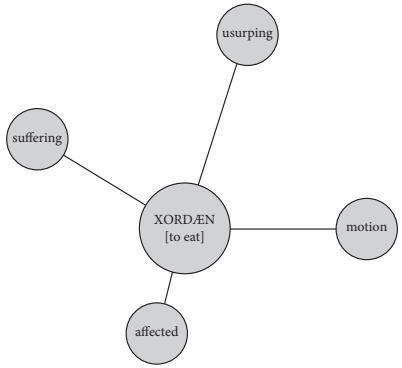

Diagram 2. Major branches in xordæn's semantic space.

5.1 XORDÆN: Affected

The richest branch, measured in terms of the number of associated islands, expresses undergoing or being affected by an action. The subject of these LVCs usually undergoes the action expressed. It is interesting to note that the group of islands that branch off to the right in the diagram are mostly used for inanimate objects, whereas those to the left are animate. Some of these LVCs can be considered as inchoative alternants of analogous LVCs constructed with *zædæn*.[9] As we will see in the next section,

9. Each LV in the system has a similarly structured semantic space (networked islands). Many of the LVCs constructed with the LV *xordæn* have alternants with the LV *zædæn* "hit" (by alternants, I refer to LVCs that have a single PV occuring with different LVs to express similar but

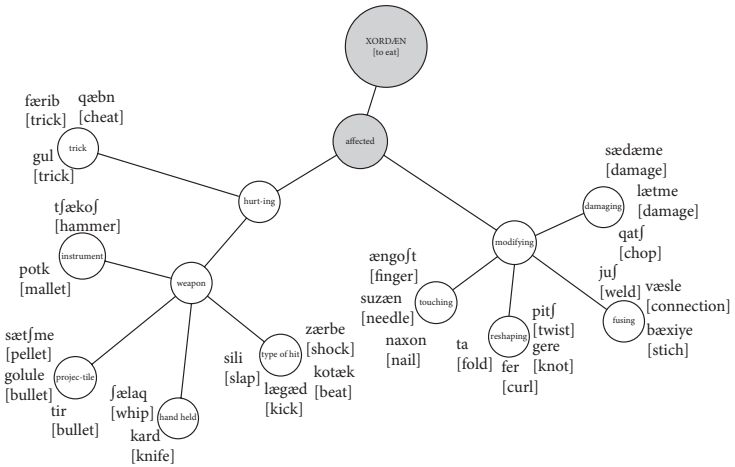

Diagram 3. XORDÆN: Affected.

several *xordæn* islands correlate highly and alternate with a cluster of islands formed with *zædæn*.

5.1.1 *XORDÆN: Affected: Modified: Fused*

Meaning: become fused or connected to parts of itself or to other entities usually through a natural process.

PV: type of connection or instrument/material used for fusing or connecting.

Remarks: Intransitive. These forms are used when the subject becomes fused or mended as a consequence of a natural process (rust, humidity, organic growth) and generally not the consequence of the actions of a conscience being. For example, the term *kuk xordæn* is rare, because stitching can only be done by a volitional external entity. Or, *juʃ* xordæn "weld or fuse" can be used for a material when the fusion is the result of heat or rust or other environmental factors, but not directly when an entity has welded the items together (though if the speaker doesn't know, care, or remember who welded it, but only assumes the action has taken place, this form can be used). In the LVCs expressing the fusion of two different entities, the second entity occurs as an indirect object.

kuk xordæn	stitch XORDÆN	be closed up by stitches
peyvænd xordæn	graft XORDÆN	be grafted (plants, organs)
væsle-pine xordæn	patch XORDÆN	be patched up

related meanings). Alternations are systematic and can be considered as cases of shared islands, connecting two or more different LVs, and in this way inter-connecting all the verbs in the system. It is beyond the scope of this article to explore the dynamics between LV spaces.

(8) *in lebas qæʃæng æst hærtʃænd besyar væsle-pine xord-e æst*
this dress beautiful is despite much patch eat-PTCP is
'This dress is beautiful even though it has been patched up quite a bit'

5.1.2 *XORDÆN: Affected: Modified: Damaged*

Meaning: be damaged or deteriorated.

PV: type of damage or wound.

