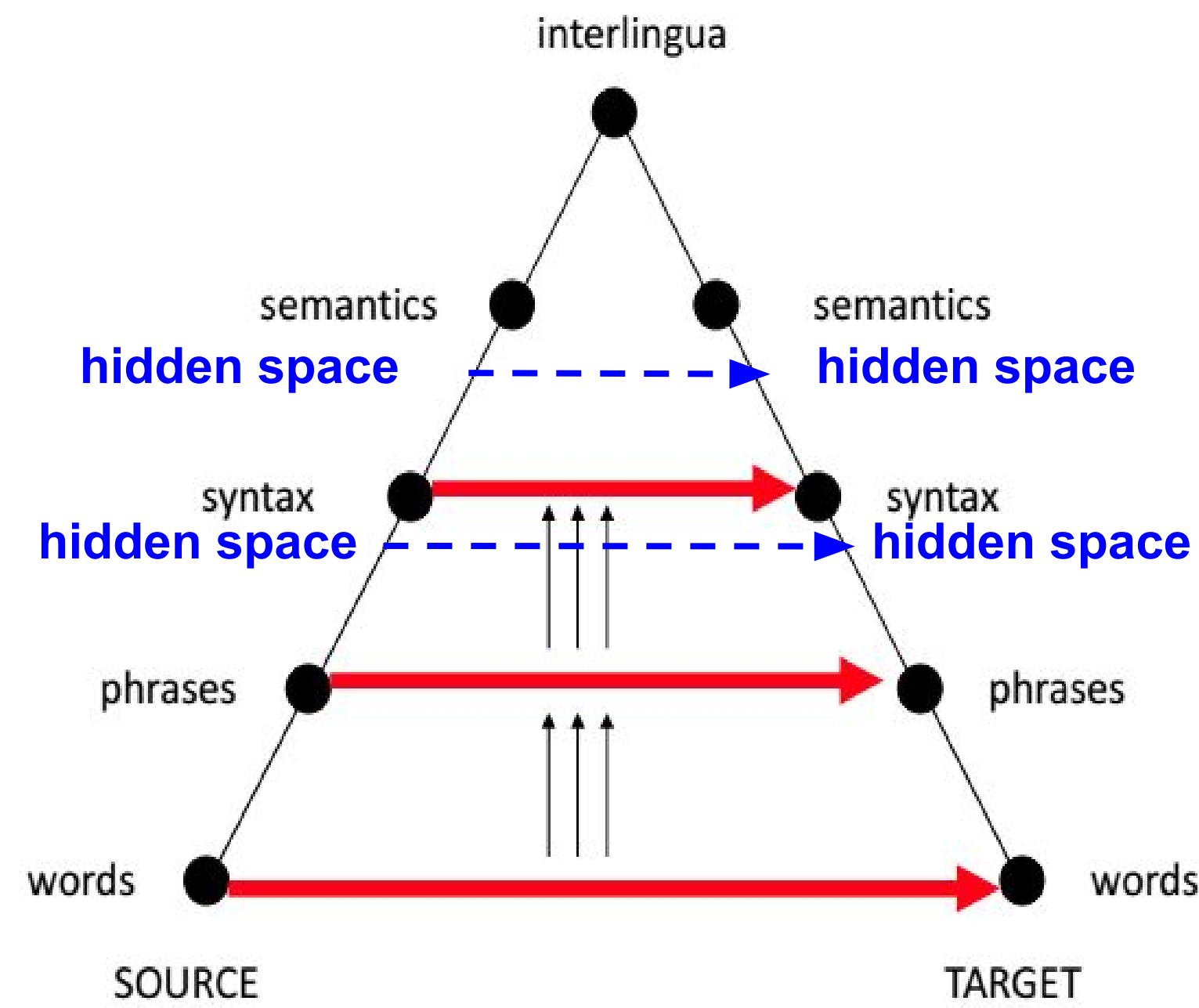


Motivation



Where is Neural Machine Translation ?

Question

1. Whether does encoder learn syntactic information of source sentence ?
2. What kind of syntactic information does the encoder learn ?
3. How much syntactic information does the encoder learn ?

Conclusion

1. NMT encoder learns quite a lot syntactic information, but still lost subtle details.
2. Not too much difference between E2F and E2G
3. NMT encoder is better than Autoencoder if you want to convert sentence into vector.

