Lighten Up: The Acquisition of Light Verb Constructions in Persian

Neiloufar Family
Institute for Cognitive Science Studies, Tehran, Iran

Introduction

Unlike languages such as English and French, where morphological processes yield a potentially unlimited number of verbs, Persian has a rather limited inventory of around 120 simple verbs. The vast majority of verbal notions are expressed through light verb constructions (LVCs). Most verbal notions expressed in other languages via simple verbs are expressed in Persian through these constructions (e.g. herf zedeen ‘word HIT’ speak, yad gerefen ‘memory GET’ learn, be donia amadenn ‘to world COME’ be born).

This special kind of productive idiom, often with unpredictable semantics, is at the foundation of the Persian verbal system. In such a language, where only a handful of lexical items are available for expressing thousands of verbal notions, children must develop a highly structured semantic space in order to navigate the ambiguity inherent to the system. This space also serves as a basis for verbal productivity.

The current study presents new data from two Persian-speaking children at ages of 1;11 and 4;1 gathered on a weekly basis over a seven month period. This study is the first to explore the productive mechanism through which children attain mastery of verbal semantics in Persian. We show that children use their knowledge of constructions, which code functional and perceptual information, to produce new verbs.

In section 1, different definitions of the term light verb are reviewed. Section 2 provides an overview of the Persian verbal system. Section 3 presents the data, and section 4 provides a preliminary analysis and discussion. Section 5 highlights potential errors, and section 6 provides a brief conclusion.

1. Light Verb Constructions
1.1. Linguistic Studies

In linguistic studies, the term light verb was first used by Jespersen (1965) to refer to English N VP constructions. These constructions consist of a nominal complement and a semantically weak verb. Examples from English include take a nap, take a drive, give a talk, give a look, have a shower, take a break, get exercise.

A light verb (LV) often contributes little or nothing to the construction that it syntactically heads. For example, English light verbs (LVs) include take and give. When used in LVCs such as take a seat, and give a talk, the meanings of the verbal elements seem to be reduced. The amount or type of information portrayed by the LV in these constructions is subject to much debate in linguistics.

1.2. Acquisition Studies

On the other hand, in past acquisition studies, the term light verb has been employed to describe a different type of linguistic entity. In these studies, light verbs are a set of semantically general verbs in a given language that have a special semantic status. These verbs (e.g. go, make, take) allegedly provide a route into adult-like language.

According to Clark (1978), children learn semantically general verbs because they express semantically general meanings. Pinker (1989) claims that these general-purpose, or light, verbs are mapped directly onto innately specified semantic primitive elements. These verbs therefore represent the simplest combination of these primitive elements and thus the basic cognitive configuration for the child.

Ninio (1999), similar to Goldberg (1998), suggests that the earliest verbs learned by a child best encode the first or default syntactic frames or constructions in which they occur. In other words, certain verbs provide the abstract syntactic schemas necessary to acquire other verbs.

Though it is widely assumed that general-purpose verbs, termed light verbs, are pivotal in language acquisition, there is no consensus on which particular verbs these are. Also, if generality explains their early acquisition, semantics must contribute to acquisition regardless of frequency. However, these same verbs tend to be highly frequent in child and child-directed speech. We will not discuss this debate further, but refer the reader to Theakston, et al. (2004) for a comprehensive review.

In this study, we focus on the type of constructions described in section 1.1, namely, a multiword construction comprised of a nonverb element combined with a light verb. The resulting construction produces a non-compositional whole. Although the light verbs in Persian tend to be semantically general and correspond to the general-purpose verbs discussed in acquisition studies, they are only considered light verbs when they occur in light verb constructions (occurring with a non-verbal complement). In other words, take in take a nap is a light verb, but when it occurs in isolation it is considered a full or heavy verb (e.g. Take that book off the table).

2. The Persian Verbal System
2.1. Verbal Structure

There are only around 120 full verbs in Persian. At first glance, Persian seems quite limited in verbal expressions. However, this deceptively small number of full verbs belies a complex system of verbal constructs. From the limited number of full verbs in Persian, several serve as LVs\(^1\), forming the basis of a great variety of verbal notions. Persian LVCs have been a challenge for syntactic, phonological, and semantic theories (see Sharifi, 1975; Vahedi-Langrudi, 1996; Karimi-Doostan, 1997; Dabir-Moghadam, 1997; Megerdoomian, 2002; Follé et al. 2004).