Remark: Intransitive. These forms express substantial physical damage sustained by the subject. This damage is usually incurred by effects of the physical environment and doesn't necessarily involve an external, conscious agent. The damage usually diminishes the value and usefulness of the subject.

asib xordæn	injury XORDÆN	be injured, be damaged
zæxm xordæn	wound XORDÆN	be damaged, wounded
lætme xordæn	setback XORDÆN	sustain setback (e.g., progress)

(9) *saltænæt pæræst-an dar enγelab lætmehaj-e ziyad xord-ænd*
royalty worshiper-PL in revolution setback-GEN much ate-3SG
'The royalists sustained much setback in the revolution'

5.1.3 *XORDÆN: Affected: Modified: Topology*

Meaning: undergo an organized, topological change.

PV: type of topological transformation.

Remarks: Intransitive. The topological change expressed by these verbs is not imposed or directly inflicted by an external entity, but rather by the environment (natural process) or unintended consequence of an action. For example, one cannot use the term *fer xordæn* to refer to someone's hair after a visit to the hair salon, though it could be used if the curls result from humidity in the air. The change usually damages the subject or at least results in an unwanted state. For example, one cannot say *gere xordæn* for a string that has been purposely tied into a knot, though the same form can be used to express a wire having gotten tangled from too much motion (a consequence of another action, e.g., a tangled telephone wire when one walks around while using the phone).

ta xordæn	fold XORDÆN	get curled
gere xordæn	knot XORDÆN	get tied in a knot
fer xordæn	url XORDÆN	get curled

(10) *hengami ke kenar-e dærya resid-im mu-hay-e*
when that side-GEN sea arrived-2PL hair-PL-GEN
ham-e-man fer xord
all-GEN-2PL curl ate
'When we arrived at the beach, all our hair got curled'

5.1.4 *XORDÆN: Affected: Modified: Surface*

Meaning: be touched with a hand or foot or an instrument, usually leaving a mark or imprint.

PV: instrument doing the touching.

Remarks: Intransitive. The instrument used for this action is usually sharp unless it is a body part (it can leave a non-negligible effect on the surface of the subject), and though the action doesn't entail *hurting*, it might have negative or damaging effects on the subject. This effect is usually not the direct intent of an action, but a consequential result. In other words, someone might touch a surface, not meaning to leave an imprint, and so the imprint is not the intent of the action: the surface can be said to have been *dæst xordæ*.

dæst xordæn	hand XORDÆN	be touched, altered
suzæn xordæn	needle XORDÆN	be touched/pierced with a needle
pa xordæn	foot XORDÆN	get hit with a foot

(11) *in æks æsl nist dast xord-e æst*
this picture original NEG-is hand ate-PTCP is
'This picture is not an original it has been altered'

5.1.5 *XORDÆN: Affected: Hurting: Weapon: Type of Hit*

Meaning: be hit with another entity's hands, feet, or head.

PV: type of hit.

Remarks: Intransitive. This action must directly affect the subject in a hurtful manner. For example, one cannot say **hol xordæn* (push eat), because the notion of *push* can imply acting on an entity without necessarily hurting it. An agent carries out the action, though not explicitly expressed.

sili xordæn	slap XORDÆN	be slapped
lægæd xordæn	kick XORDÆN	get kicked
*moʃ*t xordæn	fist XORDÆN	get punched

(12) *tæræf anyadr xodeʃ-o lus kærd ke yek sili xord*
guy so himself-ACC pest did that one slap ate
'The guy made such a pest of himself that he got slapped in the face'[10]

5.1.6 *XORDÆN: Affected: Hurting: Weapon: Hand Held*

Meaning: be wounded or penetrated by a weapon.

PV: a sharp, penetrating weapon, usually hand held.

Remarks: Intransitive. The weapon must be sharp and directly penetrate the subject. For example, one can be wounded by a gun, but the form **tofæng xordæn* (gun eat) does not exist, because it is not the gun that penetrates, but the bullets (see next island).

xænjær xordæn	saw XORDÆN	be hit with a saw
tʃaqu xordæn	knife XORDÆN	be stabbed with a knife
ʃæmʃir xordæn	sword XORDÆN	be stabbed with a sword

(13) *gozærkon væsæte mahlæke tʃaqu xord.*
passerby middle-GEN melee knife ate.
'The passerby was stabbed in the middle of the melee'

5.1.7 *XORDÆN: Affected: Hurting: Weapon: Projectile*

Meaning: be attacked by a projectile weapon.

PV: a projectile weapon.

Remarks: Intransitive. This island is similar to the previous island, but only differs in the type of weapon used.

muʃæk xordæn	missile XORDÆN	get hit by a missile
tir xordæn	bullet XORDÆN	get shot with a bullet
sætʃme xordæn	pellet XORDÆN	get shot with pellets

(14) *mohafez-e ræis jomhur tir xord*
guard-GEN boss republic bullet ate
'The president's guard got shot'

5.1.8 *XORDÆN: Affected: Hurting: Weapon: Blunt*

Meaning: be struck with a heavy or blunt instrument.

