Universal Dependencies According to BERT: Both More Specific and More General

Tomasz Limisiewicz, David Mareček, Rudolf Rosa

Goal

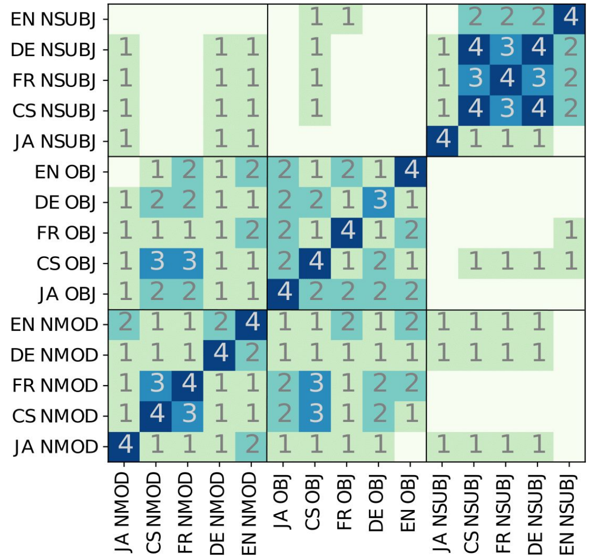
We introduce a ***head ensemble*** method, combining multiple attention heads which capture the same dependency relation label

Dependency Accuracy

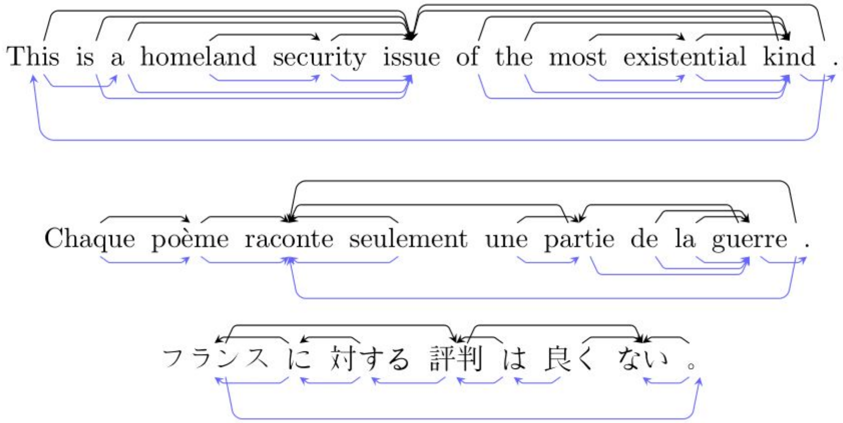
$$DepAcc_{l,d,A} = \frac{|\{(i,j) \in E_{l,d} : j = \arg\max A[i]\}|}{|E_{l,d}|}$$

- $E_{l,d}$ - all directed dependency edges
- A[i] - i^{th} row of the attention matrix

