Investigating lexical bias in Persian light verb constructions: What can we learn from priming experiments?

Pegah Faghiri, Monique Flecken and Eva van Lier (University of Amsterdam)

The verbal lexicon in Persian is overwhelmingly formed by multiword expressions (MWE) including a verb and a non-verbal element, mainly a noun, such as *bāzi kardan* 'play' (game do) or *qadam zadan* 'walk' (step hit), known as complex predicates (CPs) or light verb constructions (LVCs). There is an ongoing theoretical debate about the nature of these LVCs: The compositional view, formulated in various studies in generative syntax, is mainstream. In contrast, numerous studies, taking a lexicalist approach, have stressed the idiomatic nature and word-like properties of (a part of) Persian LVCs (see Samvelian 2018 for an overview). While some recent contributions present experimental evidence (e.g. Faghiri & Samvelian 2021), the type of data involved remains fundamentally the same: acceptability ratings vs. grammatical judgments. In particular, a language processing perspective is largely lacking (Shabani-Jadidi 2012 is a notable exception). In this study, we use a (structural) priming paradigm to investigate to what extent Persian LVCs display lexicalized behaviour. The alternation we are focusing on concerns the choice of LV involved in forming the LVC, given that a part of Persian LVCs display the possibility to alternate between two or more verbs, without affecting the meaning of the combination (see e.g. Faghiri & Samvelian 2013).

Our priming study builds on related research on the well-known dative alternation in Germanic languages. Corpus studies reveal that alternating ditransitive verbs show a statistical preference for (or occur more often in) one of the two alternating constructions. Importantly, such lexical preferences or biases manifest themselves in priming production experiments (e.g. Melinger & Dobel 2005, Bernolet & Hartsuiker 2010, Segaert et al. 2014). Specifically, presenting a verb in its dispreferred construction is shown to exert a stronger priming effect compared to that verb's preferred construction; this is called "inverse priming". However, the overwhelming majority of studies on lexical bias in priming concerns Germanic ditransitives. Therefore, our study not only offers a new, processing-based perspective on Persian LVCs, but also broadens the scope of verb-bias studies in terms of the languages and constructions covered.

As mentioned, a number of verbs can participate in forming Persian LVCs. *Kardan* 'do' is the most frequent LV and is used in many newly coined LVCs, e.g. *imeyl kardan* 'email', *čat kardan* 'chat', *fālo kardan* 'follow (in social networks)', *še(y)r kardan* 'share (in social networks)'. But there are other productive LVs as well, such as *zadan* 'hit', which is also used in newly coined LVCs, e.g. *payāmak zadan* 'send a text (on a mobile phone)', *lāyk zadan* 'like (a post on social networks)' (see Samvelian 2012). Some of these can occur with both LVs, e.g. *lāyk kardan/zadan*, *imeyl kardan/zadan*, whereas others cannot, e.g. *fālo kardan/*zadan*. As such, the LV alternation in Persian LVCs is lexically restricted, like the dative alternation.

As a pretest for our priming study, we first collected baseline data for a set of 52 LVCs potentially displaying *kardan/zadan* ('do'/'hit') alternation. We were able to use picture stimuli for 24 of these and hence included them in a constrained picture-description experiment (Fig. 1). For the remaining 28, we ran a Cloze task experiment (Fig. 2). In both experiments, participants completed sentences with a verb, i.e. the verbal element of the LVC, while the nominal element was specified – either in the picture (Exp. 1) or as part of the sentence (Exp. 2). Both experiments included 3 practice items at the beginning and 2 filler items between critical items. They were carried out online via PCIbex Farm. We collected a total number 2085 datapoints from respectively 71 (Exp. 1) and 56 (Exp. 2) native speakers of (Iranian) Persian. We excluded 15 LVCs which exclusively occurred with 'hit' or 'do' and for which we did not have corpus evidence for the possibility to occur with the other LV, as well as 7 for which the rate of 'do' and/or 'hit' were lower than the rate of a third option, and classified the 34 remaining into three "bias" types based on the rate of 'do'/'hit' in the sentences produced:

