

Language Processing for Different Domains and Genres: Project: POS Tagger Adaptation

Caroline Sporleder, Ines Rehbein

Universität des Saarlandes

Wintersemester 2009/10

10.12.2009

Data: Target Domains

- German: fiction, recipes, or
- English: fiction, recipes

Data: Source Domain

- gold standard pos-tagged data from Penn Treebank or TIGER (or TüBa), split into training and test set

POS Taggers

two of your choice

Which of the domain adaptation methods work best and why?

- 1 source only (Baseline 1)
- 2 target only (Baseline 2)
- 3 source plus small annotated set from target
- 4 source plus target with weighting
- 5 your own method