SemEval 2014 Task-3

Cross-Level Semantic Similarity

David Jurgens
Mohammad Taher Pilehvar*
Roberto Navigli

SAPIENZA
UNIVERSITÀ DI ROMA

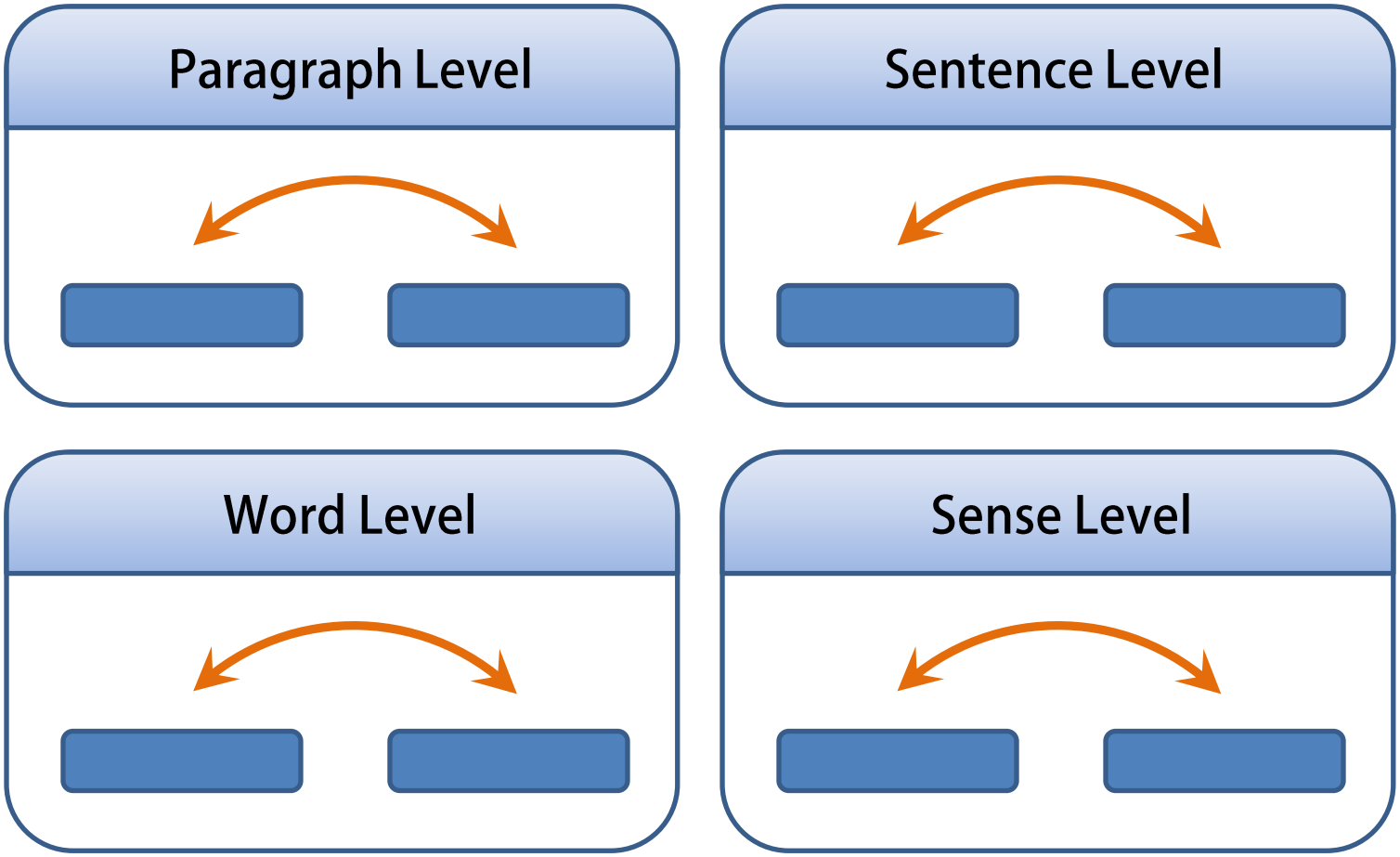
MultiJEDI
ERC 259234

Semantic Similarity

how similar are a pair of lexical items?

