# CONLL 2014

## 1Hybrid systems and type ﬁltering

### Rule-base(RBS)

### Statistics machine translation(SMT)

1. Word alignment: pialign | GIZA++ | Berkerley Aligner
2. character-level Levenshtein distance to each mapping in the phrase table
3. lexical reordering model: maximum phrase length=7
4. The IRSTLM Toolkit (Federico et al., 2008) was used to build a 4-gram target language model with Kneser–Ney smoothing
5. Decoding: Moses
6. Segmentation, tokenization, pos: NLTK

### Ranking

Language model: Microsoft’s Web LM

## 3Data-Intensive and Feature-Rich Statistical Machine Translation

Phrase-based part of the statistical machine translation system Moses

NUCLE ??

LM: WikiLM | CCLM

## 6RACAI GEC – A hybrid approach to Grammatical Error Correction

### Preprocess

Tokenization, pos, lemmatization: Bermuda software suite

### Statistical model

1. LM: Google 1T n-gram corpus
2. Distance function: weighted Levenstein algorithm

### Rule based model

## 4POSTECH grammatical error correction system