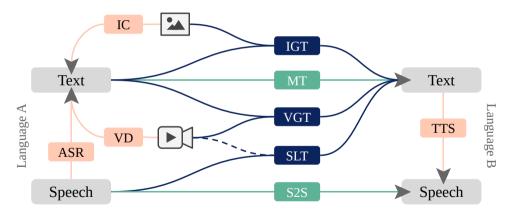
Overview of Multi-Modal MT

From survey by Sulubacak et al. (2019):



 $\mathsf{IGT} = \mathsf{image}\text{-}\mathsf{guided},\ \mathsf{VGT} = \mathsf{video}\text{-}\mathsf{guided}\ \mathsf{translation}$