These multi-word expressions consist of one of a small set of LVs combined with a preverbal element (PV) that can be nominal, adjectival, or prepositional. Each LV in Persian combines with a large variety of PVs to form different LVCs producing diverse meanings. The meaning of the resulting construction often deviates from the meaning of the sum of its constituents.

- piano zedan (piano HIT)
- gandelier (plaster OBTAIN)
- benzin zedan (gasoline HIT)
- jaf / ferdan (party OBTAIN)
- xosh zedan (dry HIT)
- xun endaxtan (blood THROW)
- play piano
- set in plaster
- pump with gas
- celebrate
- freeze in fear
- cause to bleed

Constructions similar to the above exist in many languages (consider English take a walk, give a talk, make a speech), but these often have full verb counterparts (e.g. walk, talk, speak). In Persian, LVCs are usually the only option for expressing a verbal notion. This makes it difficult to draw the line between lexical items and grammatical processes.

LVCs in Persian range from the transparent (e.g. piano zedan piano HIT to play the piano), where the contribution of the two constituent elements is relatively clear, to the idiomatic or lexical (e.g. xox zedan dry HIT to be shocked), where the meaning of the whole is seemingly unrelated to the meaning of the parts. Most LVCs occur between these two extremes. In other words, the meaning of the whole is usually motivated by the meaning of its parts, but the exact contribution of each component is unclear.

Previous findings suggest that the first verbs children acquire are general or “light” verbs (Clark, 78). In Persian, LVs are general but they later serve as a basis for hundreds of different, more sophisticated verbal notions. That is,
unlike English, LVs will not be eventually replaced by other verbs. Rather, the same verbs occur in semantically more sophisticated, or less compositional, constructions.

2.2. Productivity

Unlike English, where nouns can often be used as verbs, as in ‘I will phone you,’ (the noun ‘phone’ appears as a verb to mean ‘contact by phone’), Persian speakers must always produce verbs by combining a non-verbal element with an LV. Consider the following examples, all of which are recently created LVCs based on nouns introduced into Persian from other languages:

email zedān email HIT send an email
lotion zedān lotion HIT apply a lotion
hepatitis OBTAINT hepatitis type DO catch hepatitis
type (keyboard)
gaffe GIVE reveal by accident
mesaj dadaen message GIVE leave a message
batmitzva OBTAINT batmitzva GIVE celebrate a batmitzva
ski DODSK

Nouns appear in the language upon the introduction of a new idea, material, object, or method. These are then used as PVs in novel LVCs occurring with a pertinent LV. The types of examples above provide evidence for LVC productivity in Persian.

Initially, adult Persian learners find most Persian LVCs idiosyncratic. As we will see, however, these forms are at least partially predictable. Most LVCs systematically correspond to the semantic classes of PVs, and the level of productivity remains constrained by idiosyncratic restrictions.

2.3. Verbal Islands

Family’s analysis of Persian LVCs (2006) provides a blueprint for novel LVCs. An analytical tool is presented for mapping semantic regularities in the semantic spaces of each LV in Persian. It postulates that an LV’s semantic space is populated by notional islands. Each island has an associated construction with its own meaning, which specifies a particular LV and a constrained class of PVs. The class of PVS is defined by its functional, social, chemical, visual, tactile, and other common attributes. The islands seem to form based on certain, but not all, inherent attributes of the PV.

Variation of the PV 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. Detailed investigation shows that sets of LVCs actually share certain attributes. For example, the verb ketidān whose heavy verb meaning is to pull, serves as an LV in a variety of different constructions.

dad ketidān shout PULL
qālān ketidān hooka PULL
feraqat ketidān separation PULL
ælam ketidān banner PULL
æbru ketidān eyebrow PULL
shout smoke a hooka
suffer from longing
hoist a banner
draw in eyebrows

After studying the hundreds of LVCs produced with ketidān, patterns emerge where some of the different meanings seem to be highly related based on the properties of the PV. For example, if the PV is any smokable substance or an instrument that can be smoked out of, the resulting meaning will mean to smoke that substance or out of that instrument, respectively. Recalling that there is no lexeme with the meaning to smoke in Persian, these constructions are the only option with which one can express the notion of smoking.

hajīf ketidān hash PULL
sigār ketidān cigarette PULL
smoke hash
qālān ketidān hooka PULL
smoke cigarettes
pīp ketidān pipe PULL
smoke a hooka
smoke a pipe

baradārām ziad sigar mikeshād.
Brother-1S much cigarette PROG-pull.PRS-35.
My brother smokes a lot.

