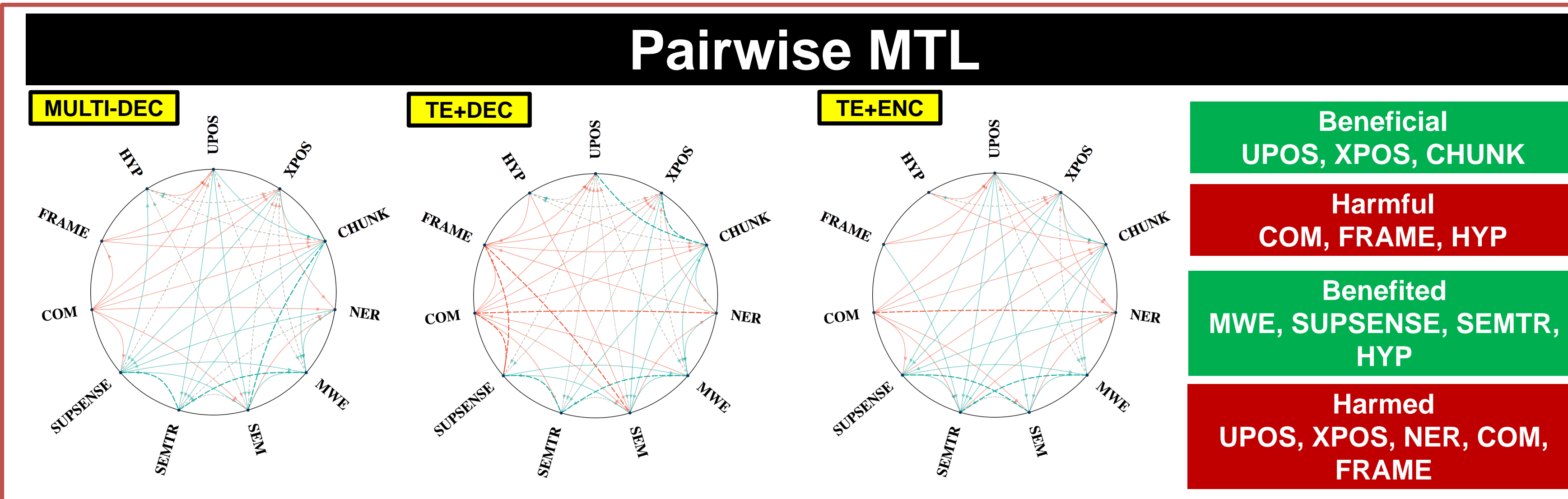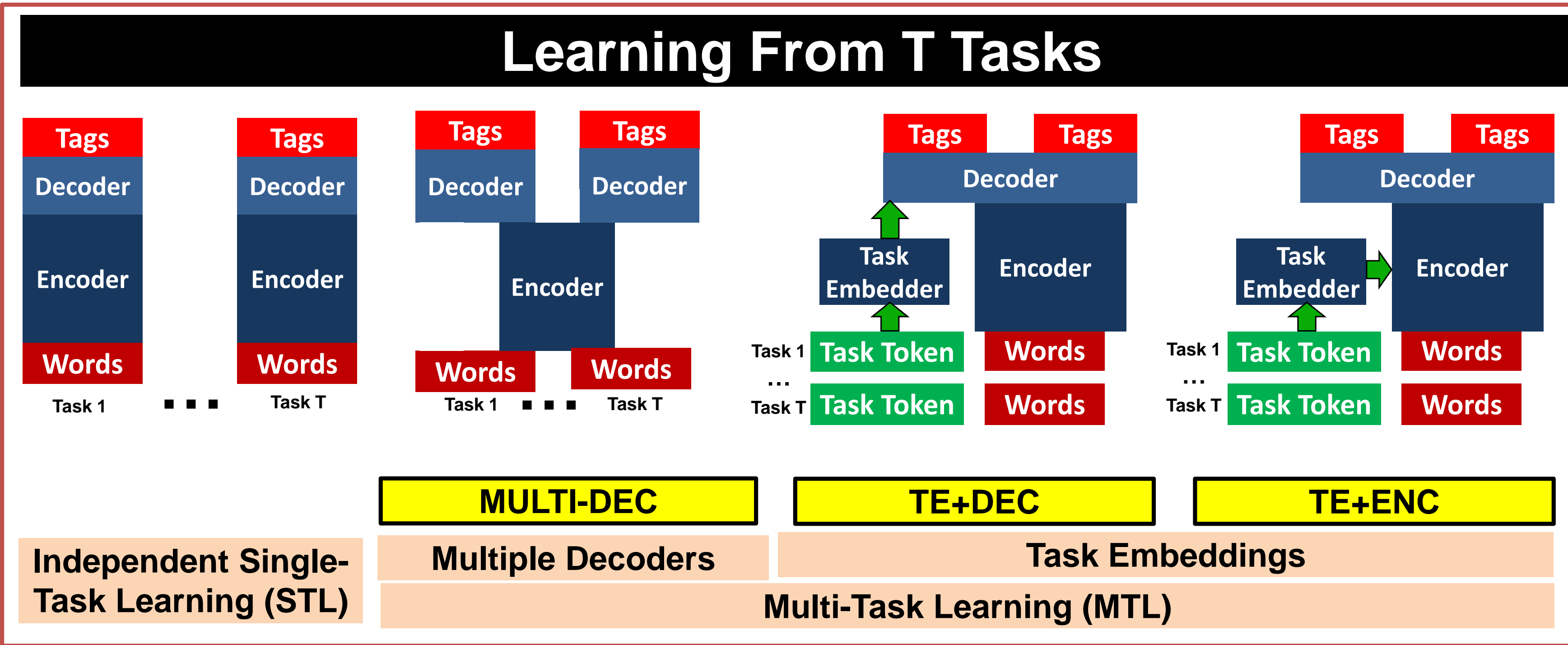