PV: a blunt instrument used for striking.

Remarks: Intransitive. The subject usually undergoes a quick blow or repetitive quick hits by the instrument.

tʃækoʃ xordæn	hammer XORDÆN	be hammered
potk xordæn	mallet XORDÆN	be hit with a mallet
guʃtkub xordæn	meat-hammer XORDÆN	be hit with a meat-hammer

(15) *in maʃin xeyli tʃækoʃ xorde æst*
this automobile much hammer eat-PTCP is
(lit. this automobile has been hammered quite a bit)
'This automobile has been repaired often'

5.1.9 *XORDÆN: Affected: Hurting: Trick*

Meaning: be tricked.

PV: trick.

Remarks: Intransitive. The subject of these forms goes through a negative process. These islands are not productive, but represent an island since there are several forms that express a similar idea.

10. In French slang (argot), one can say "il s'est mangé un pain" (lit. he ate himself some bread) to express someone getting punched (Vanhove, *personal communication*).

hoqqe xordæn	trick XORDÆN	slight, trick
kælæk xordæn	trick XORDÆN	trick
naro xordæn	double-cross XORDÆN	double-cross

(16) *bitʃare hæmiʃe sær-e bazi hoqqe mi-x*or-æd
helpless always head-GEN game trick PROG-eat-3SG
'The poor guy always gets tricked in games'

5.2 XORDÆN: Suffering

The following LVCs express suffering caused by a process or condition affecting a person physically or mentally. The cause of the suffering is usually an unintended result of an action. This is one of the only sets in the system that expresses abstract notions which otherwise mostly occur with the generic LV *kærdæn* "to do". The LVCs in these islands are all atelic, activity verbs: they express durational conditions.

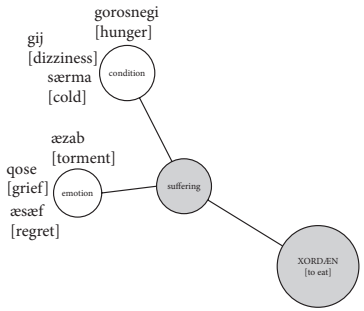

Diagram 4. XORDÆN: Suffering.

5.2.1 *XORDÆN: Suffering: Emotional*

Meaning: suffer from a negative emotion.

PV: emotion of regret, sorrow or grief.

Remarks: Intransitive. The LVCs express the durational suffering from an emotional burden. This emotion is a continuous, irrepressible, negative feeling that has to be suffered as the result of ones personal actions or experience.

nedamæt xordæn	regret XORDÆN	regret
qose xordæn	concern XORDÆN	worry, be concerned
ænduh xordæn	sorrow XORDÆN	grieve

(17) *hæmiʃe qosey-e færda-ra mi-xor-æd*
always concern-GEN tomorrow-ACC PROG-eat-3SG
'She's always worries about the future'

5.2.2 *XORDÆN: Suffering: Physical*

Meaning: suffer from a physical condition that could cause bodily damage.

PV: a natural but uncomfortable condition that causes suffering or might entail more serious ailment.

Remarks: Intransitive. These LVCs specifically express the condition or the process that causes the suffering, and not the symptoms. One cannot say **deldærd xordæn* (stomach ache eat) since this is a symptom (e.g., of hunger) and not a condition that causes suffering. Nor can one say **særgije xordæn* (vertigo eat) which is a symptom of dizziness *gij xordæn*.

gorosnegi xordæn	hunger XORDÆN	suffer from hunger
særma xordæn	cold XORDÆN	catch cold (from the cold)
giji xordæn	dizziness XORDÆN	get dizzy

(18) *u fekr mi-kon-æd æz særma særma xord-e æst*
he think PROG-do-3SG from cold cold ate-PCTP is
'He thinks he has caught a cold from the cold weather'

5.3 XORDÆN: Usurping

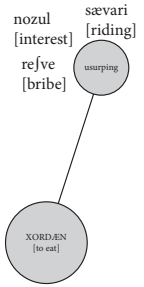

Diagram 5. XORDÆN: Usurping.

Meaning: Exploit service or property.

PV: the type of good that is being taken advantage of.