On the other hand, if the PV is an object that is usually set up or constructed along a path or perimeter, the resulting meaning will be to draw across a length.

jād ketidān road PULL
divar ketidān wall PULL
make a road
nārde ketidān fence PULL
build a wall
put up a fence

They’ve fenced their yard.

The semantic space of ketidān has about twenty islands, each with a specified class of PV. These islands form a network of different meanings allowing some predictability and some basis for productivity.

2 Abbreviations of morphemes glossed examples are as follows: PROG=progressive, NEG=negative, PRS=present, PST=past, IMP=imperative, SUBJ=subjunctive.
From these islands, other LVCs with similar meanings can be constructed. Each island’s underlying construction serves as a basis for productivity, providing rules for making new forms. Real-world knowledge about the functional attributes of the PV in these constructions plays a crucial role in judging the acceptability of the form. These constructions preempt positing highly idiosyncratic secondary meanings for LVs. Constructions and the islands they represent provide a blueprint for new verbal notions and include rich semantic and syntactic information.

Most LVs in Persian have around a couple dozen islands in their respective semantic spaces. These islands all have idiosyncratic specifications as the ones described above.

2.4. Acquisition

Persian has a special kind of productive idiom at the foundation of its verbal system. The acquisition of the system elicits a specific processing strategy that allows children to learn these constructions. The steps and difficulties of acquisition of these constructions have not been addressed. The goal of this study is to provide a first look at these constructions in child language using spontaneous samples from native Persian children.

3. Data

In order to develop a corpus of spontaneous child language, two native mono-lingual Persian speaking children were recorded. To ensure that the children were not exposed to other languages, these recordings where collected in Tehran, Iran.

One previous study by Samadi (1996) involved data collected from immigrant children in Sheffield, UK. English words are used by these children and it is not clear to what extent their English speaking environment interacts with their acquisition of Persian. These corpora are also available on the CHILDES database (McWhinney, 2000), but they are not tagged or annotated.

For the current corpus collection, recorders were provided and parents were asked to record their conversations with their children on a regular (approximately bi-weekly) basis. There are a total of 15 hours for the child Minu, and 10 hours for Lilia over a seven month period. These recordings were transcribed, synchronized to the transcription, and annotated in CHAT format.

3.1. Minu

Minu was 4;1 at the beginning of the recordings. She is an only child, living with her parents. Both parents are graduate students in philosophy. The mother was with Minu for most of the day, and the father was present during the afternoons and evenings. Minu is a relatively talkative child. The majority of recordings with Minu involve conversation between her and one or both of her parents. In these recording she is typically involved in some play activity or storytelling.

3.2. Lilia

Lilia was 1;11 at the start of the data collection. She has an older brother (5;11 at the inception of recordings) who is present in most of the recordings. Lilia’s main caretakers are her nanny and her mother who actively engage in conversation and activities with Lilia. Lilia’s parents are both professionals, but her mother is on maternity leave. The recordings are mostly made in the evenings or at dinner-time and most often only involve the mother and the brother.

4. Frequency Counts

Since no previous empirical data of spoken language have been available to study the Persian verbal system, we began by investigating the frequency of LVCs and verbs in general. For the purposes of this study, the data is grouped according to the age of the child in months. Each month represents between one and four hours of recording (a variable factor found not to affect frequency). Since no adult data exists to date, these values cannot be compared to adult speech. Also, the seventh month of recordings of Lilia have not been fully transcribed at the time of this article.

4.1. Ratio of Simple Verbs and LVCs

LVCs make up the majority of Persian verbal notions. We expected children to progressively use more LVCs as their verbal vocabulary developed. However, this wasn’t the case for either Minu or Lilia. As can be seen in Table 1, the ratio of simple verbs to LVCs fluctuates between 1.8 and 3 for Minu and 2 and 4.8 for Lilia. This ratio does not decrease over time. Furthermore, simple verbs are always more frequent than LVCs: almost always more than twice as frequent as LVCs. The same fluctuation is observed in child-directed speech (CDS) for both children.

