Formalising Multilingual Verbal MultiWord Expressions to Support Neural Machine Translation

CA21167 COST Action UniDive - (WG2 & WG3) 16-18 March 2023 - Paris-Saclay University

Johanna Monti and Maria Pia di Buono
UniOr NLP Research Group
Department of Literary, Linguistic and Comparative Studies
University of Naples L'Orientale

Italian Clitic Verbal VMWEs

Class of VMWE made up by a verb with at least one deficient pronoun (Cardinaletti and Starke, 1999) whose semantics is not derivable from its constituent parts.

1) *Prendersela* → 'to be offended',
2) *Buttarsi giù* → 'to get depressed',
3) *lasciarsi andare* → 'to relax'

The class of Italian pronouns lexicalised as verbal affixes is made up of -si pronouns, traditionally referred to as "reflexive pronouns" (*mi, ti, si, ci, vi, si*), the third person feminine of clitic complement pronouns (*la, le*) and two fixed pronominal Particles: *ci* and *ne*, generally having locative or demonstrative meaning.

We take into account only clitic VMWEs which either:
- do not exist without the clitic, such as *arrabbiarsi (*to be angry*), suicidarsi, pentirsi, vergognarsi*
- have a different meaning with respect to the non-clitic verb form, such as
- are constitutive elements of more complex idiomatic expressions, such as *darsela a gambe* 'cut and run', *farla finita* 'to end it', *darci dentro* 'knuckle down', *prenderci gusto* 'get a taste for'

Type	Source	#VMWEs
Procomplement	De Mauro	172
Idiom	De Mauro	239
IRV	PARSEME-IT	252
Total		**663**

Monti, J., di Buono, M.P., Caruso, V. (forthcoming) Ontology-based formalisation of Italian Clitic Verbal MWEs: an Approach or Supporting Machine Translation in Monti, J. Corpas Pastor, G., and Mitkov, R. (eds.) John Benjamin Publishing Company

Modelisation

Name	Features	Example	#VMWEs
MSP1	Inflecting -si pronoun	*Darsi*	413
MSP2	Fixed Pronominal particle	*Darci*	85
MSP3	Inflecting -si pronoun + fixed pronominal particle	*Andarsene*	29
MSP4	Direct object pronoun with fixed number (singular) and gender (feminine)	*Darla*	50
MSP5	Direct object pronoun with fixed number (plural) and gender (feminine)	*Darle*	12
MSP6	Inflecting -si pronoun + direct object pronoun with fixed number (singular) and gender (feminine)	*Farsela*	63
MSP7	Inflecting -si pronoun + direct object pronoun with fixed number (plural) and gender (feminine)	*Darsele*	4
MSP8	Pronominal particle + direct object pronoun with fixed number (singular) and gender (feminine)	*Farcela*	2
MSP9	Inflecting -si pronoun + direct object pronoun with fixed number (singular) and gender (masculine)	*Menarselo*	1
MSP10	Direct object pronoun with fixed number (singular) and gender (masculine)	*Farlo a fette*	1
MSP11	Fixed Pronominal particle + Fixed Pronominal particle	*Andarcene*	3

Name	Pattern	Example	#VMWEs
SyP1	VERB + -si pronoun	*Darsi*	252
SyP2	VERB + direct object pronoun	*Darla*	39
SyP3	VERB + direct object pronoun + ADJ	*Darla vinta*	15
SyP4	VERB + -si pronoun + direct object pronoun	*Darsele*	58
SyP5	VERB + -si pronoun + DET + NOUN	*Farsi un baffo*	41
SyP6	VERB + -si pronoun + ADP + NOUN	*Togliersi di mente*	11
SyP7	VERB + -si pronoun + VERB	*Farsi mantenere*	13
SyP8	VERB + -si pronoun + ADV	*Farsi sotto*	8
SyP9	VERB + pronominal particle + ADV	*Darci dentro*	10
SyP10	VERB + pronominal particle + DET + NOUN	*Metterci la faccia*	17
SyP11	VERB + pronominal particle	*Volerci*	42
SyP12	VERB + -si pronoun + pronominal particle	*Andarsene*	24
SyP13	VERB + -si pronoun + NOUN	*Farsi onore*	44
SyP14	VERB + -si pronoun + ADP + DET + NOUN	*Darsi all'ippica*	15

