

## Corpus

Petya Osenova, Aleksandar Popov and Kiril Simov (IICT, Bulgarian Academy of Sciences)  
WG1: Lexicon-Grammar Interface

### 1. Overview

The aim is to highlight the following issues:

- Realizations of different MWE types in two languages with different morphological complexity and word order
- A typology of alignment possibilities among various types of MWEs

### 2. Data and Preliminary Observations

#### Corpus:

The parallel BG-EN corpus consists of: SETIMES, PenTreeBank and CSLI dataset: 1758 sentences (31 257 tokens).

#### Types of alignments:

##### MWE-to-Word

<b>EN MWE-to-BG word</b>	<p><b>Adverbs:</b> A little bit : леко; In the forefront : начело; For example : например, примерно; In fact : всъщност; A total of : общо; In due course : своевременно; In part : отчасти; In the meantime : междувремен.</p> <p><b>Prepositions:</b> According to : според</p> <p><b>Particles:</b> Even if : дори</p> <p><b>Phrasal verbs:</b> Drive out : отклоняваше; Look for : търсят; We cleared out : очистихме; Have in mind : планират; Shut down : затваря; Pay down : плаща</p> <p><b>Nominal compounds:</b> Health care : здравеопазване; Labor federation : профсъюзите; Stock shares : акции; Interest rate : лихва; Gasoline stations : бензиностанции</p>
<b>BG MWE-to-EN word</b>	<p><b>se-verbs:</b> Complains : оплаква се; Surged : повиши се; Dropped : понижиха се; Focus : съсредоточат се; Get : сдобие се; Rose : покачи се</p> <p><b>Light verb constructions:</b> Proposed : направиха предложение; Say : изказват мнение; Resigned : подаде оставка</p> <p><b>Nominal MWE:</b> Banking : банково дело; Mining : минното дело; Cornerstone : крайгълен камък; Milestones : крайгълни камъни; Chemicals : химическите продукти; Luxuries : луксозни стоки; Toiletries : тоалетни принадлежности; Outlays : капиталовите разходи; Deadline : крайният срок; Railroad : железопътни линии; Reality : недвижимо имущество; Officials : длъжностни лица; Cash : пари в брой</p> <p><b>Various connectives:</b> Tentatively : за момента; Although : въпреки че; Ultimately : в крайна сметка; During : по време на; Nonetheless : въпреки това; Also : също така; Unlike : за разлика от; Within : в рамките на</p>

##### MWE-to-MWE

<b>EN MWE verbal constructions to BG MWE verbal constructions</b>	<p><b>Light verb, Idioms, Phrasal verbs, Se-verbs</b></p> <p>Reach a decision : взема решение; Putting pencil to paper : предприел действие; Is drawing fire : привлича критиките; Will be priced out of a job : ще загубят работата си; Have not got a clue : нямат представа; Take effect : влезе в сила; Broke even : са излезли на чисто; Take control : влезе във владение; Move back : се върнат; Jumped at : се набиват на очи; Give up : откаже се</p>
<b>MWE Connectives</b>	<p>In particular : в частност; With respect to : що се отнася до; As well as : както и; At times : от време на време; Pros and cons : доводи за и против; By contrast : за сравнение; As a whole : като цяло; On the other hand : от друга страна; In turn : на свой ред; Of course : разбира се; More and more : все повече и повече; In particular : в частност</p>
<b>EN MWE nominal compounds to BG MWE nominal phrases</b>	<p>Free market : свободния пазар; Real estate : Недвижимото имущество; Tough line : твърда позиция; Law enforcement : силите на реда; Face amount : номинална стойност</p>

##### MWE-to-phrase

<b>EN MWE to BG phrase</b>	<p><b>EN Nominal compounds to BG noun phrases</b></p> <p>Labor economist : специалист по икономика на труда; Life style : начин на живот</p> <p><b>EN Verbal idioms to BG phrases</b></p> <p>To be in a long wait : дълго да чакат</p>
<b>BG MWE to EN phrase</b>	<p><b>BG prepositional MWE to EN phrase</b></p> <p>These results compare : за сравнение</p>

### 3. Mapping to PARSEME Classifications

Within MWE-MWE relations the correspondences are:

#### Straightforward

- Light verbs in one language often correspond to similar constructions in the other
- Complex prepositions in English tend to have structurally similar counterparts in Bulgarian
- Nominal MWEs of the kind A N seem to be often preserved across the two languages
- The MWE construction V NP (not mentioned in WP1) also seems to be sometimes preserved in translation
- Multi-word adverbial constructions
- Conjunctions composed of multiple words

#### Cross-language specific types

- English phrasal verbs often correspond to Bulgarian se-verbs
- N N compounds in English can map to A N compounds in Bulgarian
- N N in English can also be translated as N PP in Bulgarian, where the meaning carried by the first noun is expressed through a PP
- N and N constructions in English can apparently translate in similar coordinated constructions in Bulgarian, but constructed out of different POS constituents
- An idiomatic clausal construction (V NP PP) can be translated with a light verb construction in Bulgarian
- V AP (not in WP1) in English can be translated with minimal changes into V AdvP in Bulgarian
- V PP (not in WP1) in English can be translated as V NP in Bulgarian

### 4. Conclusions

**MWE-to-MWE** and **MWE-to-word** are more frequent than **MWE-to-phrase**

Such knowledge can be used in:

- designing automatic translation systems
- singling out best practices in human translation
- can illuminate the ways the two languages differ in expressing meaning