<table>
<thead>
<tr>
<th>Table 1: Ratio in MINU’s Speech</th>
<th>Table 2: Ratio in LILIA’s Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGE</td>
<td>CHI</td>
</tr>
<tr>
<td>4;1</td>
<td>1.82</td>
</tr>
<tr>
<td>4;2</td>
<td>1.88</td>
</tr>
<tr>
<td>4;3</td>
<td>3.09</td>
</tr>
<tr>
<td>4;4</td>
<td>1.98</td>
</tr>
<tr>
<td>4;5</td>
<td>2.37</td>
</tr>
<tr>
<td>4;6</td>
<td>2.98</td>
</tr>
<tr>
<td>4;7</td>
<td>2.19</td>
</tr>
</tbody>
</table>

avg: 2.33 2.12
4.2. Most Frequent LVs

In Minu's speech, the LVs that have the highest tokens are karden 'to do' (avg: 45%), joden 'to become' (avg: 25%) and dalten 'to have' (avg: 10%). The rest of the LVs occur less than 10% of the time in all the recordings.

The only difference observed in Lilia's speech is that there are sessions during which dalten occurs more frequently than joden. This could be because of the high frequency of the verb dust dalten (friend HAVE 'to like/love') in Lilia's speech. It is important to collect adult speech samples to compare frequencies.

4.3. Most Frequent Simple Verbs

A final observation in frequency counts is that some of the most frequent simple verbs in children's speech do not correspond to the LVs in the language. The most frequent simple verbs in both children's speech in order of frequency are goftan 'to say', xastan 'to want', and tevanestan 'to be able'. This shows that the LVs in Persian are not necessarily the most frequent verbs in a child's language. There must be another explanation for their emergence as LVs.

5. Errors

Certain types of errors point to productivity in the children's use of LVCs. Productivity in Persian is semantically based, so we expected semantic overgeneralizations to emerge.

One type of error is when the child replaces an LV in an LVC with the most frequent and general LV karden 'to do'. In the example below, Minu uses the erroneous LVC qelt karden 'somersault DO' instead of the correct form qelt zedan 'somersault HIT' to express doing a somersault.

nobæti *qelt mikonim. (MINU, 4:1.19)
turn-taking somersault PROG-do.PRS-1P.
[target: Let's take turns doing somersaults.]

The semantic basis for this error could be overgeneralizing the unergative use of the verb karden 'to do'. However, the error is more likely due to the high frequency and general meaning of karden. Also, its full verb meaning doesn't occur as frequently, while the full version of zedan 'to hit' does and might cause semantic interference.

Another type of observed error more directly supports the island analysis described above. In these cases, the child invents an LVC to express a novel verbal notion. In the example below, Minu is describing her day at the playground and describes unsuccessful attempts to go up the slide. She says she kept slipping because her feet were slippery. She makes up the LVC sor dalten 'slipperiness HAVE'.

pam xeili sor dasht. (MINU, 4:5)
foot-1S much slipperiness have.PST-3S.
[target: My feet slipped a lot.]

According to our analysis, Minu uses her knowledge of an island in the semantic space of the verb dalten productively. When dalten occurs with a PV that expresses an intrinsic characteristic, the resulting meaning of the construction is to have that intrinsic characteristic. Some examples of this island include:

tæhæmol dalten tolerance HAVE be tolerant
næm dalten moisture HAVE be moist
tab dalten warp HAVE be warped

In lebas hænzæ naæm dare.
This clothing still moisture have.PRS-3S.
This clothing is still wet (or moist).

Minu hasn't fine-tuned her knowledge of this island and combines the noun, sor 'slipperiness' with dalten to express being slippery, while slipperiness is not an intrinsic property but a surface property, thus unacceptable as a PV for dalten. The correct way to express the meaning Minu intended would be with the verb to be (e.g. is slippery).

Another example is when Lilia uses her knowledge of an island in the semantic space of aamedan 'to come' to produce a new LVC. In the example below, Lilia picks up the telephone and hears a dial tone. She uses the PV buq 'beep' in this construction to express that she hears a beep.

buq miad . (Lilia, 2;1)
beep PROG-come.PRS.3S .
[target: I hear a beep.]

This island only takes PVs expressing basic non-tactile stimuli with no specified source or specific quality (e.g. smell, sound, light) to express perceiving that stimulus.

bu aamedan smell COME smell
seda aamedan sound COME sound
nur aamedan light COME perceive light shining
Tu un otaq bu miad.
In that room smell PROG.come.PRS.3S.
It smells in that room.