Remarks: Intransitive. The LVCs in this island express the notion of taking advantage of another person's labor or property. Here, the original meaning of *xordæn*, "eat," emerges in a metaphorical expression denoting gluttony.

reʃve xordæn	bribe XORDÆN	accept a bribe
nozul xordæn	interest XORDÆN	charge interest
pul xordæn	money XORDÆN	embezzle or extract money

(19) *hoquq-e kæm ba'es-e reʃve xord-æn ziyad ʃod-e æst*
salary-GEN small cause-GEN bribe eat-INF much become-PCTP is
'Low salaries have become the cause of much bribery'

5.4 XORDÆN: Agitated

In these LVCs, the subject undergoes certain types of motion. The motion is usually unintentional on the part of the subject and often repetitive.

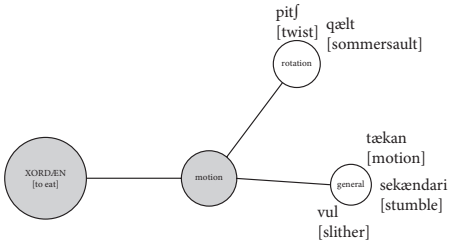

Diagram 6. XORDÆN: Motion.

5.4.1 *XORDÆN: Agitated: General*

Meaning: move.

PV: type of movement.

Remarks: Intransitive. These LVCs express non-goal oriented movement, usually non-volitional. The movement results from an internal, uncontrollable condition or an external agent, such as twitching from muscle spasms (internal) or being shaken by someone to be woken up (external).

tækan xordæn	movement XORDÆN	jerk, shake, wag
telo telo xordæn	sway XORDÆN	sway
vul xordæn	fidget XORDÆN	fidget

(20) *bætʃe æz bihoselegi hæmæʃ vul mi-xord*
child from boredom constantly fidget PROG-eat-3SG
'The kid constantly fidgeted from boredom'

5.4.2 *XORDÆN: Agitated: Rotation*

Meaning: rotate.

PV: type of rotational movement.

Remarks: Intransitive. Similar to the previous island, this set of LVCs express uncontrollable motions, but involve the rotation or turning of the subject.

pitʃ xordæn	roll XORDÆN	be rolled
qælt xordæn	flip XORDÆN	get flipped
mælʲæq xordæn	somersault XORDÆN	go into a somersault, flip over

(21) *maʃin-e mosabeqe seta mælʲæq xord*
car-GEN race three somersault ate
'The race car flipped over three times'

6. Concluding remarks

Each construction has strong restrictions on the lexical items that can occur within it. The idiosyncratic meanings of the constructions emerge from the interaction of the items in the construction with the construction itself. In other words, the construction is stored as a lexical entry, but its existence gives rise to the possibility of making new verbs. It also allows for a more efficient storage of the massive amounts of verbal notions produced by each LV. This type of meaning construction can be considered as a case of *semi-compositionality*, where the meaning of the whole is motivated by the meaning of the parts, but calculated in a more sophisticated way than summation of its components. Each construction encodes semantic information, as shown here, as well as elaborate syntactic constraints, which will be investigated in future studies. The exact contribution of each element is seldom clear, further analysis must be done to find an algorithm for verb formation in this system.

It seems that we cannot observe any basic overarching traits unique to the LV *xordæn*. Only very general properties can be assigned to *the majority* of the islands: proto-patient subject, inchoative. However, these don't differentiate *xordæn* from other LVs in the system that also have these properties, like *gereftæn* "to obtain," or *oftadæn* "to fall." For this reason, a bottom-up approach is crucial for understanding this system. Since no general properties can capture the global behavior of the LV, taking a top-down approach (by assigning general properties to each LV) would not produce significant patterns. Furthermore, individual LVCs have contradicting patterns within each data set of all the LVCs possible with a single LV. A top-down approach would not shed light on the correct contexts a particular LVC would occur in, nor would it allow for precise and unambiguous rules for productivity.

Constructions like those presented in the previous section provide a basis for productivity, they are *semi-productive*. Semi-productivity occurs when a process displays systematic behavior, though the amount of productivity is constrained by semantic or other restrictions. In Persian, the fact that the constructions that define the islands involve particular semantic information signals semi-productivity. Production of new forms is limited by the LV and the restrictions on the PV. Furthermore, there are certain syntactic constraints on the PVs. For example, anaphora or pronominal elements cannot replace the PV in context. This shows that the semantics and syntax of the constructions are to a certain extent constrained, placing restrictions on productivity.

Verb formation in languages such as English or French, follows syntactic rules. In Persian, it is semantically based. This study provides a basis for investigations comparing and analyzing processes of verb formation as they have evolved in different languages, especially its effects on acquisition (comparing the ease and patterns of acquisition).