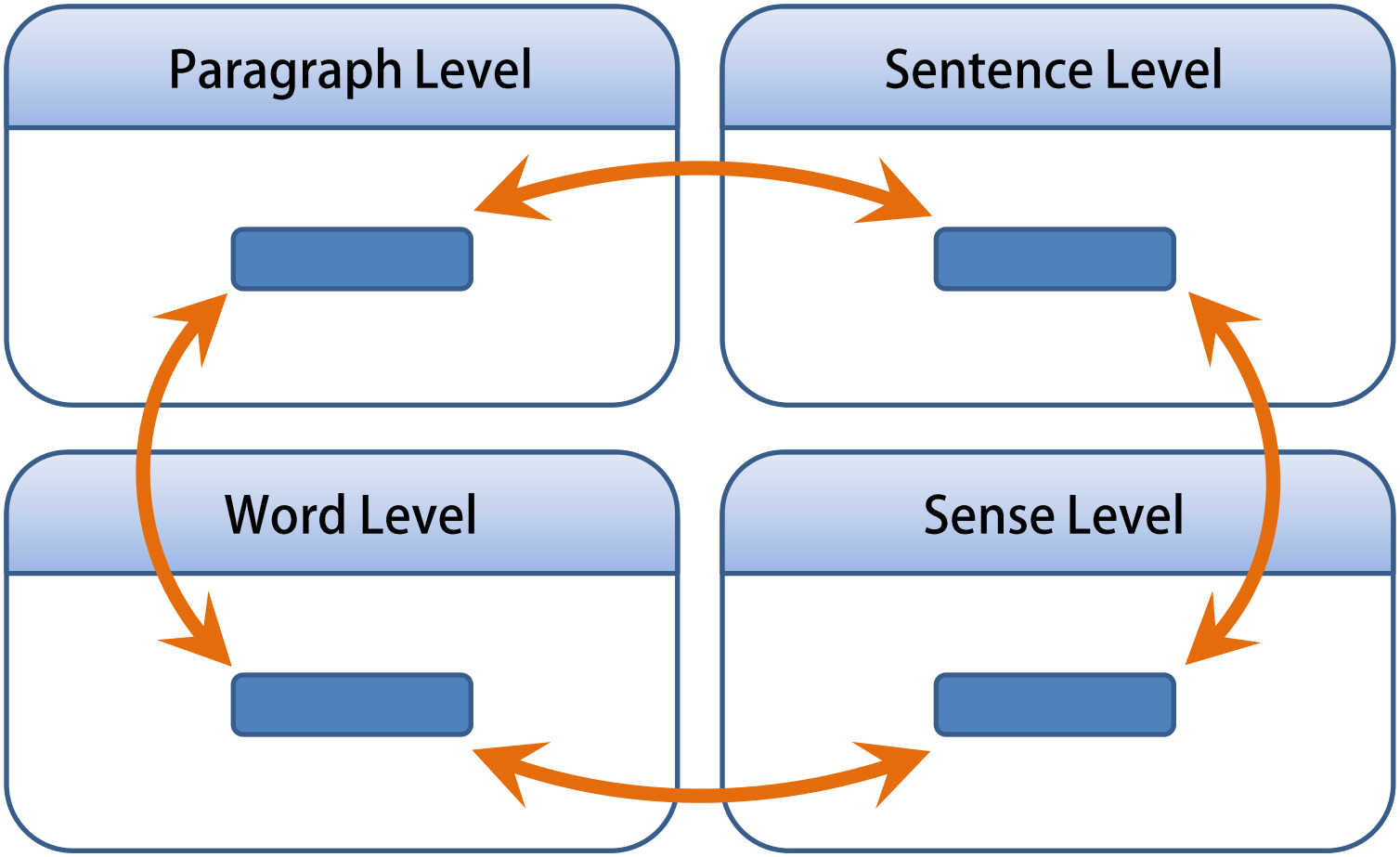
Semantic Similarity

Mostly focused on similar types of lexical items



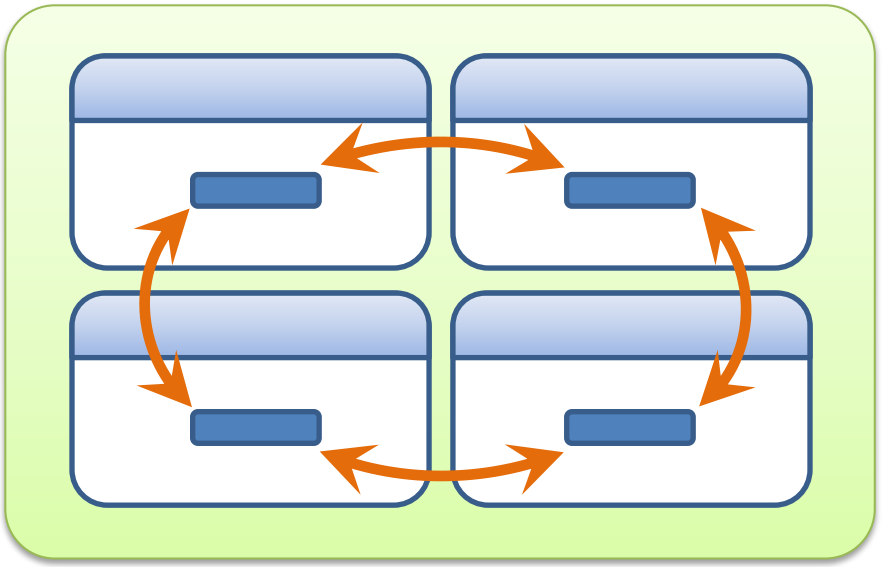
Semantic Similarity

What if we have different types of inputs?



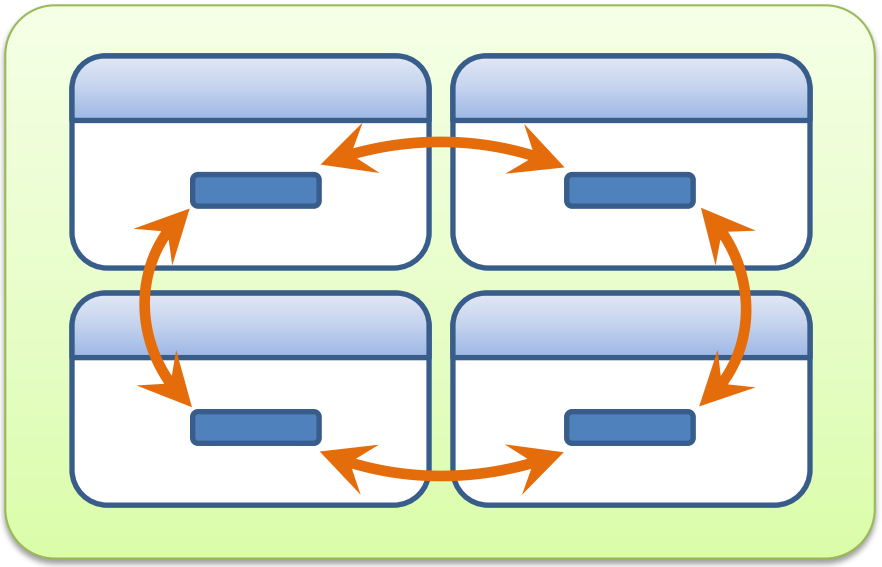
CLSS: Cross-Level Semantic Similarity