According to the above analysis, Lilia hasn’t understood the semantic restrictions on this island, namely, that the PV may not be a specific type of sound, but the description of the basic stimulus.

A third example of an error occurs when Minu is tickling her mother. While she previously has used the correct verb for tickle (i.e. gelqelak amana ‘tickle COME’), she overgeneralizes her knowledge of a semantic island in gereftan ‘to get’ and invents a new LVC.

maman, injat hæm gelqelak migire? (Minu, 4;7)
mommy, here-2S also tickle PROG-get.PRS-3S
[target: Mommy, are you ticklish here, too?]

One explanation may be her confusion about the semantic island in the space of gereftan which expresses the sudden seizure by a psychological or biological state or pressure. The subject of these forms experiences a state or condition that seize or overwhelsms. PVs include og gag, sorfe cough, $al urine. The resulting constructions express the anticipation of releasing the urge expressed by the PV. In her error, Minu might be considering the urge to laugh upon the act of tickling. She seems to collapse the action and effect into one verbal concept. In fact laugh, kænde ‘laughter’, is in the same island as gag and cough.

Another explanation could be that she is extending another similar construction with gereftan, which expresses the seizure of pain (e.g. dærd gereftan ‘pain GET’). The PV dærd ‘pain’ can also occur with amana ‘to come’ to express a similar meaning. As mentioned above, the PV gelqelak ‘tickle’ occurs with amana. Since the PV dærd occurs with both gereftan and amana, Minu could be making the analogy producing this erroneous alternation (i.e. dærd amana- dærd gereftan, gelqelak amana- *gelqelak gereftan).

In another example, Minu overgeneralizes the alternation of kærend ‘to do’ and jodan ‘to become’. These LVCs occur with the same PV, producing the causative alternation, where the LVC with jodan expresses the inchoative counterpart of the causative kærend. Consider the following dialogue between Minu and her aunt (they are drawing and sticking stickers on paper).

Minu: in bayaæ kændæ beshe, næ?
This must ripped-off SUBJ-become.PRS-3S, no?
This has to be ripped off, no?
Minu: kændæsh konæm [*]?
Ripped-off-3S do.PRS-1S?
Should I rip it off?

Aunt: kændæsh bokon [*],
Ripped-off-3S IMP-do.PRS.2S.
Rip it off.
Minu: bekænemesh.
SUBJ-rip-off.PRS-1S-3S.
I’ll rip it off.

Here, Minu is using her knowledge that one of kærend’s islands (i.e. the causative) alternates with one of jodan’s (i.e. the inchoative). Since she uses the correct form of the verb kænde jodan ‘ripped-off become’ where kænde is the participial adjective form of rip off. Then, immediately after, she erroneously uses the form kænde kærend according to the island alternation. In this case, however, the actual simple verb kændæ ‘to rip off’ should preempt the production of the LVC. Her aunt finds her utterance cute, and repeats it, and only after does Minu herself reflect, become aware that she has made a mistake, and then produce the correct form. Her production of the erroneous form points to the productive property of semantic spaces, which includes relevant characteristics of islands.

The above types of errors are not common. In our data, we had a total of 15 errors for Minu and 4 for Lilia. Given the amount of data for each child, it would mean that Minu made this type of error once per hour, and Lilia once every two hours. Minu’s higher creativity in LVC production suggests that the abstraction and creation of nuanced semantic classes (e.g. of PVs), emerges at a later stage in development.

6. Conclusion

In this study, we utilized newly acquired data from Persian speaking children in order to investigate the frequency of and semantic errors in LVCs. We found that the frequency of LVCs fluctuates in children’s speech, and in child directed speech. In other words, LVCs do not progressively take over a child’s verbal vocabulary as expected from linguistic hypotheses. The LVC kærend ‘do’ and jodan ‘become’ make up the majority of tokens in LVC production.

From error analysis, it is clear that children use their knowledge of islands of productivity in Persian verbal space. They fine-tune their semantic knowledge of semantic constraints of LVCs as well as the idiosyncratic semi-productive properties of the islands they belong to. This productive capacity seems to be more prevalent in the older child. We conclude that children progressively learn about idiosyncratic semantic class distinctions and use this knowledge creatively.
References


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