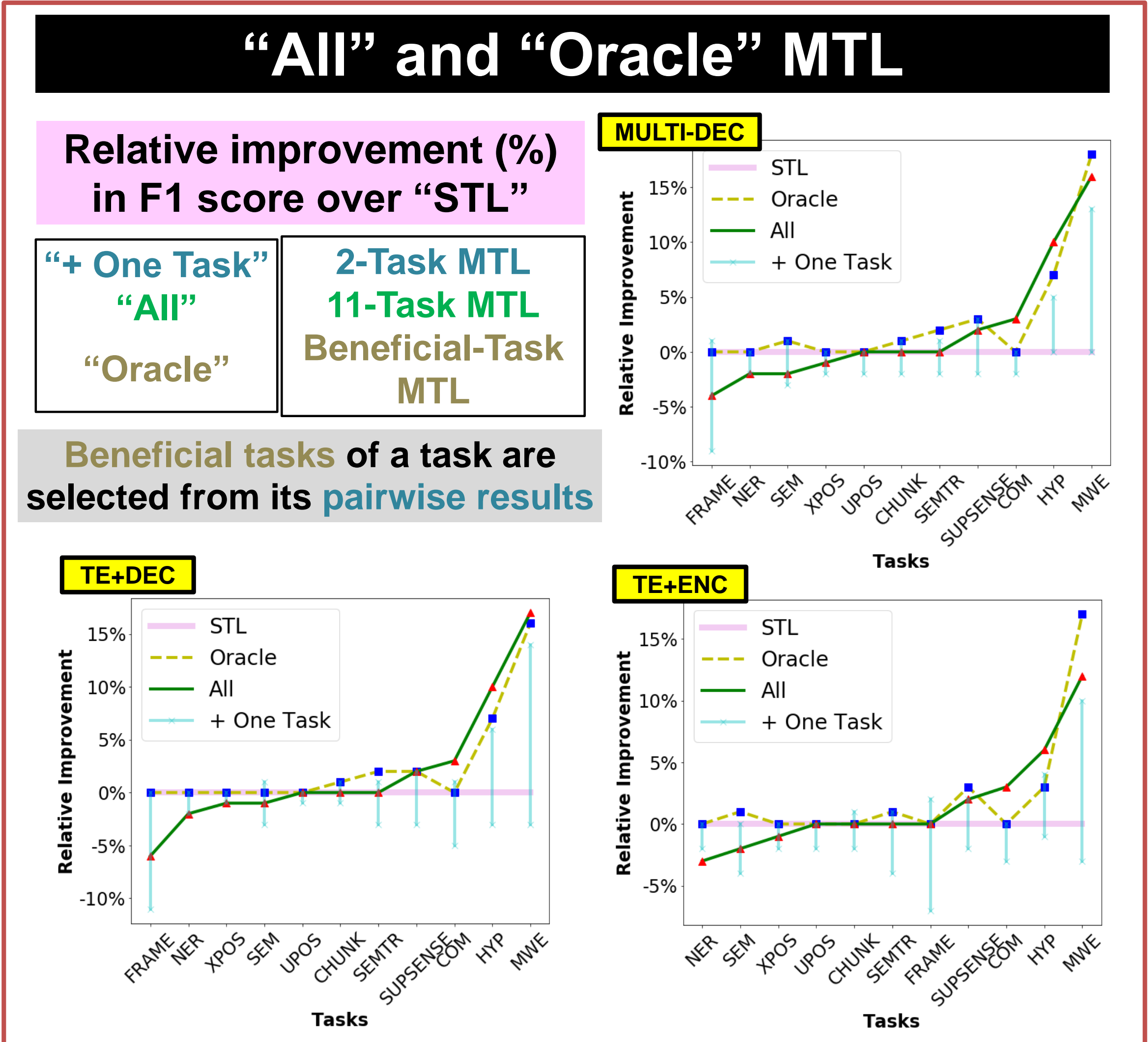# Multi-Task Learning for Sequence Tagging: An Empirical Study

