Linguistic Knowledge Can Enhance Encoder-Decoder Models (If You Let It)

Alessio Miaschi, Felice Dell'Orletta, Giulia Venturi

Institute for Computational Linguistics "Antonio Zampolli" (CNR-ILC), Pisa ItaliaNLP Lab – www.italianlp.it



Introduction

- Motivations:
 - Understanding "how linguistic concepts that were common as features in NLP systems are captured in neural networks" (Belinkov & Glass, *Transactions of the Association for Computational Linguistics 2019*) has been the focus of many recent studies
 - Fine-tuning on a intermediate supporting task and then on the target task consecutively is highly beneficial to improve pre-trained model's performance (Weller et al., *ACL 2022*)

Introduction

- Motivations:
 - Understanding "how linguistic concepts that were common as features in NLP systems are captured in neural networks" (Belinkov & Glass, *Transactions of the Association for Computational Linguistics 2019*) has been the focus of many recent studies
 - Fine-tuning on a intermediate supporting task and then on the target task consecutively is highly beneficial to improve pre-trained model's performance (Weller et al., *ACL 2022*)

Does a step of intermediate fine-tuning on linguistic tasks enhance the prediction on a target task that strongly relies on linguistic knowledge?

- Two-step approach:
 - Fine-tune the T5 models on several intermediate tasks
 - Multi- and single-task fine-tuning
 - Fine-tune the Linguistically-Informed (LI) models on the target task
- We saved checkpoints every 5 epochs, in order to monitor the impact of the approach at increasing snapshots of the models
- We tested the approach both in Italian and English and in a cross-lingual scenario



- Two-step approach:
 - Fine-tune the T5 models on several intermediate tasks
 - Multi- and single-task fine-tuning
 - Fine-tune the Linguistically-Informed (LI) models on the target task
- We saved checkpoints every 5 epochs, in order to monitor the impact of the approach at increasing snapshots of the models
- We tested the approach both in Italian and English and in a cross-lingual scenario



- Two-step approach:
 - Fine-tune the T5 models on several intermediate tasks
 - Multi- and single-task fine-tuning
 - Fine-tune the Linguistically-Informed (LI) models on the target task
- We saved checkpoints every 5 epochs, in order to monitor the impact of the approach at increasing snapshots of the models
- We tested the approach both in Italian and English and in a cross-lingual scenario



- Two-step approach:
 - Fine-tune the T5 models on several intermediate tasks
 - Multi- and single-task fine-tuning
 - Fine-tune the Linguistically-Informed (LI) models on the target task
- We saved checkpoints every 5 epochs, in order to monitor the impact of the approach at increasing snapshots of the models
- We tested the approach both in Italian and English and in a cross-lingual scenario



Data

• Intermediate tasks:

- 10 morpho- and syntactic characteristics of a sentence
 - selected on the degree of correlation between sentence-level complexity judgments and their values

						Crowdsourcing task: How dime	cuit is this
		Features	Corr	Features	Corr	sentence?	•
		Italian		English			
Profiling_UD [.]		char_per_tok	0.28	upos_dist_NUM	0.35	20 Italian and English native speakers were	
Tronning-OD.		upos_dist_ADJ	0.21	dep_dist_nummod	0.31	recruited through CrowdFlower to read each	CrowdFlower
extraction of feature		upos_dist_NUM	0.19	upos_dist_SYM	0.27	sentence and rate now unitcut it was	
		lexical_density	0.17	upos_dist_AUX	0.25		
values from ITA e ENG		dep_dist_aux	0.17	dep_dist_compound	0.25	Sentence:	
		dep_dist_mark	0.16	upos_dist_PRON	0.24	I wonder when we'll be able to relax.	
UD treebank 📲 🔤 🛛		aux_mood_dist_Ind	0.14	upos_dist_DET	0.23		
	/	obj_post	0.14	subord_prop_dist	0.17	How difficult is this sentence?	
		upos_dist_PUNCT	0.13	aux_form_dist_Fin	0.16	1 2 3 4 5	6 7
		subord_prop_dist	0.12	aux_mood_dist_Ind	0.14	very easy	very difficult

- Prediction of their distribution in the Italian and English versions of the Universal Dependency Treebanks extracted with Profiling-UD
- Target task:
 - corpus of 1,440 Italian and 2,400 English sentences manually rated by 20 crowdsourced workers for the level of perceived complexity on 1-7 Likert scale (Brunato et al., EMNLP 2018)

Models and Evaluation

Models:

Language	Model	Parameters				
English	t5-small	60M				
	t5-base	220M				
	t5-large	770M				
Italian	it5-small	60M				
	it5-base	220M				
	it5-large	738M				

Evaluation:

- We used Spearman correlation score as evaluation metric:
 - Intermediate tasks: Correlation between the gold value of each feature in the Italian or English treebank and the predicted value of the models for the intermediate tasks.
 - **Target task**: Correlation between average judgments of complexity and the complexity scores obtained with the fine-tuned LiT5 models.

