

Table 1: The Transformer generalizes well to English constituency parsing (Results are on Section 23 of WSJ)

Parser	Training	WSJ 23 F1
Vinyals & Kaiser et al. (2014) [?]	WSJ only, discriminative	88.3
Petrov et al. (2006) [?]	WSJ only, discriminative	90.4
Zhu et al. (2013) [?]	WSJ only, discriminative	90.4
Dyer et al. (2016) [?]	WSJ only, discriminative	91.7
Transformer (4 layers)	WSJ only, discriminative	91.3
Zhu et al. (2013) [?]	semi-supervised	91.3
Huang & Harper (2009) [?]	semi-supervised	91.3
McClosky et al. (2006) [?]	semi-supervised	92.1
Vinyals & Kaiser et al. (2014) [?]	semi-supervised	92.1
Transformer (4 layers)	semi-supervised	92.7
Luong et al. (2015) [?]	multi-task	93.0
Dyer et al. (2016) [?]	generative	93.3