Soravit Changpinyo, Hexiang Hu, and Fei Sha
University of Southern California (USC)

Tasks

UPOS	Universal Part of Speech Tagging
XPOS	English Part of Speech Tagging
Chunk	Syntactic Chunking / Shallow Parsing
NER	Named Entity Recognition
MWE	Multi-Word Expression Identification
SUPSENSE	Streusel Supersense Tagging
SEM	SemCor Supersense Tagging
SEMTR	Semantic Trait Identification
COM	Sentence Compression
FRAME	Frame Target Identification
HYP	Hyperlink Prediction

Datasets & Examples

Dataset	Example
Universal Dependencies v1.4	once again , thank you all for an outstanding accomplishment . ADV ADV PUNCT VERB PRON DET ADP DET ADJ NOUN PUNCT
CoNLL-2000	once again , thank you all for an outstanding accomplishment . RB RB , VBP PRP DT IN DT JJ NN .
	the carrier also seemed eager to place blame on its american counterparts . B-NP E-NP S-ADVP S-VP S-ADJP B-VP E-VP S-NP S-PP B-NP I-NP E-NP O
CoNLL-2003	6. pier francesco chili (italy) ducati 17541 O B-PER I-PER E-PER O S-LOC O S-ORG O
Streusel 4.0	had to keep in mind that the a / c broke , i feel bad it was their opening ! B I B I I O O B I I O O O O O O O O O O
	this place may have been something sometime ; but it way past it " sell by date " . O n.GROUP O O v.stative O O O O O O p.Time p.Gestalt O v.possession p.Time n.TIME O O
SemCor	a hypothetical example will illustrate this point . O adj.all noun.cognition O verb.communication O noun.communication O
	he wondered if the audience would let him finish . O Mental O O Object O Agentive O BoundedEvent O
Broadcast News 1	he made the decisions in 1995 , in early 1996 , to spend at a very high rate . KP KP DEL KP DEL DEL DEL DEL DEL DEL KP KP KP KP DEL KP KP KP
FrameNet 1.5	please continue our important partnership . O B-TARGET O B-TARGET O O
Hyper-Text Corpus	will this incident lead to a further separation of civilizations ? O O O O O O O O B-HTML B-HTML B-HTML O

Learning From T Tasks

MULTI-DEC	TE+DEC	TE+ENC
Multiple Decoders	Task Embeddings	

Independent Single-Task Learning (STL)

Multi-Task Learning (MTL)

Pairwise MTL

MULTI-DEC   TE+DEC   TE+ENC

Beneficial
UPOS, XPOS, CHUNK

Harmful
COM, FRAME, HYP

Benefited
MWE, SUPSENSE, SEMTR, HYP

Harmed
UPOS, XPOS, NER, COM, FRAME

"All" and "Oracle" MTL

Relative improvement (%) in F1 score over "STL"

"+ One Task"	2-Task MTL
"All"	11-Task MTL
"Oracle"	Beneficial-Task MTL

Beneficial tasks of a task are selected from its pairwise results

MULTI-DEC — STL, Oracle, All, + One Task

TE+DEC — STL, Oracle, All, + One Task

TE+ENC — STL, Oracle, All, + One Task

Main and Interesting Results

- There are no significant differences among MTL approaches.
- Pairwise MTL results are *generally* correlated to MTL results with more than two tasks.
- Selecting "Oracle" with our "greedy" strategy works well.
- Results on COM are surprising; All pairwise MTL results are bad (Oracle = STL) but All MTL result is good.
- Task embeddings in TE+DEC reveal clustering syntactic & semantic tasks.

TE+DEC