1. Do-bias LVCs: LVCs with more than two-third (66.7%) of do- responses

2. Hit-bias LVCs: LVCs with less than one-third (33.4%) of do- responses
3. No-bias LVCs: other/in-between LVCs

11 LVCs were labelled as having a do-bias (e.g. *atse* 'sneeze', *mangane* 'staple', *imeyl* 'email', *emzā* 'sign'), 14 as having a hit-bias (e.g. *telefon* 'phone', *baxiye* 'stich', *susu* 'twinkle', *taxmin* 'estimate') and 9 as having no-bias (e.g. *rang* 'paint', *vasle* 'patch', *mohr* 'stamp', *jamʔ* 'sum').

Thus, not only do our pretest data confirm LV alternations but they also show that LVCs can display lexical biases and that the choice of the LV is not necessarily motivated by syntactic and/or semantic constraints. If Persian LVCs display different lexical preferences and these preferences are indeed part of speakers knowledge, then we would expect to see priming effects similar to what has been observed for dative alternations. In particular, we expected a difference in the strength of priming effects, depending on the prime LVC bias type (i.e. do-bias, hit-bias and no-bias) and the priming condition: prime LVCs with a lexical bias were predicted to exert a stronger priming effect when used with their dispreferred LV (i.e. the inverse priming effect); Also, we expected target LVCs to show resistant to priming depending on the strength of their lexical bias, in particular do-/hit- bias target LVC types should be in general more resistant to priming than no-bias target LVCs. We expected to see these difference in the distribution of LVs produced by the participants and/or in their response latencies (see e.g. Segaert et al. 2014).

To test these predictions, we carried out a priming experiment via PCIbex Farm using a task similar to Ziegler et al. (2018). In each trial participants read a sentence (prime item) and then saw a picture (target item) that they had to describe (by typing) in one sentence using the given noun. They also answered a recognition memory question (cover task) after each item. We included three priming conditions (baseline do-prime, hit-prime) and used a 3x3x3 mixed design with priming condition and prime LVC bias type as a within-items factor and target LVC bias type as a between-items factor – across 9 experimental lists. For example:

1. Do-prime: *mesvāk kardan* 'Ali brush-**did** his teeth after dinner.' (simplified)
2. Hit-prime: *mesvāk zadan* 'Ali brush-**hit** his teeth after dinner.'
3. Baseline: 'Ali had a toothache after dinner'

Accordingly, with these combinations we get three different priming types depending on the prime LV expectedness: "expected" (when the prime bias type and the prime LV match, e.g. a do-bias LVC used with 'do'), "unexpected" (when the prime bias type and the prime LV mismatch, e.g. a do-bias LVC used with 'hit'), "neutral" (for all no-bias primes). Our stimuli consisted of 12 target pictures corresponding to 4 no-bias, 4 hit-bias and 4 do-bias LVCs from Experiment 1. Each target was paired with three different prime sentences, one of each type. In total 17 different prime LVCs were used. Each list contained all the 12 target items and across all lists each target item occurred an equal number of times with each prime of the three prime types as well as with each of the three priming conditions. Each list contained a different subset of 12 prime items appearing only once in one of the three conditions. In addition, each list included a set of 39 filler trials, which sums to 51 total trials.

129 (Iranian) Persian speakers participated in the experiment, of which results from 4 were excluded. We used the target sentence completion time span between two consecutive items as a measure of response latency (RL) for each item. We normalized RLs for participants across all the items only excluding the first one. We coded participants responses (1500 in total) according to the verb used: *kardan* (776), *zadan* (610), other (96) and incomplete (18), and filtered out irrelevant responses (114=7.6% of the data). For each target LVC, we calculated its baseline bias as the log-odds for *kardan* 'do' responses in the baseline condition (see Bernolet & Hartsuiker, 2010:457). We analysed the data using logistic mixed models for LV (with *kardan* = 1 and *zadan* = 0) and linear mixed models for RL.