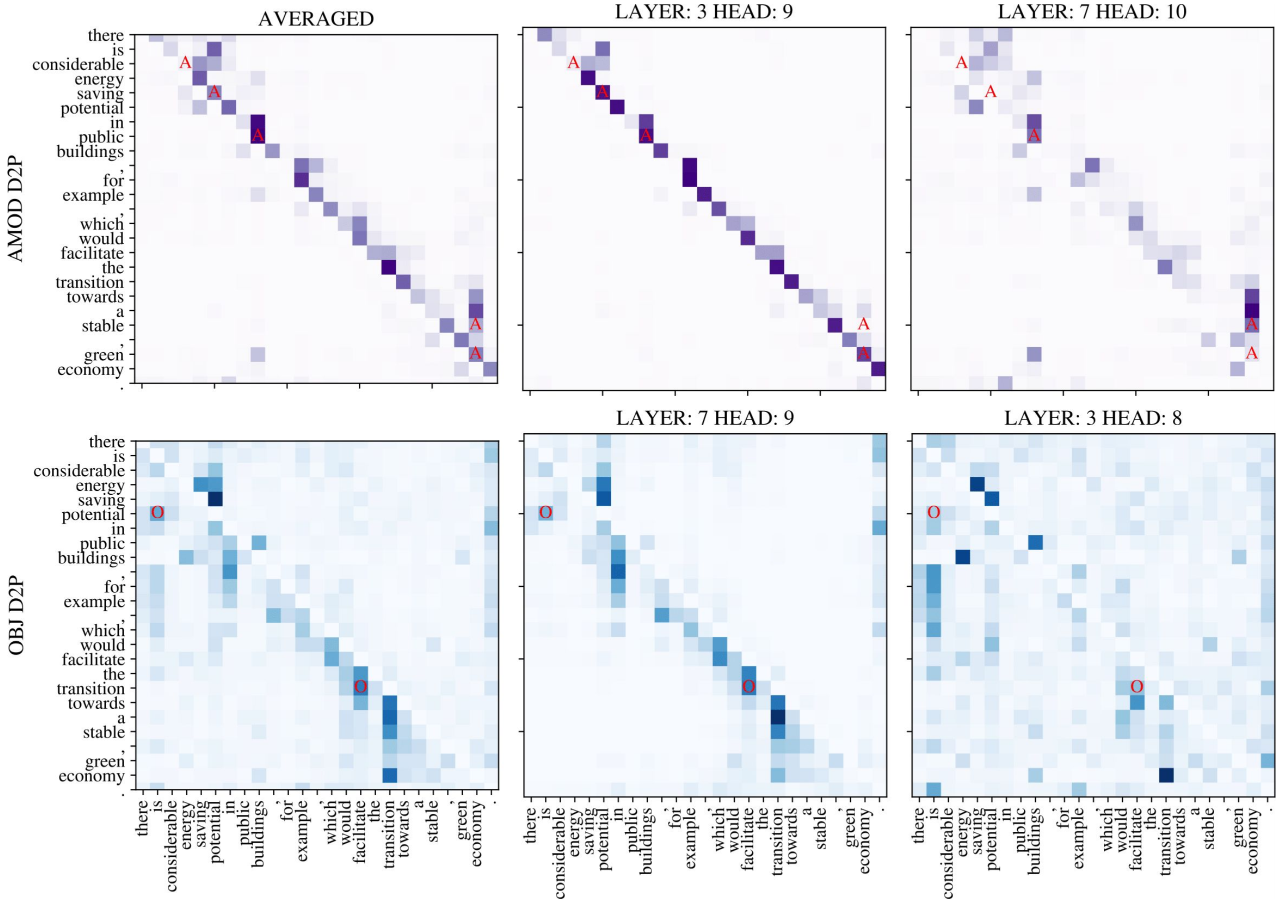
Ensembles Overlap



	JA NMOD	CS NMOD	FR NMOD	DE NMOD	EN NMOD	JA OBJ	CS OBJ	FR OBJ	DE OBJ	EN OBJ	JA NSUBJ	CS NSUBJ	FR NSUBJ	DE NSUBJ	EN NSUBJ
EN NSUBJ							1	1				2	2	2	4
DE NSUBJ	1			1	1		1				1	4	3	4	2
FR NSUBJ	1			1	1		1				1	3	4	3	2
CS NSUBJ	1			1	1		1				1	4	3	4	2
JA NSUBJ	1			1	1						4	1	1	1	
EN OBJ		1	2	1	2	2	1	2	1	4					
DE OBJ	1	2	2	1	1	2	2	1	3	1					
FR OBJ	1	1	1	1	2	2	1	4	1	2					1
CS OBJ	1	3	3	1	1	2	4	1	2	1		1	1	1	1
JA OBJ	1	2	2	1	1	4	2	2	2	2					
EN NMOD	2	1	1	2	4	1	1	2	1	2	1	1	1	1	
DE NMOD	1	1	1	4	2	1	1	1	1	1	1	1	1	1	
FR NMOD	1	3	4	1	1	2	3	1	2	2					
CS NMOD	1	4	3	1	1	2	3	1	2	1					
JA NMOD	4	1	1	1	2	1	1	1	1		1	1	1	1	


This is a homeland security issue of the most existential kind .
Chaque poème raconte seulement une partie de la guerre .
フランス に 対する 評判 は 良く ない 。

extracted trees **edges below**, gold trees **edges above**.



Dependency Tree Extraction

- Trees are extracted from averaged ***head ensembles*** by an ***MST*** algorithm. Similar approach to **(Raganato and Tiedemann, 2018)**
- Extracted trees are directed and labeled

Key Findings

1. Using head ensembles instead of single heads improves:

a.	Average DepAcc:	67.8%→	74.1%
b.	UAS:	37.2%→	52.0%
c.	LAS:	N/A	21.7%

2. We have observed many-to-many relationship between heads and syntactic functions
3. The method is effective for 9 typologically diverse languages