A new type of similarity task



CLSS: Cross-Level Semantic Similarity

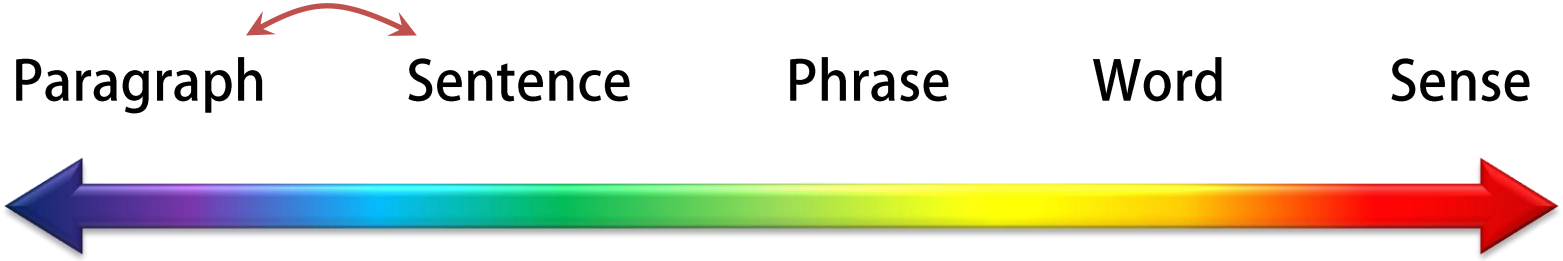
A new type of similarity task



- Multiple types of comparison
- Incorporate multiple genres of text
- Push towards computing the similarity of anything

