

English			
system	precision	recall	F1 score
rbal	30.1%	24.3%	26.8%
lbal	27.8%	20.8%	23.8%
rand.init	25.1%	20.0%	22.3%
en $\rightarrow$ de	35.4%	30.6%	32.8%
en $\rightarrow$ fr	35.4%	30.2%	32.6%
German			
system	precision	recall	F1 score
rbal	39.1%	31.3%	34.8%
lbal	38.1%	27.6%	32.0%
rand.init	33.7%	25.9%	29.3%
de $\rightarrow$ en	46.1%	39.6%	42.6%
de $\rightarrow$ fr	46.7%	40.9%	43.6%
French			
system	precision	recall	F1 score
rbal	34.3%	28.7%	31.3%
lbal	32.5%	25.4%	28.5%
rand.init	26.1%	24.4%	25.3%
fr $\rightarrow$ en	44.4%	39.7%	41.9%
fr $\rightarrow$ de	46.9%	41.7%	44.2%

Table 2: Scores of baseline trees and our extracted trees using all attention heads, evaluated against standard syntactic parse trees.