Name	Features and patterns	Type	Example	#VMWEs
SD1	MSP1+SyP1	Pronominal intransitives	*Annoiarsi*	252
SD2	MSP1+SyP13	Idiom	*Darsi importanza*	44
SD3	MSP1+SyP14	Idiom	*Farsi dal nulla*	15
SD4	MSP1+SyP5	Idiom	*Darsi un contegno*	41
SD5	MSP1+SyP7	Idiom	*Farsi conoscere*	13
SD6	MSP2+SyP11	Pro-complement w/ pronominal particle	*Contarci*	42
SD7	MSP2+SyP9	Idiom	*Darci dentro*	10
SD8	MSP4+SyP2	Pro-complement w/ direct object	*Darla*	31
SD9	MSP3+SyP12	Pro-complement w/ -si pronoun and pronominal particle	*Andarsene*	24
SD10	MSP4+SyP3	Idiom	*Farla finita*	12
SD11	MSP6+Syp4	Pro-complement w/ -si pronoun and direct object	*Cavarsela*	53
SD12	MSP1+SyP6	Idiom	*Darsi da fare*	11
SD13	MSP2+SyP10	Idiom	*Darci dentro*	17

Ontolex-based Translation
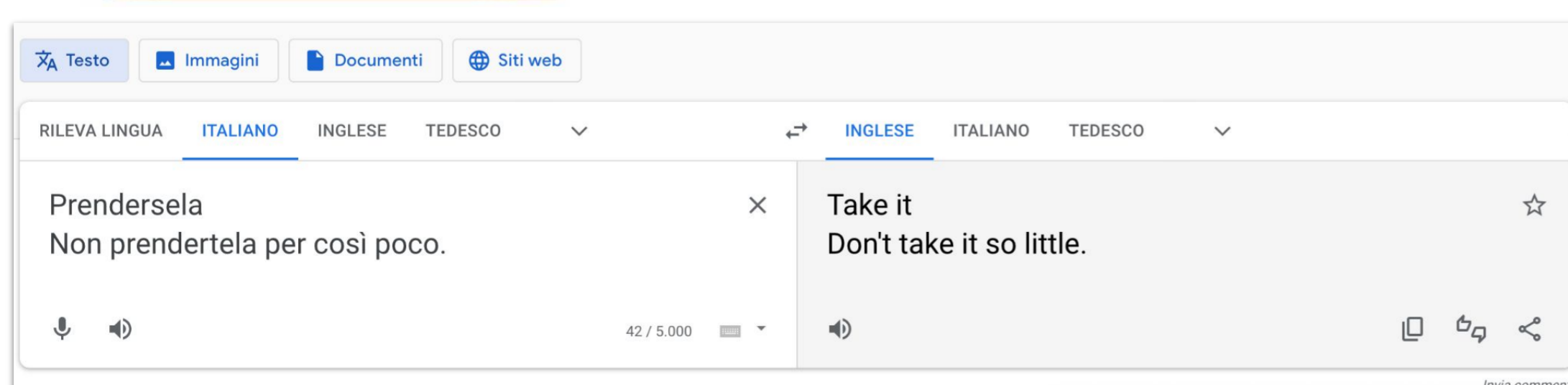

```
:prendersela v it sense-to_be_offended_v_en_sense      a
vartrans:Translation ;
    vartrans:source :prendersela v it sense ;
    vartrans:target :to_be_offended_v_en_sense .

# Examples
:prendersela v it sense ex a lexicog:UsageExample;
rdf:value "Non prendertela per così poco."@it;
rdf:value "Don't be offended for so little."@en.
```