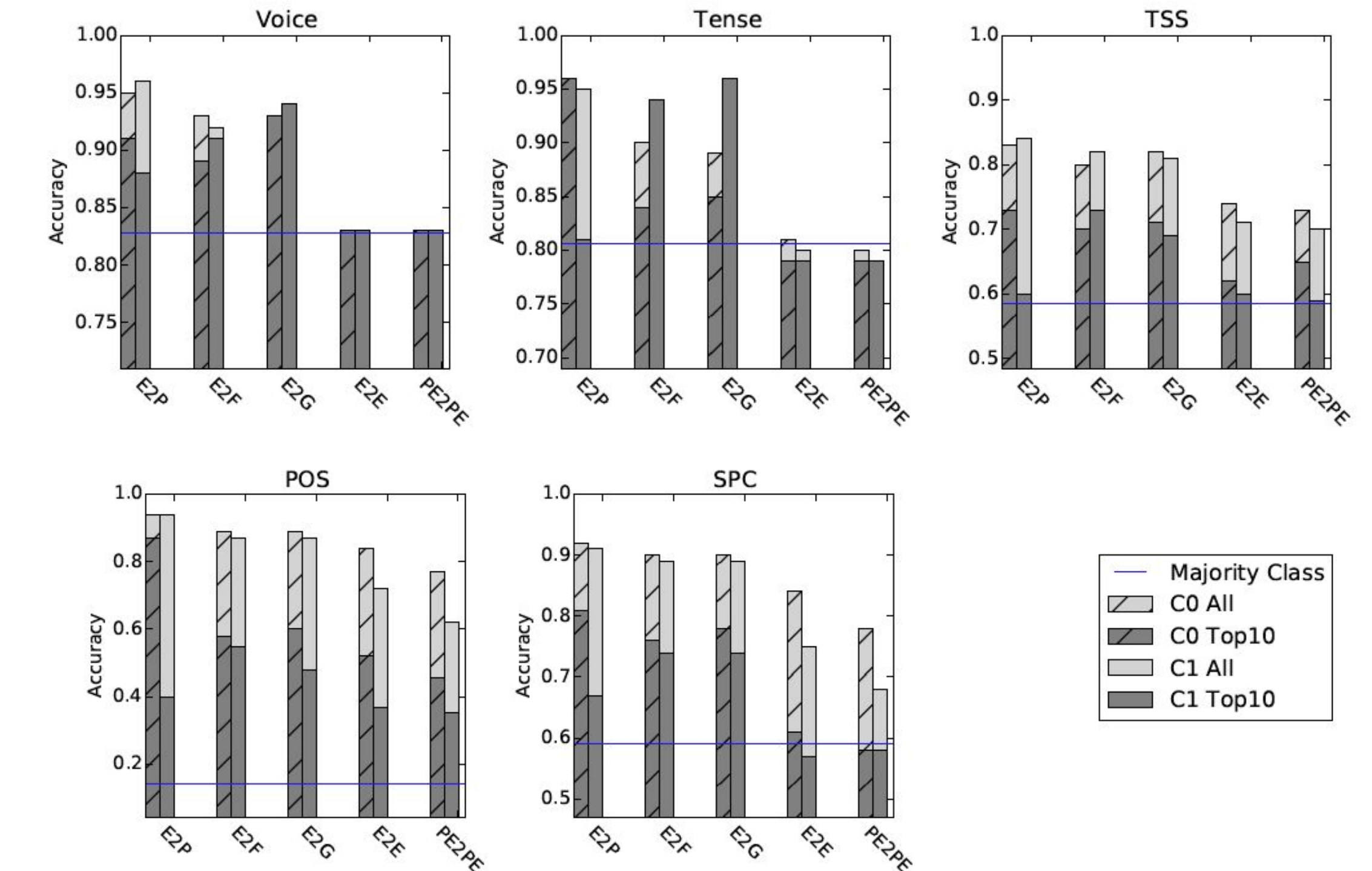