Enhancing T5 with Linguistic Features

								itu	nan						
it5-small					it5-base					it5-large					
All	0.41	0.49	0.53	0.55	0.56	0.53	0.64	0.73	0.76	0.77	0.6	0.72	0.75	0.81	0.83
aux_mood_dist_Ind	0.17	0.31	0.34	0.38	0.4	0.36	0.73	0.81	0.86	0.87	0.59	0.81	0.87	0.89	0.9
char_per_tok	0.0056	-0.046	0.06	0.061	0.13	0.15	0.28	0.36		0.53	0.15	0.31	0.42	0.6	0.63
dep_dist_aux	0	0	0	0.14	0.17	0	0.12	0.68	0.81	0.85	0.074	0.59	0.71	0.81	0.8
dep_dist_mark	0	0	0.091	0.21	0.23	0	0.38	0.59	0.65	0.74	0.021		0.76	0.77	0.82
lexical_density	0.0054	0.14	0.15	0.2	0.17	0.21	0.22	0.22	0.25	0.29	0.18	0.18	0.17	0.2	0.19
obj_post	0.18	0.31	0.38	0.41	0.41	0.35	0.38	0.42		0.5	0.46	0.54	0.59	0.68	0.69
subord_prop_dist	0.51	0.52	0.58	0.63	0.64	0.63	0.68	0.77	0.8	0.79	0.59	0.7	0.71	0.75	0.77
upos_dist_ADJ	0.14	0.18	0.22	0.18	0.22	0.26	0.39	0.44	0.44	0.45	0.24	0.29	0.39	0.53	0.58
upos_dist_NUM	0	0	0	0	0	0	0.34	0.93	0.94	0.94	-0.024	0.91	0.9	0.92	0.92
upos_dist_PUNCT	-0.15	0.13	0.22	0.21	0.25	0.17	0.3	0.41	0.51	0.54	0.2	0.24	0.38	0.61	0.76
	5	10	15	20	25	5	10	15	20	25	5	10	15	20	25
t5-small					t5-base										
		1	t5-smal	I				t5-base					t5-large		
All	0.45	0.51	t5-smal 0.66	0.79	0.87	0.54	0.78	t5-base 0.88	0.89	0.9	0.89	0.92	t5-large 0.93	0.93	0.93
All aux_form_dist_Fin	0.45 0.55	0.51 0.66	t5-smal 0.66 0.76	0.79 0.84	0.87 0.85	0.54 0.69	0.78 0.74	t5-base 0.88 0.9	0.89 0.91	0.9 0.94	0.89 0.9	0.92 0.92	t5-large 0.93 0.94	0.93 0.95	0.93 0.95
All aux_form_dist_Fin aux_mood_dist_Ind	0.45 0.55 0.46	0.51 0.66 0.63	t <mark>5-smal</mark> 0.66 0.76 0.79	0.79 0.84 0.86	0.87 0.85 0.89	0.54 0.69 0.72	0.78 0.74 0.72	t5-base 0.88 0.9 0.86	0.89 0.91 0.9	0.9 0.94 0.9	0.89 0.9 0.92	0.92 0.92 0.93	t5-large 0.93 0.94 0.93	0.93 0.95 0.95	0.93 0.95 0.94
All aux_form_dist_Fin aux_mood_dist_Ind dep_dist_compound	0.45 0.55 0.46 0	0.51 0.66 0.63 0	t5-smal 0.66 0.76 0.79 0.14	0.79 0.84 0.86 0.35	0.87 0.85 0.89 0.52	0.54 0.69 0.72 0	0.78 0.74 0.72 0.16	t5-base 0.88 0.9 0.86 0.57	0.89 0.91 0.9 0.57	0.9 0.94 0.9 0.61	0.89 0.9 0.92 0.53	0.92 0.92 0.93 0.62	t5-large 0.93 0.94 0.93 0.64	0.93 0.95 0.95 0.63	0.93 0.95 0.94 0.68
All aux_form_dist_Fin aux_mood_dist_Ind dep_dist_compound dep_dist_nummod	0.45 0.55 0.46 0 0	0.51 0.66 0.63 0	t5-smal 0.66 0.76 0.79 0.14 0	0.79 0.84 0.86 0.35 0.5	0.87 0.85 0.89 0.52 0.7	0.54 0.69 0.72 0 0	0.78 0.74 0.72 0.16 0.65	t5-base 0.88 0.9 0.86 0.57 0.8	0.89 0.91 0.9 0.57 0.8	0.9 0.94 0.9 0.61 0.81	0.89 0.9 0.92 0.53 0.73	0.92 0.92 0.93 0.62 0.74	t5-large 0.93 0.94 0.93 0.64 0.83	0.93 0.95 0.95 0.63 0.8	0.93 0.95 0.94 0.68 0.81
