

# **Annotation of Multiwords in BulTreeBank (current status and perspectives)**

Petya Osenova and Kiril Simov  
IICT-BAS

PARSEME Meeting, Athens, 2014



# MWE annotation: current state

- *fixed-expressions (complex POS)* - 1081 tokens within 214 000 tokens

## (IN)VISIBLE MWE:

- *semi-fixed expressions*:
  - Idioms - treated syntactically (no difference between: *kick the door* and *kick the bucket*)
  - Proper names – treated as constructs
- *syntactically-flexible expressions* – treated syntactically

# MWE annotation: perspectives

- **Selection-based** – depends on the lexical meaning of the elements, selected by the head (*lose time* = idiom, but *lose wallet* = phrase)
- **Construction-based** – idioms ‘*from needle to thread*’ (from the beginning to end)
- **Catena-based** – for idiosyncratic cases and non-constituent expressions

(**VPS** Той (**VPC-C** (**V-C** рутна) (**N-C** камбаната))) =  
He kicked the bucket.

