Hybridization: Pre-AMTA 2006

• Preprocessing:
  – Morphology, NE’s $\rightarrow$ any MT method
  – Reordering rules $\rightarrow$ EBMT, SMT, CBMT

• Postprocessing:
  – Any set of MT methods $\rightarrow$ MEMT

• Integrated:
  – SMT + EBMT (phrase tables, decoders) $\approx$ SMT++
  – Syntax “SMT” (e.g. ISI, UMD, …)
Hybridization: Post-AMTA 2006

- **Preprocessing:**
  - Morphology, reordering, … $\rightarrow$ any MT method
  - Linguistic S\(\rightarrow\)T normalization beyond reordering

- **Postprocessing:**
  - Any set of MT methods $\rightarrow$ MEMT
  - Lattice (SMT, EBMT) $\rightarrow$ CBMT flood + COD

- **Integrated:**
  - Shared phrase tables, syntax in SMT, …
  - Semantic units (beyond NE’s: propositions, etc…)
Beyond Hybridization
Exploiting Cheap Plentiful Data

• Exploit Monolingual Corpora Better
  – E.g. CMBT

• Recruit Vast Quantities of Bilinguals (Web)
  – To produce bilingual texts
  – To correct MT output \(\rightarrow\) generalized automated improvement of underlying MT system

• Beg, borrow, buy and steal TM’s?