All aux_form_dist_Fin aux_mood_dist_Ind dep_dist_compound dep_dist_nummod subord_prop_dist	0.45 0.55 0.46 0 0	0.51 0.66 0.63 0 0 0 0.72	t5-smal 0.66 0.76 0.79 0.14 0 0.75	0.79 0.84 0.86 0.35 0.5 0.81	0.87 0.85 0.89 0.52 0.7 0.85	0.54 0.69 0.72 0 0 0 0.64	0.78 0.74 0.72 0.16 0.65 0.78	t5-base 0.88 0.9 0.86 0.57 0.8 0.87	0.89 0.91 0.9 0.57 0.8 0.87	0.9 0.94 0.9 0.61 0.81 0.85	0.89 0.9 0.92 0.53 0.73 0.86	0.92 0.93 0.62 0.74 0.9	t5-large 0.93 0.94 0.93 0.64 0.83 0.89	0.93 0.95 0.95 0.63 0.8 0.89	0.93 0.95 0.94 0.68 0.81 0.88
All aux_form_dist_Fin aux_mood_dist_Ind dep_dist_compound dep_dist_nummod subord_prop_dist upos_dist_AUX	0.45 0.55 0.46 0 0 0 0.67 0	0.51 0.66 0.63 0 0 0 0.72 0	t5-smal 0.66 0.76 0.79 0.14 0 0.75 0.57	0.79 0.84 0.86 0.35 0.5 0.81 0.84	0.87 0.85 0.89 0.52 0.7 0.85 0.89	0.54 0.69 0.72 0 0 0 0.64 0.17	0.78 0.74 0.72 0.16 0.65 0.78 0.77	t5-base 0.88 0.9 0.86 0.57 0.8 0.87 0.87 0.9	0.89 0.91 0.9 0.57 0.8 0.87 0.93	0.9 0.94 0.9 0.61 0.81 0.85 0.94	0.89 0.9 0.92 0.53 0.73 0.86 0.9	0.92 0.93 0.62 0.74 0.9 0.96	t5-large 0.93 0.94 0.93 0.64 0.83 0.89 0.94	0.93 0.95 0.95 0.63 0.8 0.89 0.97	0.93 0.95 0.94 0.68 0.81 0.88 0.96
All aux_form_dist_Fin aux_mood_dist_Ind dep_dist_compound dep_dist_nummod subord_prop_dist upos_dist_AUX upos_dist_DET	0.45 0.55 0.46 0 0 0 0.67 0 0	0.51 0.66 0.63 0 0 0.72 0.72 0 -0.011	t5-smal 0.66 0.76 0.79 0.14 0 0.75 0.57 0.33	0.79 0.84 0.86 0.35 0.5 0.81 0.81 0.84 0.62	0.87 0.85 0.89 0.52 0.7 0.85 0.89 0.81	0.54 0.69 0.72 0 0 0.64 0.17 0.14	0.78 0.74 0.72 0.16 0.65 0.78 0.77	t5-base 0.88 0.9 0.86 0.57 0.8 0.87 0.9 0.84	0.89 0.91 0.9 0.57 0.8 0.87 0.93 0.84	0.9 0.94 0.9 0.61 0.81 0.85 0.94 0.88	0.89 0.92 0.53 0.73 0.86 0.9 0.75	0.92 0.92 0.93 0.62 0.74 0.9 0.96 0.87	t5-large 0.93 0.94 0.93 0.64 0.83 0.89 0.94 0.92	0.93 0.95 0.63 0.63 0.89 0.89 0.97	0.93 0.95 0.94 0.68 0.81 0.88 0.96 0.93
All aux_form_dist_Fin aux_mood_dist_Ind dep_dist_compound dep_dist_nummod subord_prop_dist upos_dist_AUX upos_dist_DET upos_dist_NUM	0.45 0.55 0.46 0 0 0.67 0 0 0 0	0.51 0.66 0.63 0 0 0.72 0 -0.011 0	t5-smal 0.66 0.76 0.79 0.14 0 0.75 0.57 0.57 0.33	0.79 0.84 0.86 0.35 0.5 0.81 0.84 0.62 0.76	0.87 0.85 0.89 0.52 0.7 0.85 0.89 0.81 0.9	0.54 0.69 0.72 0 0 0.64 0.17 0.14 0.23	0.78 0.74 0.72 0.16 0.65 0.78 0.77 0.74 0.85	t5-base 0.88 0.9 0.86 0.57 0.8 0.87 0.9 0.84 0.92	0.89 0.91 0.9 0.57 0.8 0.87 0.93 0.84 0.91	0.9 0.94 0.61 0.81 0.85 0.94 0.88 0.91	0.89 0.9 0.92 0.53 0.73 0.86 0.9 0.75 0.89	0.92 0.93 0.62 0.74 0.9 0.96 0.87 0.92	t5-large 0.93 0.94 0.93 0.64 0.83 0.89 0.94 0.92 0.93	0.93 0.95 0.63 0.89 0.89 0.97 0.89 0.94	0.93 0.95 0.94 0.68 0.81 0.88 0.96 0.93 0.94
