Formalizing MultiWords as Catenae in a Treebank and in a Lexicon

Kiril Simov and Petya Osenova

IICT-BAS

PARSEME Meeting, Malta, 19-20 March 2015

Representation in the Treebank (CONLL format)

No	Wf	Le	POS	ExPOS	GramFeat	Head	Rel	Catena
1	Те	те	P	Pp	num=pll case=nom	3	subj	_
2	си	си	P	Pp	form=possesive	3	clitic	c_1
3	затварят	затварям	V	Vpi	num=pll pers=3	0	Root	c_1
4	очите	око	N	Nc	num=pll def=y	3	obj	c_1
5	пред	пред	R	R	_	3	indobj	_
6	истината	истина	N	Nc	num=sgl def=y	5	prepobj	_

Representation in the Lexicon

[**form:** < затварям си очите >

catena:

No	Wf	Le	POS	ExPOS	GramFeat	Head	Rel
1	_	затварям	V	Vpi	_	0	CRoot
2	си	си	P	Pp	form=possesive	1	clitic
3	очите	око	N	Nc	number=pl‖ definite-ness=y	1	obj

semantics:

No1: { run-away-from_rel(e,x_0,x_1), fact(x_1), [1](x_1) }

valency:

No1: < :indobj: x/Prep :prepobj: y/N[1] ‖ x ∈ { пред, за } >

]