End-to-end vs. Component Eval.

- Similar to black vs. glass box evaluation and translation vs. task-based evaluation.
- Evaluation of a single component may not correlate with overall performance of the system.

Pre-processing	Symmetrization	Alignment Error Rate	BLEU
Lemmas + singletons	Intersection	**14.6**	30.8
Lemmas	Intersection	15.0	29.8
Lemmas	Union	17.2	**32.0**
Lemmas + singletons	Union	17.4	31.9
Baseline (word forms)	Union	25.5	29.8
Baseline (word forms)	Intersection	27.4	28.2

Data by Bojar et al. (2006). See also e.g. Lopez and Resnik (2006).