

A Macedonian E-dictionary of Simple Verb Forms

Aleksandar Petrovski

Tetovo, Macedonia

sise@mt.net.mk

The construction of verbs

Word form: prodav-a-m (PR+1+s)



prodav-a (PR+3+s)

Canonical forms: nos-i (PR+3+s)

dojd-e (PR+3+s)

Inflectional categories

Category		NooJ Code	Example
Gender	Masculine	m	gledal
	Feminine	f	gledala
	Neuter	n	gledalo
Number	Singular	s	gleda
	Plural	p	gledaat
Person	First	1	gledam
	Second	2	gledash
	Third	3	gleda
Tense	Present	PR	gleda
	Aorist	AO	padna
	Imperfect	IP	gledashe
Form	Imperative	Imprt	gledaj
	L-form (active participle)	Lform	gledal
	Verbal adjective (passive participle)	Vadj	gledan
	Verbal noun (substantive)	Vnoun	gledanje
	Verbal adverb (gerund)	Vadv	gledajkji

Inflectional rules

DAVA = (m/PR+1+s + sh/PR+2+s + <E>/PR+3+s + me/PR+1+p + te/PR+2+p) + at/PR+3+p
+ v/IP+1+s + she/IP+2+s + she/IP+3+s + vme/IP+1+p + vte/IP+2+p + a/IP+3+p
+ l/Lform+m+s + la/Lform+f+s + lo/Lform+n+s + le/Lform+p
+ (<E> + <B2>)(j/Imprt+2+s + jte/Imprt+2+p)
+ jkji/Vadv + n/Vadj + nje/Vnoun
);

ZAPRE = (am/PR+1+s + esh/PR+2+s + e/PR+3+s + eme/PR+1+p + ete/PR+2+p + at/PR+3+p
+ e(v/AO+1+s + <E>/AO+2+s + <E>/AO+3+s + vme/AO+1+p + vte/AO+2+p + a/AO+3+p)
+ e(b/IP+1+s + she/IP+2+s + she/IP+3+s + vme/IP+1+p + vte/IP+2+p + a/IP+3+p)
+ e(l/Lform+m+s + la/Lform+f+s + lo/Lform+n+s + le/Lform+p)
+ i/Imprt+2+s + ete/Imprt+2+p
+ en/Vadj
);

Syntactic categories

Category		NooJ code
Aspect	Imperfective	Impf
	Perfective	Perf
Reflexiveness	Reflexive	Ref
	Not reflexive	Iref
Transitivity	Transitive	Tr
	Not transitive	Itr

Part of the dictionary

продава, V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf

продаде, V+FLX=DODADE+Iref+Ref+Tr+Perf

продемонстрира, V+FLX=ZDUVA+Iref+Tr+Perf

продене, V+FLX=SKINE+Iref+Ref+Tr+Perf

продени, V+FLX=IZNOSI+Ref+Perf

проденува, V+FLX=DAVA+Iref+Tr+Ref+Impf

продефилира, V+FLX=ZDUVA+Iref+Itr+Perf

продира, V+FLX=IMA+Iref+Itr+Impf

продискутира, V+FLX=ADRESIRA+Iref+Tr+Itr+Impf+Perf

продлаби, V+FLX=IZNOSI+Iref+Itr+Tr+Perf

The paradigm of 'prodava' (23 word forms)

- prodavam,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+PR+1+s
- prodavash,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+PR+2+s
- prodava,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+PR+3+s
- prodavame,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+PR+1+p
- prodavate,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+PR+2+p
- prodavaat,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+PR+3+p
- prodavav,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+IP+1+s
- prodavashe,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+IP+2+s
- prodavashe,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+IP+3+s
- prodavavme,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+IP+1+p
- prodavavte,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+IP+2+p
- prodavaa,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+IP+3+p
- prodaval,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+Lform+m+s
- prodavala,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+Lform+f+s
- prodavalo,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+Lform+n+s
- prodavale,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+Lform+p
- prodavaj,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+Impprt+2+s
- prodavajte,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+Impprt+2+p
- prodaj,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+Impprt+2+s
- prodajte,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+Impprt+2+p
- prodavajkji,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+Vadv
- prodavan,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+Vadj
- prodavanje,prodava,V+FLX=DAVA+Iref+Ref+Tr+Itr+Impf+Vnoun

Dictionary figures

Group	Lemmas	Inflectional rules	Word forms	Inflectional factor
a-group	11,653	11	263,033	22.57
i-group	5,340	31	149,831	28.06
e-group	2,988	52	87,312	29.22
Other	5	2	44	8.80
Total	19,986	96	500,220	25.03

Ambiguities

- The same lemma for different PoS
mora, pravda
pravda, N+FLX=DEVOJKA
pravda, V+FLX=IMA+Iref+Ref+Tr+Impf
- The lemma of one word the same with the word form of another lemma
luta (V+PR+3+s, A+f+s+Dn+pst)
- The same word forms for different inflectional categories
prodavashe(IP+2+s, IP+3+s)
zapre (PR+2+s, PR+3+s, AO+2+s, AO+3+s)

Applying dictionaries - results

		Verb dictionary	Complete dictionary
1.	Different tokens	15,660	15,660
2.	Unknowns	10,980	1,808
3.	Annotations	6,773	16,921
4.	Unique recognized tokens (1-2)	4,680	13,852
5.	Level of ambiguity (3/4)	45%	22%

Present situation

PoS	Dictionary entries	Word forms	Inflectional factor	Inflectional rules
Nouns	34,480	315,338	9,15	148
Adjectives	9,907	468,285	47,27	33
Pronouns	47	150	3,19	21
Verbs	19,986	500,220	25,03	96
Adverbs	2,797	9,259	3,31	4
Numerals	88	364	4,14	6
Prepositions	63	63	1	1
Conjunctions	62	62	1	1
Particles	63	63	1	1
Interjections	142	142	1	1
Total	67,635	1,293,946	19,13	312

Future work

- Addition of proper nouns
- Developing syntactic graphs to solve the ambiguities
- Addition of compound words