CLSS: Comparison Types

Paragraph to Sentence



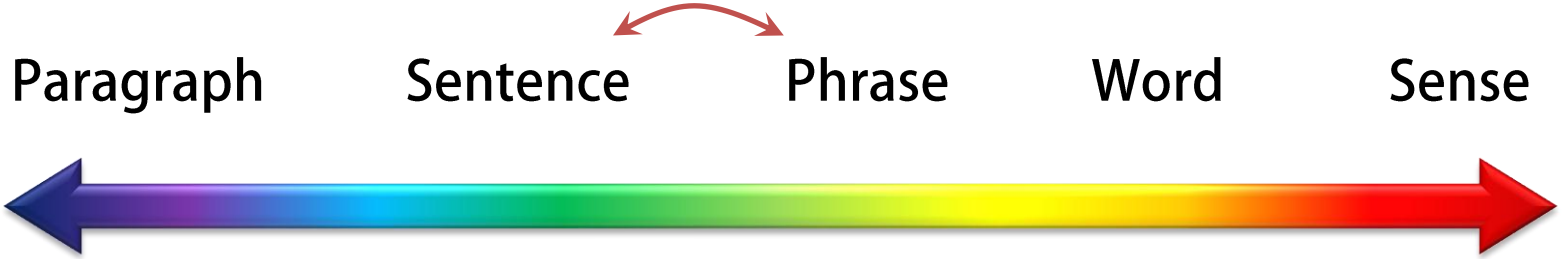
CLSS: Comparison Types

Paragraph to Sentence

Sentence to Phrase

The 30-year-old woman has had no contact with the outside world.

