

xLiTe: Cross-Lingual Technologies  
NIPS 2012 WORKSHOP  
<http://km.aifb.kit.edu/ws/xlite/>

## Inducing Cross-Lingual Semantic Representations of Words, Phrases, Sentences and Events.

Ivan Titov - Machine Learning for Natural Language Processing - Saarland University - Germany

Cross-lingual representations of linguistic units (e.g., words or phrases) can facilitate transfer of annotation from resource-rich to resource-poor languages and have many potential multilingual applications (e.g., machine translation and cross-lingual information retrieval). In this talk, I will discuss our ongoing work which aims to induce cross-lingual representations relying primarily on monolingual unannotated texts readily available for many languages.

From the learning standpoint, our approaches maximize the likelihood of monolingual unannotated texts but also use a form of regularization which favors agreement on a smaller collection of parallel data (i.e. sentences along with their translations). I will address the induction of different types of cross-lingual representations (clusters and distributed representations) for different types of units (words, phrases and predicate-argument structures). We show that these models induce linguistically-plausible semantic representations and that cross-lingual induction both helps to induce better representations for individual languages and benefits various cross-lingual applications. Specifically, I will consider direct transfer of a classifier for a document classification task from one language to another, and show preliminary results in the context of low resource machine translation.

Joint work with Alexandre Klementiev, Binod Bhattarai, Chris Callison-Burch and Ann Irvine.