

Deep RNNs Encode Soft Hierarchical Syntax: Supplementary Materials

1 Extended Results

A full list of the accuracy results from the syntax prediction experiments can be found in Table 1.

Source model	Prediction task	MFT	Layer				
			0	1	2	3	4
DP	POS	0.8801	0.9990	0.9964	0.9853	0.9434	0.8962
	Parent	0.8190	0.8681	0.9177	0.9347	0.9384	0.9349
	Grandparent	0.5422	0.5721	0.7538	0.7920	0.8094	0.8253
	Great-Grandparent	0.4511	0.4648	0.5909	0.6659	0.6826	0.7183
SRL	POS	0.8801	0.8732	0.9063	0.8691	0.7833	0.5392
	Parent	0.8190	0.7983	0.8727	0.8892	0.8835	0.7870
	Grandparent	0.5422	0.5041	0.6812	0.7394	0.7549	0.6325
	Great-Grandparent	0.4511	0.4412	0.5415	0.6159	0.6449	0.5493
MT	POS	0.8801	0.8618	0.9274	0.9198	0.9191	0.9195
	Parent	0.8190	0.8019	0.8975	0.9040	0.9088	0.9083
	Grandparent	0.5422	0.5368	0.7072	0.7311	0.7572	0.7776
	Great-Grandparent	0.4511	0.4361	0.5631	0.5909	0.6303	0.6752
LM	POS	0.8801	0.8608	0.9093	0.9359	0.9304	0.9165
	Parent	0.8190	0.8126	0.8724	0.9232	0.9137	0.9000
	Grandparent	0.5422	0.5237	0.6251	0.7862	0.7606	0.7189
	Great-Grandparent	0.4511	0.4249	0.4702	0.6423	0.6302	0.5971

Table 1: Table of accuracy results for the syntax feature prediction experiments with best performing layer in each source model/ prediction task pair in bold. “DP” refers to the dependency parsing model.