30-year-old female recluse



CLSS: Comparison Types

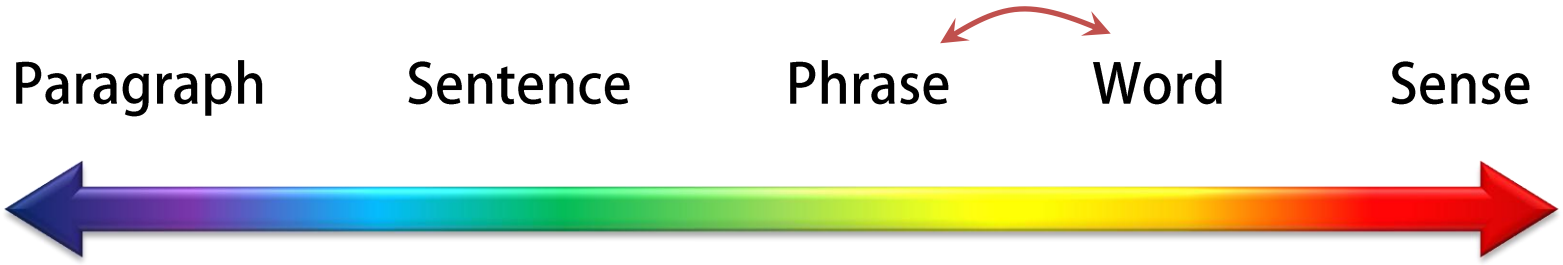
Paragraph to Sentence

Sentence to Phrase

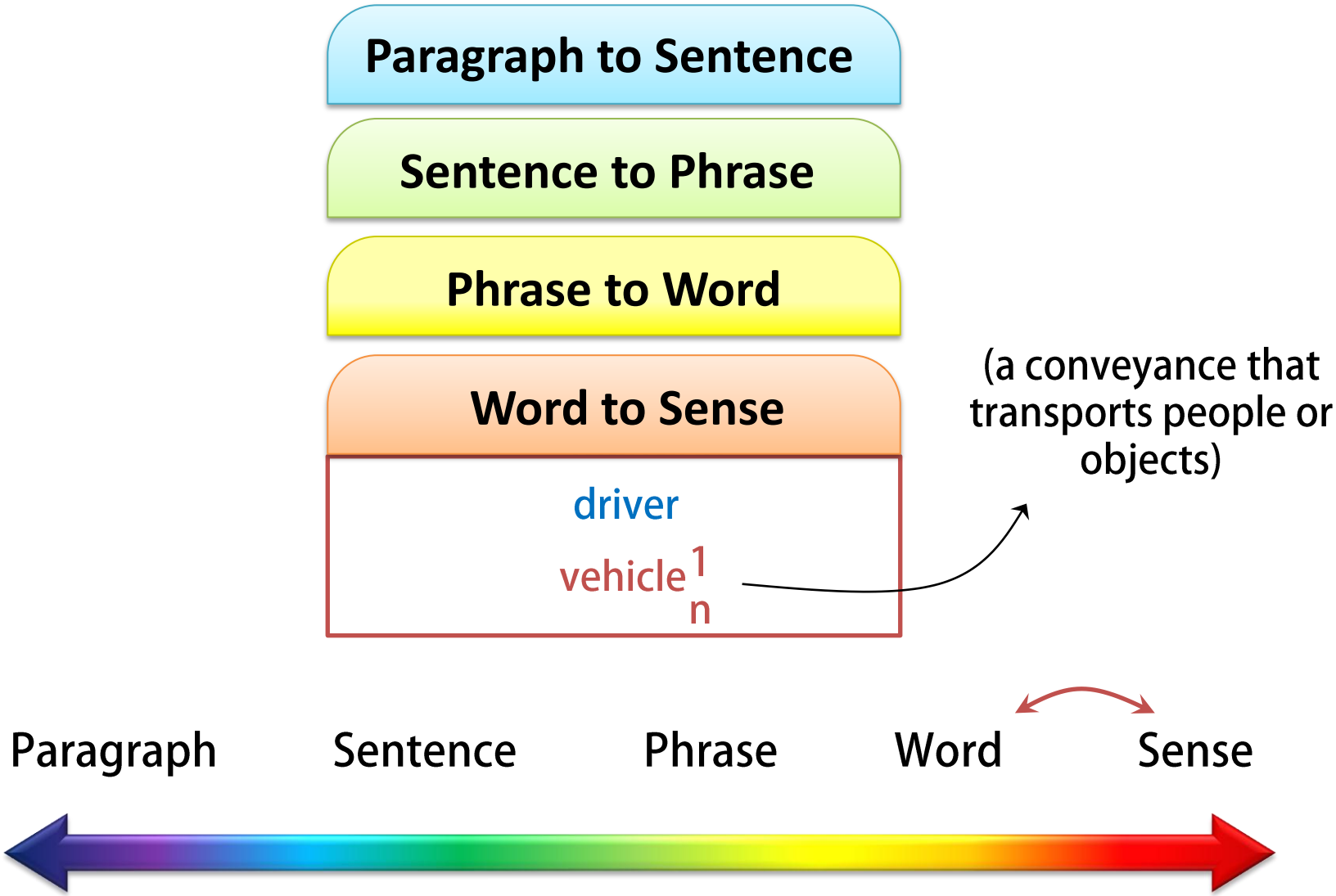
Phrase to Word

a large, expensive house

mansion

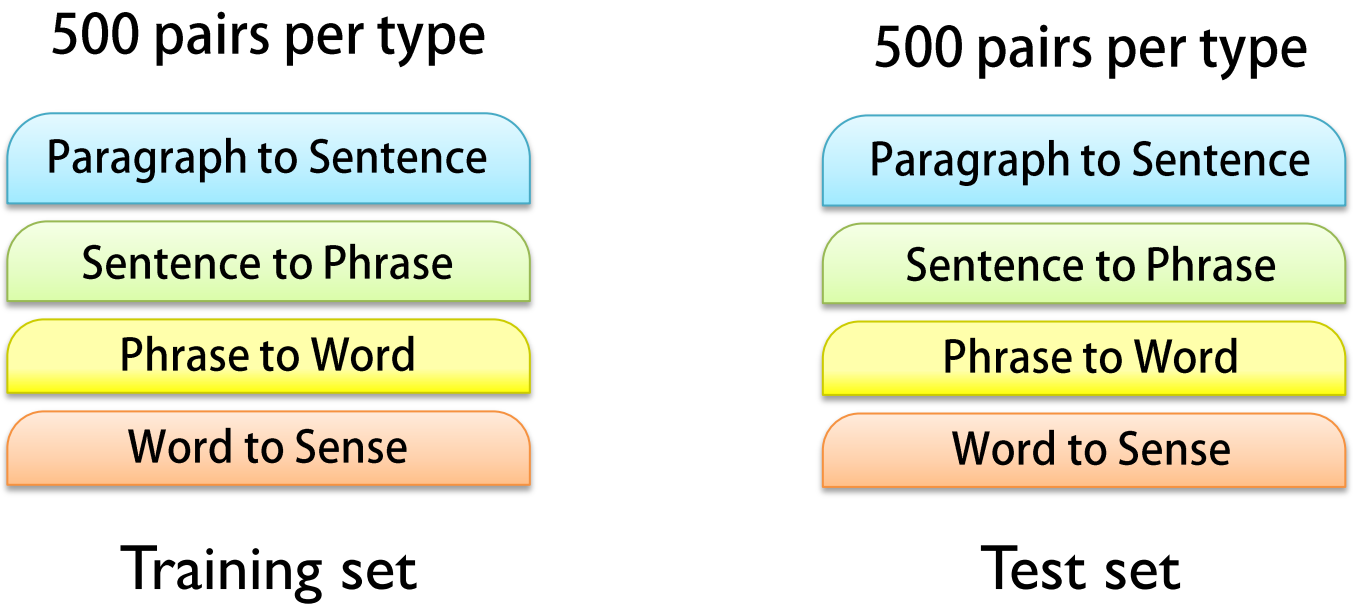


CLSS: Comparison Types



Task Data

4000 pairs in total

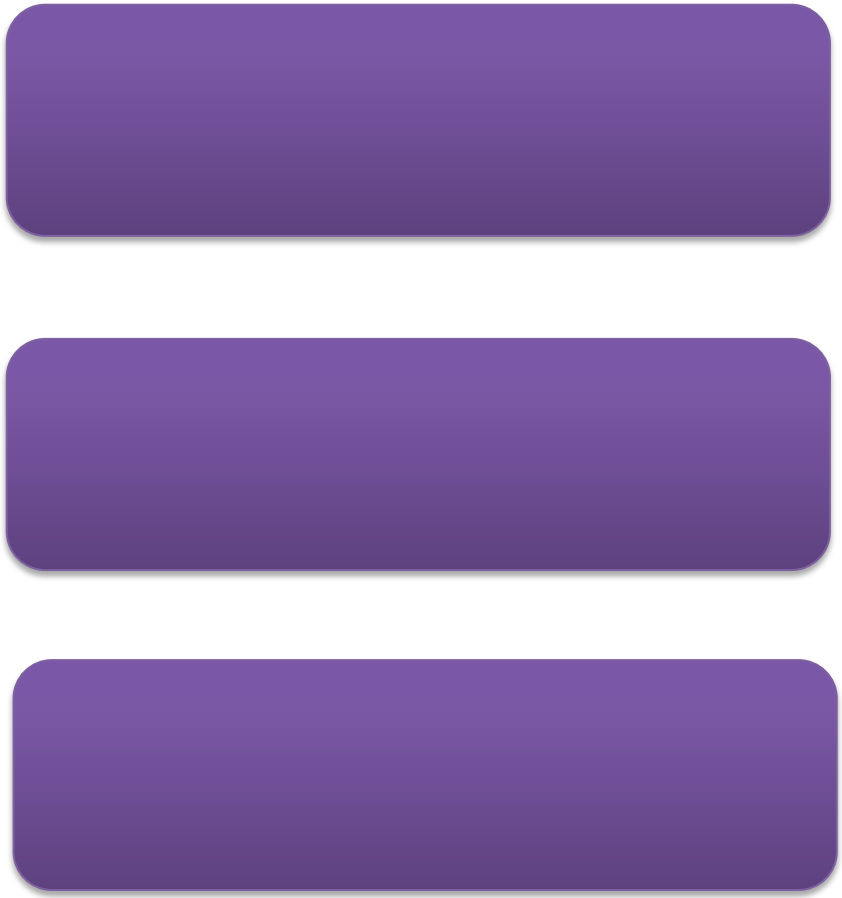


Task Data

A wide range of domains and text styles

Paragraph to Sentence	Sentence to Phrase	Phrase to Word
Newswire	Newswire	Newswire
Travel	Travel	Slang
Question Answering	Question Answering	Idiomatic
Scientific	Scientific	Lexicographic
Metaphoric	Slang	Descriptive
Review	Idiomatic	Search

Word to Sense pairs



Word to Sense pairs

"Regular"

"central" *vs.* essential#a#1

"tyre" *vs.* automobile#n#1

Word to Sense pairs

"Regular"