The result of this analysis has been a fresh insight into several general linguistic issues, such as meaning construction, productivity, and compositionality as they

are manifested in the Persian verbal system. The descriptions in the last section highlighted some of the main semantic components of the islands that populate the semantic space of the LV *xordæn*.

Xordæn, or the notion of "eat" is common to all languages. In Persian, the same lexeme expressing this notion, has evolved to help express many different verbal notions. Some of these seemingly farfetched notions actually reemerge in totally unrelated languages for the same lexeme. This raises an important point for typological studies: languages grouped together with similar syntax might have little common semantic structure, whereas those with similar semantics might diverge in their syntactic structures. The cross-linguistic study of verbs like *xordæn* is necessary to gain a better understanding of the semantic structures common to different languages.

This study also demonstrates that the distinction between the grammar and the lexicon is not very clear, or discrete, and that the space between the two poles is constantly being filled with new constructions. As these include both grammatical and lexical information, they further blur the distinction.

Thus, only deeper and more thorough analysis of these types of phenomena will give us a better chance of understanding the cognitive processing involved. We would expect that inter-language comparative analysis would yield even more insights because it can highlight certain possible universal developments of such phenomena (see Bonvini; Boyeldieu; Hénault this volume). This is not a far-fetched expectation, since we know similar parallel developments in other human faculties across different cultures.

References

Afshar, G., Hakami, N. & Hakami, N. 2002. *Contemporary Persian Dictionary*. Tehran: Farhang Moaser Publishers.

Dabir-Moghaddam, M. 1990. On postposition ra in Persian. *Iranian Journal of Linguistics* 7(1): 2–60.

Dowty, D. 1991. Thematic proto-roles and argument selection. *Language* 67(3): 547–619.

Family, N. 2006. Explorations of Semantic Space: The case of Light Verb Constructions in Persian. Ph.D. dissertation, EHESS, Paris.

Fillmore, C.J. & Kay, P. 1996. Construction Grammar. Ms, University of California at Berkeley, Department of Linguistics.

Ghomeshi, J. 1997. Non-projecting nouns and the ezafe construction in Persian. *Natural Language and Linguistic Theory* 15: 729–788.

Ghomeshi, J. 2003. Plural marking, indefiniteness, and the noun phrase. *Studia Linguistica* 57(2): 47–74

Goldberg, A. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago IL: Chicago University Press.

Goldberg, A. 1996. Words by default: Optimizing constraints and the Persian complex predicate. In *Berkeley Linguistic Society* 22: 132–146.

Goldberg, A. & Jackendoff, R. 2004. The English resultative as a family of constructions. *Language* 80: 532–568.

Haim, S. 1995. *The Larger Persian English Dictionary*. Tehran: Farhang Masser Publishers.

Karimi, S. 1997. Persian complex verbs: Idiomatic or compositional. *Lexicology* 3(2): 273–318.

Karimi-Doostan, M.R. 1997. Light Verb Constructions in Persian. Ph.D. dissertation, University of Essex.

Lambton, A.K.S. 1961. *Persian grammar*. Cambridge: UCP.

Lazard, G. 1970. Etude quantitative de l'évolution d'un morphème: l'Enclitique *ra* en persan. In *Mélanges Marcel Cohen*, 381–388. The Hague: Mouton.

Lazard, G. 1992. *A Grammar of Contemporary Persian*. Cosa Mesa CA: Mazda Publishers.

Mahootian, S. 1997. *Persian*. London: Routledge.

Megerdoomian, K. 2002. Beyond Words and Phrases: A Unified Theory of Predicate Composition. Ph.D. Dissertation, University of Southern California.

Roberts, J. 2005. *Ra* in Persian and information structure. In *Proceedings SIL International*. (http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg.html)

Samiian, V. 1994. The ezafe construction: Some implications for the theory of X-bar syntax. In *Persian Studies in North America*, M. Marashi (Ed.), 17–41. Betheda MD: Iranbooks.

Stevenson, Fazly, A. & North, R. 2004. Statistical measures of the semi-productivity of light verb constructions. In *2nd ACL Workshop on Multiword Expressions: Integrating Processing*, 1–8. (http://www.cs.toronto.edu/~suzanne/publications.html)

Vahedi-Langrudi, M.M. 1996. The Syntax, Semantics and Argument Structure of Complex Predicates in Modern Farsi. Ph.D. dissertation, University of Ottawa.

Wierzbicka, A. 1982. Why can you have a drink when you can't *have an eat? *Language* 58: 753–799.