

Computational Understanding of Hindi: The case of Verb Groups

-Narayan Choudhary, PhD Student, Jawaharlal Nehru University, New Delhi

Verb Groups: Constitutes of all the lexical elements that represent the verbal content in a sentence. Also called ‘complex predicate’ or ‘complex verb’. Verb Group (VG) is recognised as a chunk in morphological parsing (Markantonatoul, Stella et. al. 2005; Watanabe, Taro et. al. 2003).

Verb Groups in Hindi: Verbal content is spread over one or more than one lexical constituents in a sentence. Identification of the verb group involves morpho-syntactic analysis.

Information Contained in a Verb Group: Besides the semantic intent, a verb group contains grammatical information about tense, aspect and mood (TAM) and also sheds light on (and can also predict) gender, number and person (GNP) of the subject of the sentence (or clause).

A Verb Root can appear in 25 different forms, reflecting different TAM and GNP markings

VR/IMP	INF_MSG	INF_F_SG	INF_PL	IMPF_MSG	IMPF_FSG	IMPF_F_SGH	IMPF_PL	PFT_MSG	PFT_F_SG	PFT_F_SGH	PFT_PL /OPT_3SG	FUT_3MSG	FUT_2MSG	FUT_3_FSG	FUT_3_FSGH	FUT_PL	FUT_2MSG_GNH	FUT_2FS_GNH	FUT_1MSG	FUT_1_FSG	OPT_1SG	IMP_PLH	IMP_H	IMP_2S_GNH
रह	रहना	रहनी	रहने	रहता	रहती	रहतीं	रहते	रहा	रहा	रहा	रहे	रहेगा	राहएगा	रहेगी	रहगा	रहग	रहोगे	रहोगी	रहूँगा	रहूँगी	रहूँ	रह	राहए	रहो