All aux_form_dist_Fin aux_mood_dist_Ind dep_dist_compound dep_dist_nummod subord_prop_dist upos_dist_AUX upos_dist_DET upos_dist_PRON	0.45 0.55 0.46 0 0 0 0 0 0 0 0 0 0	0.51 0.66 0.63 0 0 0.72 0 -0.011 0 0.11	t5-smal 0.66 0.76 0.79 0.14 0 0.75 0.57 0.33 0.19 0.53	0.79 0.84 0.86 0.35 0.5 0.81 0.84 0.62 0.76 0.66	0.87 0.85 0.52 0.7 0.85 0.89 0.81 0.9 0.83	0.54 0.69 0.72 0 0.0 0.04 0.17 0.14 0.23 0.26	0.78 0.74 0.72 0.16 0.65 0.78 0.77 0.74 0.85 0.84	t5-base 0.88 0.9 0.86 0.57 0.8 0.87 0.9 0.84 0.92 0.92	0.89 0.91 0.57 0.8 0.87 0.93 0.84 0.91 0.92	0.9 0.94 0.9 0.61 0.81 0.85 0.94 0.88 0.91 0.92	0.89 0.92 0.53 0.73 0.86 0.9 0.75 0.89	0.92 0.93 0.62 0.74 0.9 0.96 0.87 0.92 0.93	t5-large 0.93 0.94 0.64 0.83 0.89 0.94 0.92 0.93 0.95	0.93 0.95 0.63 0.63 0.89 0.97 0.89 0.94 0.95	0.93 0.95 0.94 0.68 0.81 0.88 0.96 0.93 0.94
All aux_form_dist_Fin aux_mood_dist_Ind dep_dist_compound dep_dist_nummod subord_prop_dist upos_dist_AUX upos_dist_DET upos_dist_PRON upos_dist_SYM	0.45 0.55 0.46 0 0 0 0 0 0 0 0 0 0 0	0.51 0.66 0.63 0 0 0.72 0 -0.011 0 0.111 0	t5-smal 0.66 0.76 0.79 0.14 0.75 0.57 0.33 0.19 0.53 0	0.79 0.84 0.86 0.35 0.5 0.81 0.84 0.62 0.76 0.66 0	0.87 0.89 0.52 0.7 0.85 0.89 0.81 0.9 0.83 0.83	0.54 0.69 0.72 0 0.64 0.17 0.14 0.23 0.26 0	0.78 0.74 0.72 0.16 0.65 0.78 0.77 0.74 0.85 0.84 0.84	t5-base 0.88 0.9 0.86 0.57 0.8 0.87 0.9 0.84 0.92 0.92 0.9	0.89 0.91 0.9 0.57 0.8 0.87 0.93 0.84 0.91 0.92 0.38	0.9 0.94 0.9 0.61 0.81 0.85 0.94 0.88 0.91 0.92 0.65	0.89 0.92 0.53 0.73 0.86 0.9 0.75 0.89 0.89 0.27	0.92 0.93 0.62 0.74 0.9 0.96 0.87 0.92 0.93	t5-large 0.93 0.94 0.93 0.64 0.83 0.89 0.94 0.92 0.93 0.95 0.8	0.93 0.95 0.63 0.83 0.89 0.97 0.89 0.94 0.95	0.93 0.95 0.94 0.68 0.81 0.88 0.96 0.93 0.94 0.94

Italian

Predicting Complexity with LI Models



Selected Findings

• Informing models linguistically over several epochs allows them to progressively improve their degree of language proficiency.

• The method of linguistic enhancement is particularly effective, especially when applied to smaller models and in scenarios with limited availability of target training data.

• Small models, refined through intermediate fine-tuning, can frequently surpass the performance of larger models that have not undergone this intermediate refinement process.

Thanks for the attention!





Istituto di Linguistica Computazionale "Antonio Zampolli"