Syntax and Semantics in Quality Estimation of Machine Translation

Rasoul Kaljahi†‡, Jennifer Foster†, Johann Roturier‡

Presented by: Joachim Wagner†

†NCLT, School of Computing, Dublin City University, Ireland
{rkaljahi, jfoster, jwagner}@computing.dcu.ie

‡Symantec Research Labs, Dublin, Ireland
johann_roturier@symantec.com
Introducing a new QE data set: **SymForum**

- 4500 En-Fr machine translation segments
- User-generated content - Not edited!
- Technical software support domain - Not newswire!
- Publicly available
Syntax-based quality estimation using:

- Constituency and dependency syntax
- Tree kernels and hand-crafted features
Semantic-based quality estimation using:

- Semantic role labelling
- Proposition subtree kernels
- Augmented syntactic tree kernels
- Hand-crafted features
- Predicate-argument structure match
Predicate-Argument Structure Match: PAM

- A new QE metric
- Based on word-alignment of predicates and arguments of source and target
- Used both as estimations and as features
Manual analysis of PAM

- Finding what hinders PAM performance
Combining QE systems:

- All methods used in this work
- With WMT 17 surface feature
Results in the poster