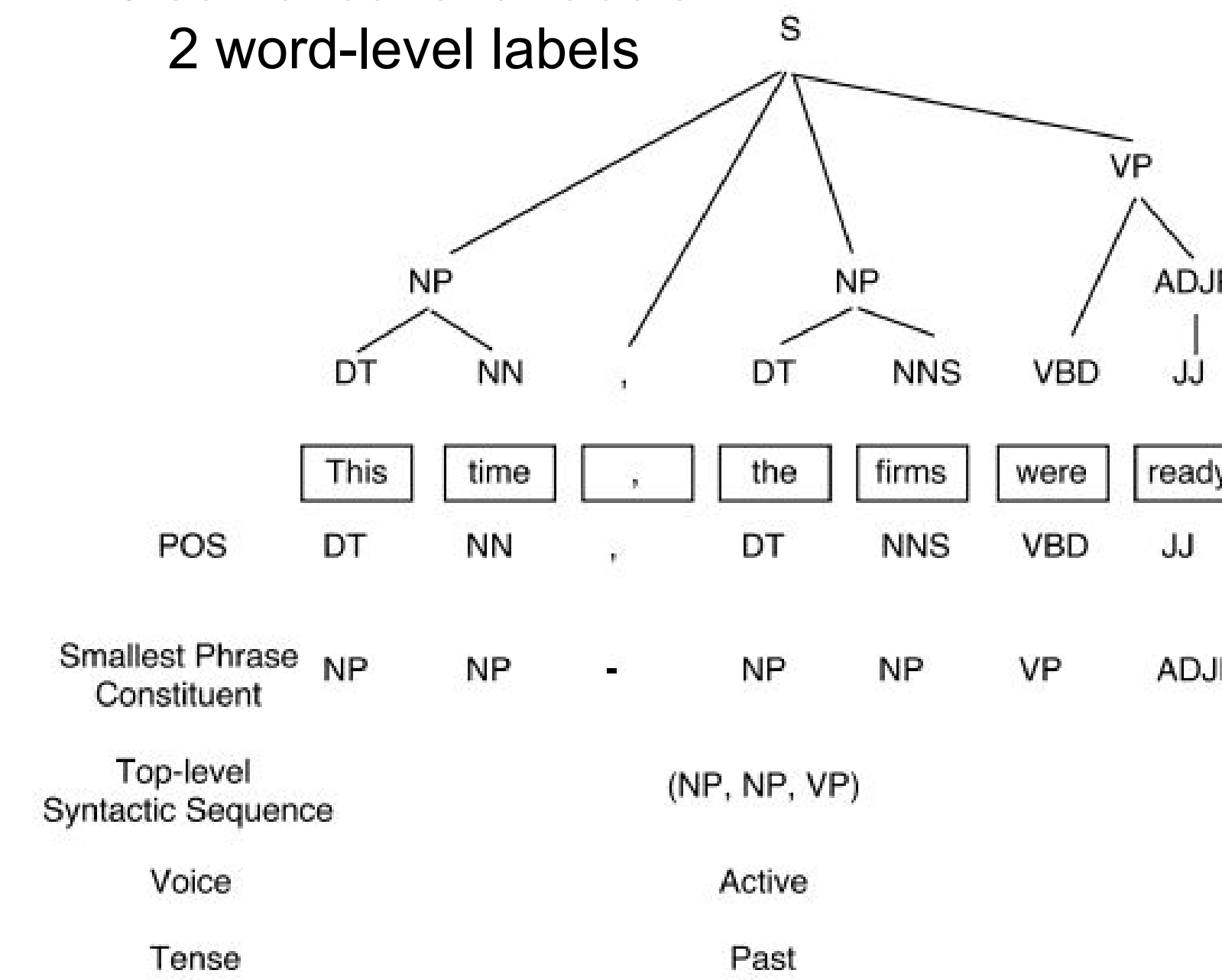
Models