"central" *vs.* essential#a#1

"tyre" *vs.* automobile#n#1

Word not in WordNet

"zombify" *vs.* resurrect#v#3

"drank" *vs.* opiate#n#1

Word to Sense pairs

"Regular"	"central" *vs.* essential#a#1 "tyre" *vs.* automobile#n#1
Word not in WordNet	"zombify" *vs.* resurrect#v#3 "drank" *vs.* opiate#n#1
Sense not in WordNet	"red" *vs.* communist#a#1 "shiraz" *vs.* grape#n#1

Rating Scale

4 -- *Nearly* identical

3 -- Similar, but not identical

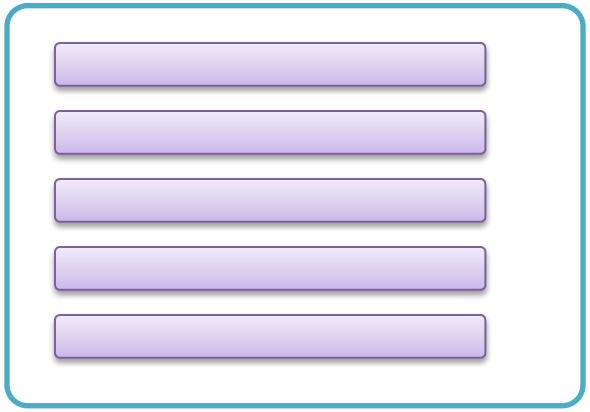
2 -- Related but not similar

1 -- On the same topic, but not closely related

0 -- Completely unrelated

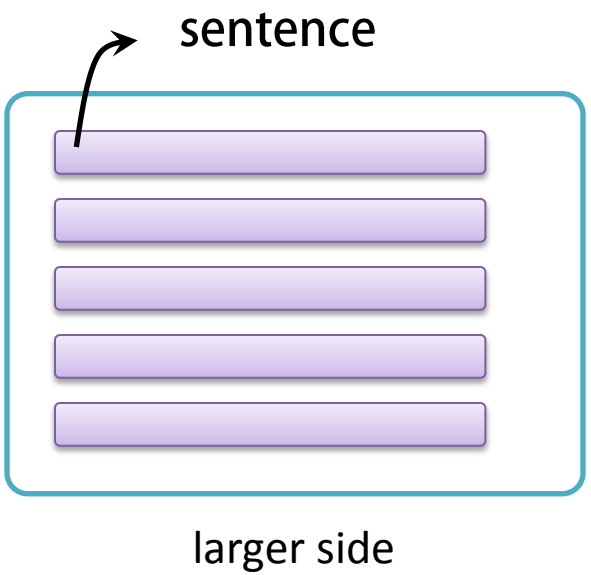
Crafting an idealized similarity distribution