With respect to the choice of LV, the results show a clear main effect of target bias, but we do not see any priming effects including the expected inverse priming effect: prime sentences with 'do' or 'hit' LVs trigger more 'do' responses (which is the overall more frequent choice) than the baseline for target LVCs that do not (already) have a bias for 'do', regardless of the LV used in the prime (Fig. 3) or the prime bias type (Fig. 4). RLs, on the other hand, are sensitive to both priming conditions (Fig. 5) and prime bias types (Fig. 6). While in the baseline condition, RLs are comparable for *kardan* 'do' and *zadan* 'hit' responses i) with 'do' primes RLs are lower with 'do' responses and with 'hit' primes RLs are lower with 'hit' responses and ii) with do-bias primes RLs are lower with 'do' responses, whereas with hit-bias primes RLs are lower with 'hit' responses. The results of our model fitted to a subset of the data excluding the no-bias prime type and the baseline condition (641 datapoints) showed i) a 3-way significant interaction between LV choice, priming condition and priming type and ii) a 4-way significant interaction including target bias. These results show that response latencies are differentially affected by the priming condition and the prime expectedness type: when participants choose 'do' with hit-bias targets they are significantly faster when primed with an unexpected 'hit' than when primed with an expected 'hit' (while with 'do' primes, RLs are comparable). Similarly, when participants choose 'hit' with hit-bias targets they are much faster when primed with an unexpected 'do' than when primed with an expected 'do' (while with 'hit' primes, RLs are comparable).

Our study confirms that some Persian LVCs show LV alternations that cannot be explained based on syntactic and/or semantic rules and that these LVCs display lexical biases in terms of which verb they preferably appear with. Moreover, the results of our priming experiment suggest that speakers have knowledge of these biases, which, when primed, manifests not only in their usage, i.e. significant difference in the rate of 'do' depending on the target LVC's bias, but also in their response latencies, which are clearly affected by inverse priming.

References

Bernolet, S. & R. J. Hartsuiker. 2010. Does verb bias modulate syntactic priming? *Cognition 114*: 455–461.

Faghiri, P. & P. Samvelian. 2021. The issue of "separability" in Persian complex predicates. In *One-to-many relations in morphology, syntax, and semantics,* pp.117–149. Language Science Press.

Melinger, A. & Ch. Dobel. 2005. Lexically driven syntactic priming. *Cognition 98*: B1–B20.

Samvelian, P. 2012. *Grammaire des prédicats complexes. Les constructions nom-verbe*. Paris: Lavoisier.

Samvelian, P. 2018. Specific features of Persian syntax. The Ezāfe Construction, Differential Object Marking and Complex Predicates. In A. Sedighi & P. Shabani-Jadid (eds.), The Oxford handbook of Persian linguistics, pp 226–269.

Samvelian, P. & P. Faghiri. 2013. Introducing PersPred, a syntactic and semantic database for Persian Complex Predicates, In *Proceedings of The 9th Workshop on Multiword Expressions (MWE2013),* pp 11-20.

Segaert, K., Weber, K., Cladder-Micus, M. & Hagoort, P., 2014. The influence of verb-bound syntactic preferences on the processing of syntactic structures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(5): 1448.

Shabani-Jadidi, P., 2012. Processing Compound Verbs in Persian. Ph.D. thesis, University of Ottawa.

Ziegler, J., Snedeker, J., & Wittenberg, E. 2018. Event structures drive semantic structural priming, not thematic roles: Evidence from idioms and light verbs. *Cognitive Science 42*(8): 2918-2949.