Lower bound Models: E2E PE2PE
 Translation Models: E2F E2G
 Upper bound models: E2P

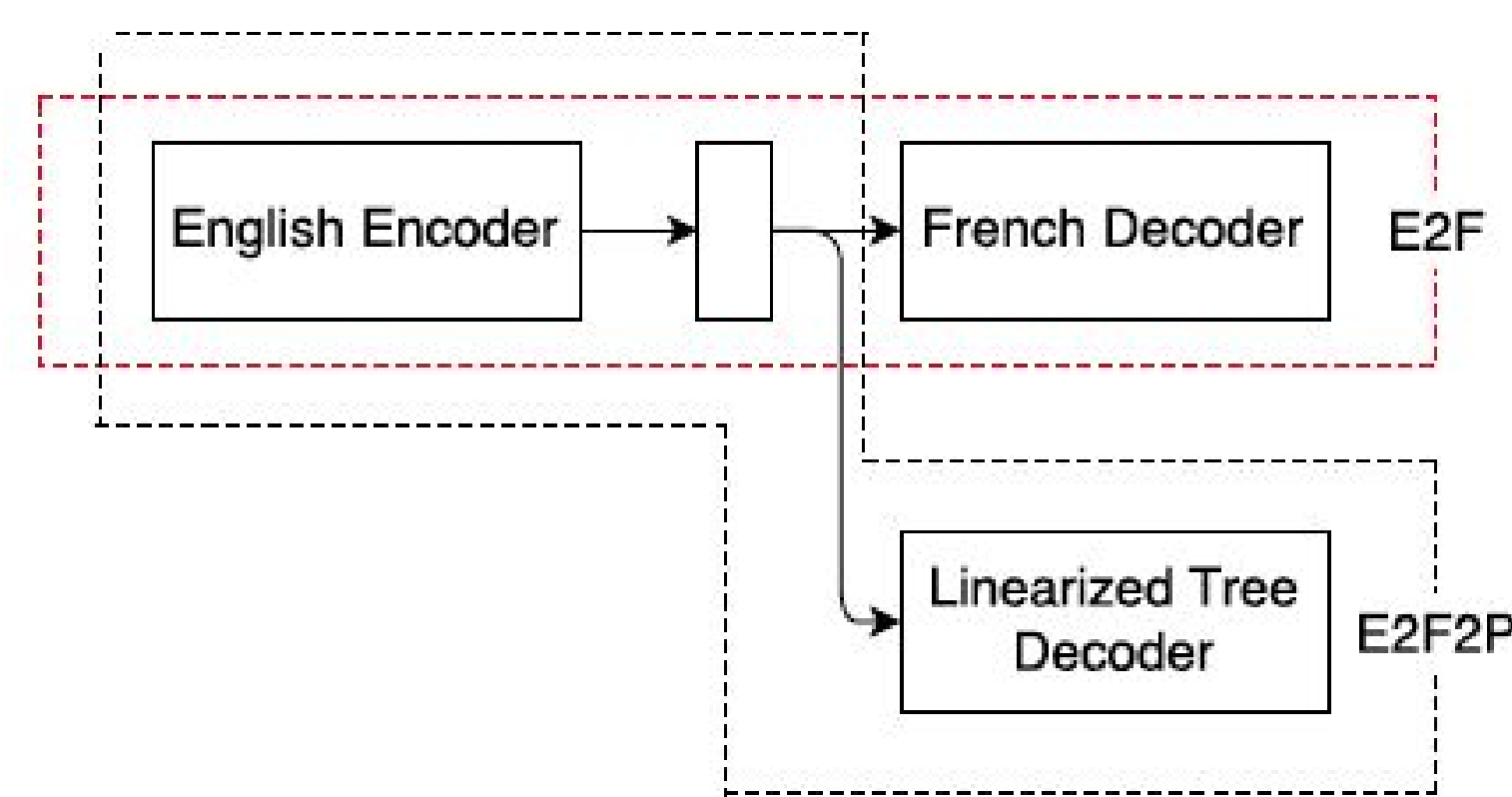
Model	Source	Target
E2E	I like it .	I like it .
PE2PE	it I . like	it I . like
E2F	I like it .	J'aime ça.
E2G	I like it .	Ich mag das.
E2P	I like it .	(S (NP PRP) _{NP} (VP VBP (NP PRP) _{NP})VP .) _S

Method 1 : Syntactic labels = logistic-regression(hidden states)