Summary of the TAM Templates and TAM Tags of the Verb Groups in Hindi

Active Voice				Passive Voice			
TAM Tags	Template	V2s Affecting VG Templates	TAM Tags	Template	TAM Tags	Template	
prs.sim.dcl	prs_aux	MAUX /s k/	fut.sim.abil	V1+s k fut	pas_fut.sim.dcl	VR_pft+ja_fut	
pst.sim.dcl	pst_aux		prs.impf.abil	V1+s k_impf+prs_aux	pas_prs.impf.dcl	VR_pft+ja_impf+prs_aux	
fut.sim.dcl	VR_fut		pst.impf.abil	V1+s k_impf+pst_aux	pas_pst.impf.dcl	VR_pft+ja_impf+pst_aux	
prs.impf.dcl	VR_impf+prs_aux		prs.pft.abil	V1+s k_pft+prs_aux	pas_prs.pft.dcl	VR_pft+ja_pft+prs_aux	
pst.impf.dcl	VR_impf+pst_aux		pst.pft.abil	1.)V1+s k_pft+pst_aux	pas_pst.pft.dcl	VR_pft+ja_pft+pst_aux	
prs.pft.dcl	VR_pft+prs_aux			2.)V1+s k_pft		VR_pft+ja_pft	
pst.pft.dcl	1.)VR_pft+pst_aux		fut.pft.abil	V1+s k_pft+ho_fut	pas_fut.pft.dcl	VR_pft+ja_pft+ho_fut	
	2.)VR_pft		prs.prog.abil	V1+s k+r h_pft+prs_aux	pas_prs.prog.dcl	VR_pft+ja+r h_pft+prs_aux	
fut.pft.dcl	VR_pft+ho_fut		pst.prog.abil	V1+s k+r h_pft+pst_aux	pas_pst.prog.dcl	VR_pft+ja+r h_pft+pst_aux	
prs.prog.dcl	VR+r h_pft+prs_aux		fut.prog_prsm.abil	V1+s k+r h_pft+ho_fut	pas_fut.prog.prsm	VR_pft+ja+r h_pft+ho_fut	
pst.prog.dcl	VR+r h_pft+pst_aux	MAUX /pa/	fut.sim.abil	V1+pa_fut	pas_fut.sim.abil	VR_pft+ja+s k_fut	
fut.prog_prsm	VR+r h_pft+ho_fut		prs.impf.abil	V1+pa_impf+prs_aux	pas_prs.impf.abil	VR_pft+ja+s k_impf+prs_aux	
prs.impf_ince.dcl	1.)VIN+1 g_impf+prs_aux		pst.impf.abil	V1+pa_impf+pst_aux	pas_pst.impf.abil	VR_pft+ja+s k_impf+pst_aux	
prs.pft_ince.dcl	2.)VIN+1 g_pft+prs_aux		prs.pft.abil	V1+pa_pft+prs_aux	pas_prs.pft.abil	VR_pft+ja+s k_pft+prs_aux	
pst.impf_ince.dcl	1.)VIN+1 g_impf+pst_aux		pst.pft.abil	1.)V1+pa_pft+pst_aux	pas_pst.pft.abil	VR_pft+ja+s k_pft+pst_aux	
pst.pft_ince.dcl	2.)VIN+1 g_pft+pst_aux			2.)V1+pa_pft		VR_pft+ja+s k_pft	
	3.)VIN+1 g_pft		fut.pft.abil	V1+pa_pft+ho_fut	pas_fut.pft.abil	VR_pft+ja+s k_pft+ho_fut	
fut.sim_ince.dcl	1.)VIN+1 g_fut		prs.prog.abil	V1+pa+r h_pft+prs_aux			
	2.)VIN+1 g_pft+ho_fut		pst.prog.abil	V1+pa+r h_pft+pst_aux			
prs.impf_cnt.dcl	1.)VR_IMP+ r h_impf+prs_aux		fut.prog_prsm.abil	V1+pa+r h_pft+ho_fut			
prs.pft_cnt.dcl	2.)VR_IMP+ r h_pft+prs_aux	MAUX /de/	fut.sim.perm	VINF+de_fut			
pst.impf_cnt.dcl	1.)VR_IMP+ r h_impf+pst_aux		prs.impf.perm	VINF+de_impf+prs_aux			
pst.pft_cnt.dcl	2.)VR_IMP+ r h_pft+pst_aux		pst.impf.perm	VINF+de_impf+pst_aux			
	3.)VR_IMP+ r h_pft		prs.pft.perm	VINF+de_pft+prs_aux			
fut.cnt.dcl	VR_impf+r h_fut		pst.pft.perm	1.)VINF+de_pft+pst_aux			
prs.impf_dur.dcl	1.)VR_IMP+ja_impf+prs_aux			2.)VINF+de_pft			
prs.pft_dur.dcl	2.)VR_IMP+ja_pft+prs_aux		fut.pft.perm	VINF+de_pft+ho_fut			
prs.prog_dur.dcl	3.)VR_IMP+ja_prog+prs_aux		prs.prog.perm	VINF+de+r h_pft+prs_aux			
prs.prog_dur.dcl	VR_impf+c_l_pft+a+r h_pft+prs_aux		pst.prog.perm	VINF+de+r h_pft+pst_aux			
pst.impf_dur.dcl	1.)VR_IMP+ja_impf+pst_aux		fut.prog_prsm.perm	VINF+de+r h_pft+ho_fut			
pst.pft_dur.dcl	2.)VR_IMP+ja_pft+pst_aux	MAUX /p /	fut.sim.oblg	VINF+p_fut			
	4.)VR_IMP+ja_pft		prs.impf.oblg	VINF+p_impf+prs_aux			
pst.prog_dur.dcl	3.)VR_IMP+ja_prog+pst_aux		pst.impf.oblg	VINF+p_impf+pst_aux			
pst.prog_dur.dcl	VR_impf+c_l_pft+a+r h_pft+pst_aux		prs.pft.oblg	VINF+p_pft+prs_aux			
fut.dur.dcl	VR_impf+ja_fut		pst.pft.oblg	1.)VINF+p_pft+pst_aux			
prs.impf_frq.dcl	VR_pft+k_r_impf+prs_aux			2.)VINF+p_pft			
pst.impf_frq.dcl	1.)VR_pft+k_r_impf+pst_aux/		fut.pft.oblg	VINF+p_pft+ho_fut			
pst.pft_frq.dcl	2.)VR_pft+k_r_impf		prs.prog.oblg	VINF+p_r h_pft+prs_aux			
fut.impf_frq.dcl	VR_pft+k_r_impf+ho_fut		pst.prog.oblg	VINF+p_r h_pft+pst_aux			
0.0.imp	VR		fut.prog_prsm.oblg	VINF+p_r h_pft+ho_fut			
0.0.opt	VR_opt	MAUX /ho/	fut.sim.oblg	VINF+ho_fut			
fut.impf_prsm	VR_impf+ho_fut		prs.impf.oblg	VINF+ho_impf+prs_aux			
fut.impf_dur_prsm	VR_impf+ja+r h_pft+ho_fut		pst.impf.oblg	VINF+ho_impf+pst_aux			
pst.impf.cfct	VR_impf+c_l_pft+a+r h_pft+ho_fut		prs.pft.oblg	VINF+ho_pft+prs_aux			
pst.impf.cfct	VR_impf+ho_impf		pst.pft.oblg	1.)VINF+ho_pft+pst_aux			
pst.impf.cgnt	VR_IMP			2.)VINF+ho_pft			
fut.impf.cgnt	VR_impf+ho_prsm_aux		fut.pft.oblg	VINF+ho_pft+ho_fut			
			prs.prog.oblg	VINF+prs_aux			
			pst.prog.oblg	VINF+pst_aux			
			fut.prog_prsm.oblg	VINF+ho_fut			
		MAUX /cahiye/	prs.sugg.dcl	VINF+cahiye			
			pst.sugg.dcl	VINF+cahiye+pst_aux			

