

## **Knowledge- and resource-poor identification and translation of multiword expressions across language pairs**

Ruslan Mitkov

The correct identification, interpretation and translation of multiword expressions in both general and specialised languages, is vital for the successful operation of most Natural Language Processing applications and computer-aided tools supporting various users including phraseologists, translators, interpreters, terminologists, language learners and teachers,. As parallel corpora are scarce and in order to benefit from the wider availability of comparable corpora (which can be also compiled with a specific task in mind), there is a pressing need to develop approaches which can extract and translate multiword terms from comparable corpora. In this presentation the speaker will propose a novel methodology based on computing semantic similarity to extract and translate multiword units from comparable corpora termed as 'knowledge- and resource-poor' as it is not dependent on the use of any dictionaries or linguistic knowledge. While this particular study focuses on English and Spanish, the methodology is not restricted to any particular pair of languages.