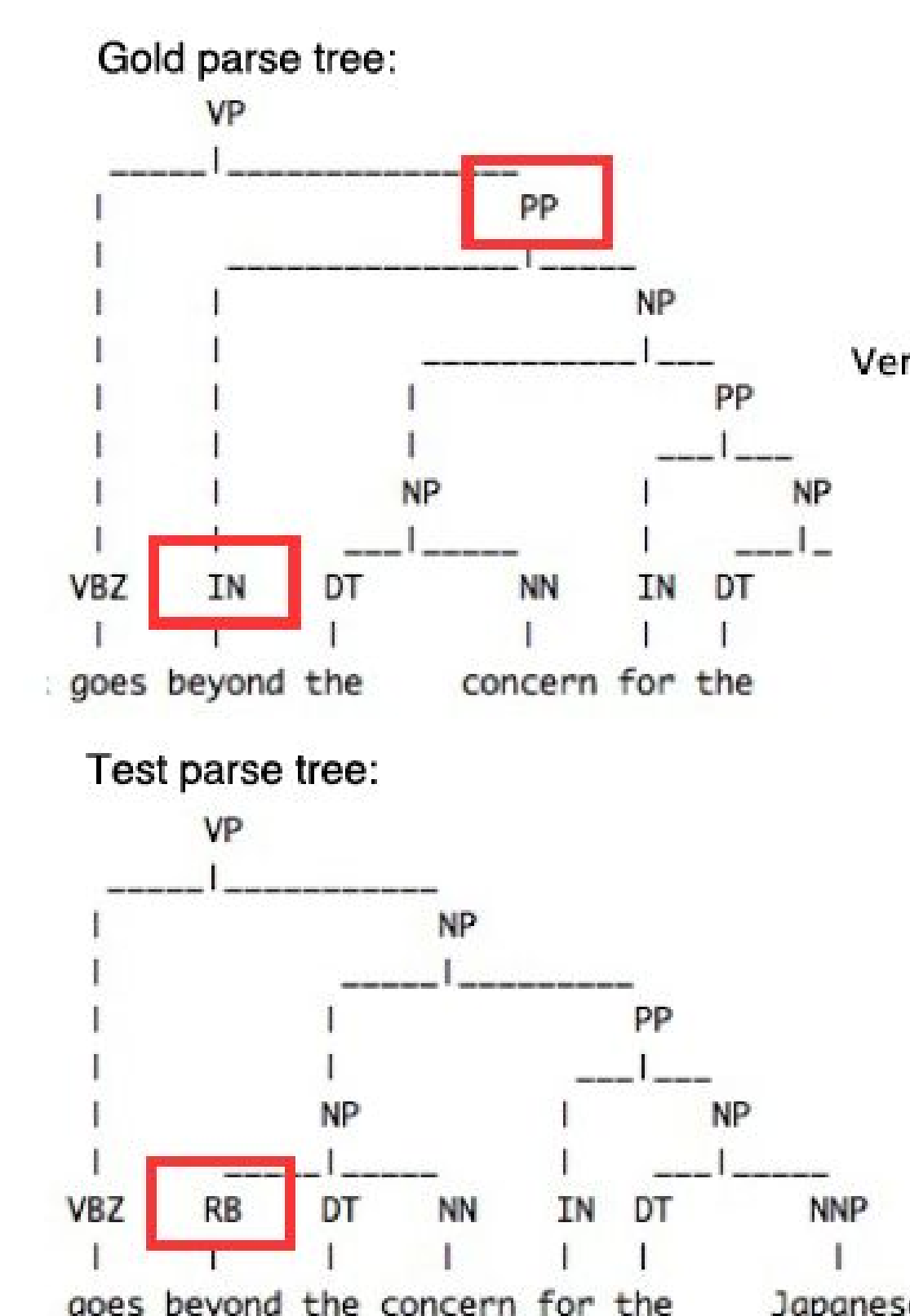
3 sentence-level labels
 2 word-level labels



Method 2 : Extract whole parse tree from NMT encoder



Model	Labeled F1	POS Tagging Accuracy
PE2PE2P	58.67	54.32
E2E2P	70.91	68.03
E2G2P	85.36	85.30
E2F2P	86.62	87.09
E2P	93.76	96.00



	E2P (Ave. Bracket Err)	E2F2P (Ratio)	E2G2P (Ratio)	E2E2P (Ratio)	PE2PE2P (Ratio)
Co-ordination	0.081	1.14	1.05	1.78	0.22
VP Attach	0.024	1.36	1.27	4.55	5.64
PP Attach	0.242	1.44	1.27	1.82	2.27
Verb taking wrong arguments	0.035	1.44	2.75	3.50	2.44
Modifier Attach	0.205	1.46	2.03	1.69	1.82
Unary	0.123	1.98	2.25	3.21	3.71
NP Internal	0.053	2.17	1.83	3.17	3.58
Noun boundary error	0.022	2.20	3.20	3.10	5.10
Different label	0.137	2.52	2.42	5.00	5.26
Single Word Phrase	0.150	2.74	3.31	5.01	5.12
Sense Confusion	0.057	16.58	17.77	23.77	32.19

Analysis done by Berkeley Parser Analyzer (Kummerfeld et al., 2012)