↑ Colour Codes:

One Word VG
Two Word VG
Three Word VG
Four Word VG
Five Word VG

Algorithm: How to identify the VG

- Step 1:** Input PoS tagged text in Hindi.
- Step 2:** Start searching for verbal word (as defined in the list provided) from the right of the sentence boundary.
- Step 3:** When a verbal word is found, match it with a template and store it.
- Step 4:** Continue the search looking rightward for more verbal words and match it with the TAM template. Continue it till the last verbal word found in the sequence.
- Step 5:** From among the templates matched, choose the longest verb sequence matched with the TAM template and mark the boundary of the verb sequence within square braces '[']'.
- Step 6:** Assign the VG Tag corresponding to the TAM Template matched at the end of the square bracket, prefixed with /VG./.

Constituents of a Verb Group in Hindi

- Main Verb:** Mandatory. the constituent that carries the main semantic intent (action or state) of the verb group.
- Auxiliary Verb:** Optional. Contains information on the TAM.
- V2** (also called ‘explicator’ or ‘vector’ or ‘light’ verb): Optional. Adds (additional) aspectual or modal shades to the intent.
- Conjunct Verb:** Optional. The noun or adjective (i.e. the conjunct) that gives the intent to the verb and is always followed by a V2.
- Conjunctive Participle:** Optional. A verb root, invariably followed by the morpheme ‘k r’.
- Particles:** Optional. A closed set of indeclinables that can be inserted amidst the constituents of verb group.

Verbal Elements in a Verb Group

As shown in the Summary table, we find that there are verb groups in Hindi that can have upto a total of five verbal words in Active voice, contained in 47 VG templates.

- One Word Verb Group: contained in 7 templates
- Two Word Verb Group: contained in 15 templates
- Three Word Verb Group: contained in 19 templates
- Four and Five word Verb Group: contained in three templates each

Passive Voice construction in Hindi is limited to the main three aspects of imperfect, perfect and progressive. The only other mood that can have a passive construction is the capability /s k/. The total templates added is ten.

In a VG with a compound verb, it is the V2 that always takes the TAM and agreement markings while the V1 remains in the root form. However, there are a few V2s that allow their V1 to be in non-root form. These are: *r h, s k, cuk, pa, l g, de, ja, c l, ho, p* and *k r*. These V2s add aspectual or modal values and also add to the templates. Including these additional templates, we get a total of 115 templates.

VG Tags

Out of these 115 VG templates, we draw a total of 104 TAM tags. These tags contain information about the tense, aspect and mood of the VG. We get 11 lesser TAM tags than the VG templates because in the past perfective cases, we get only one TAG because the two representations in this case basically denote the same TAM category.

VG Tags are denoted as: VG+TAM Tag

One can add markings for GNP as well into the tag. VGs identified with GNP tags as well will give information about the subject agreement as well and can also predict it to some extent.