Fig. 1 Screenshot of an Item from Experiment 1 (English translations are added for illustration)

Fig. 2 Screenshot of an item from Experiment 2 (English translations are added for illustration)

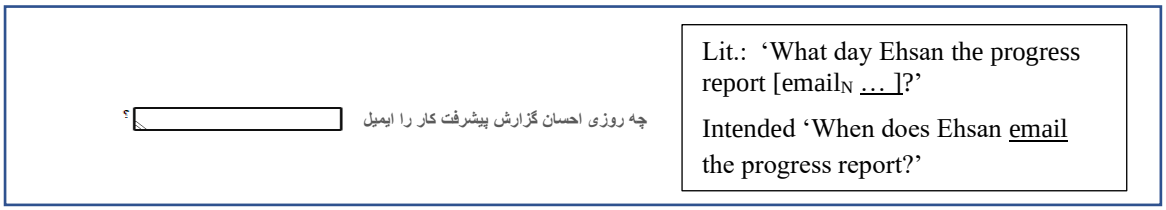

Fig. 3 Mean rate of 'do' by target bias type and prime conditions (Baseline, Do and Hit)

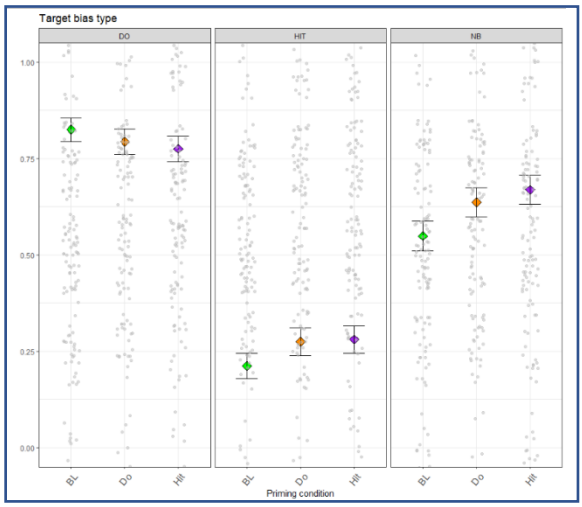

Fig. 4 Mean rate of 'do' by prime bias type and prime conditions (Baseline, Do and Hit)

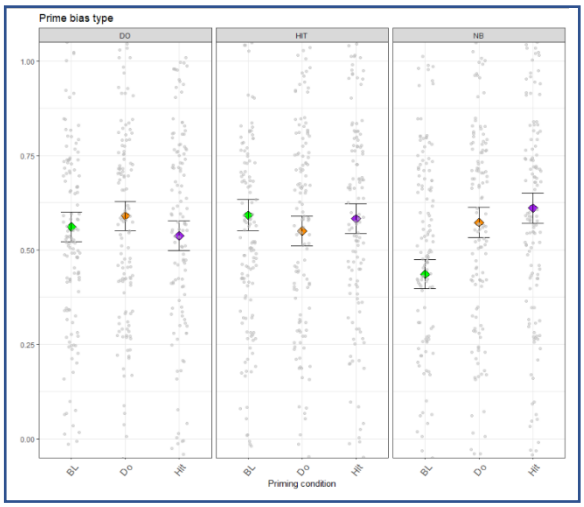

Fig. 5 Mean RLs by prime condition and response (kardan 'do' vs. zadan 'hit')

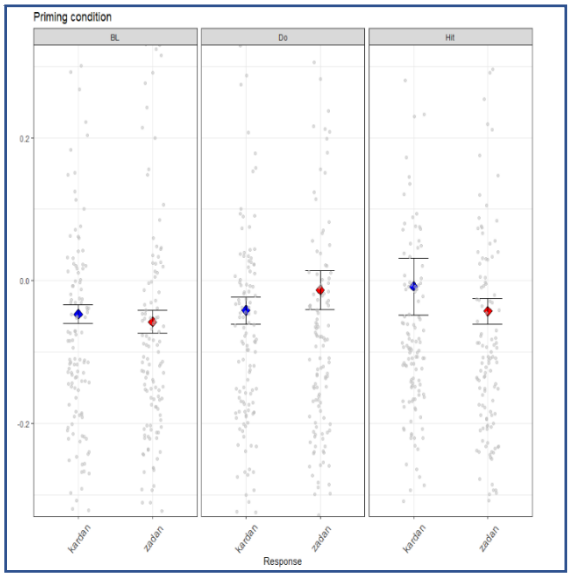

Fig. 6 Mean RLs by prime bias type and response (kardan 'do' vs. zadan 'hit')