

### **Guide:**

This corpus is a synchronic monolingual linguistic corpus of Persian Light Verb Constructions. It represents Light Verb Constructions (LVCs) consisting of 21 productive Persian LVs i.e. da:shtan (to have), kardan (to do), shodan (to become), gashtan (to turn), goza:shtan (to put), keshidan (to pull), didan (to see), da:dan (to give), bakhshidan (to give), to grant), gereftan (to get,) yaftan (to obtain), ?a:madan (to come), ?a:vardan (to bring), residan (to arrive), raftan (to go), ?ofta:dan (to fall), ?anda:khtan (to throw, bordan (to take), khordan (to collide), zadan (to hit), bastan (to tie).

The annotation has been done at word level. In line with Distributed Morphology (Halle & Marantz, 1993 and Marantz, 2013), preverbal elements are considered as categoryless elements (annotated as PVs) and light verbs as categorizers (annotated as LVs). In addition, the present and past lemmas of each LVC are given. SEP /INSEP is used as another tag for the cases that there is a gap between preverbal element and light verb. It should be noted that in line with Karimi-Doostan (2011) the cases in which preverbal elements and light verbs are broken by Negation Particle (neg), the imperfective morpheme (mi), modals and auxiliaries such as ba:yad (should, must), future auxiliary xa:stan (will) and progressive auxiliary da:s'tan (to have), as well as clitic pronouns like -es' (it) are annotated as INSEP. Table 1 represents these tags and the colors used for each one of them.

Tag's Guide		
Lemma	Past and present lemma of each LVC	White
LV	Light Verb	Green
PV	Preverb	Red
SEP	Separable	
INSEP	Inseparable	

Table1. Tag's guide

The first row in each excels file represents the Past and presents lemma of the LVCs, the second row the LVs and the third row the PVs.