Crafting an idealized similarity distribution

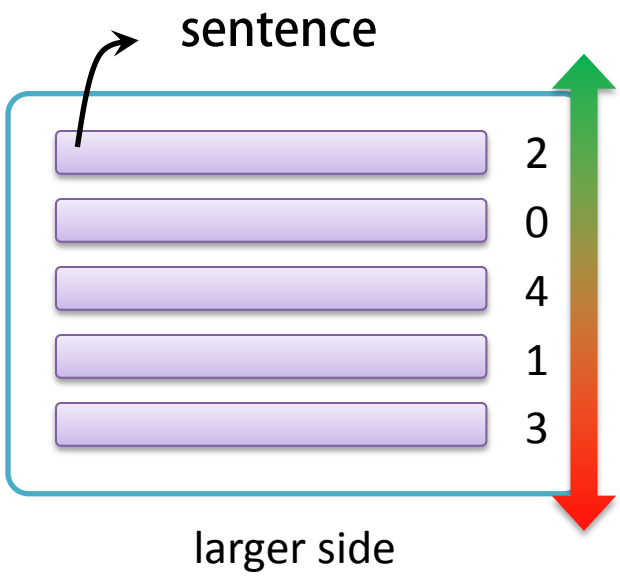


larger side

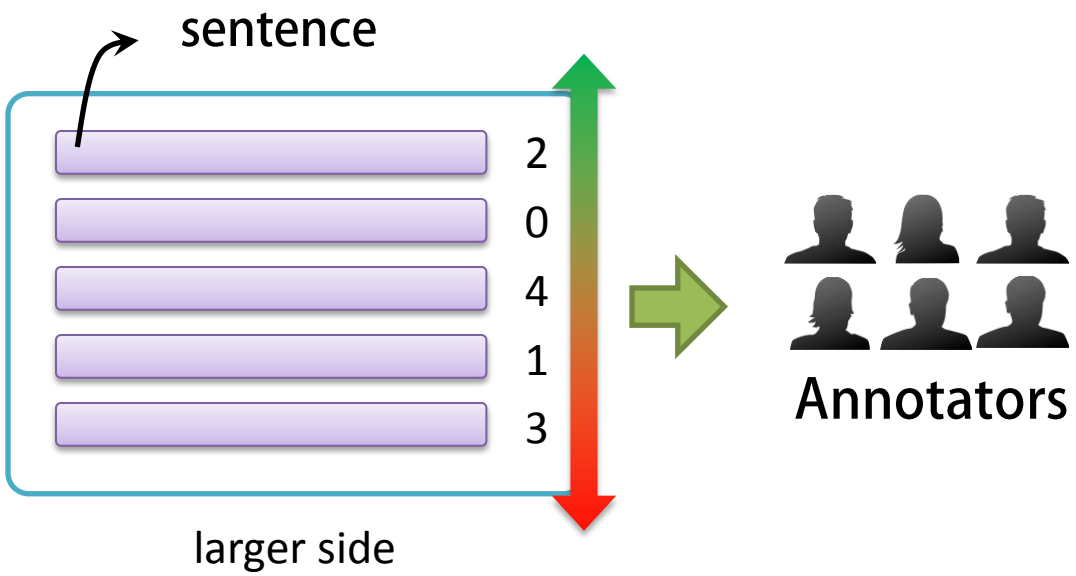
Crafting an idealized similarity distribution



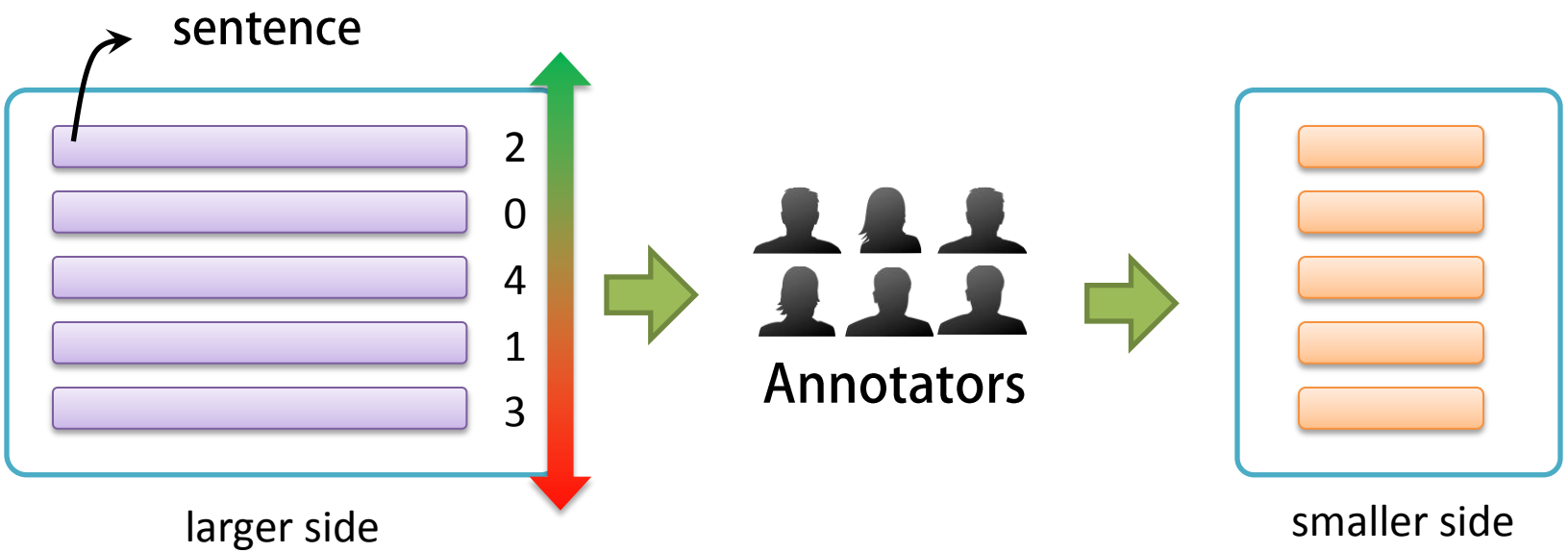
Crafting an idealized similarity distribution



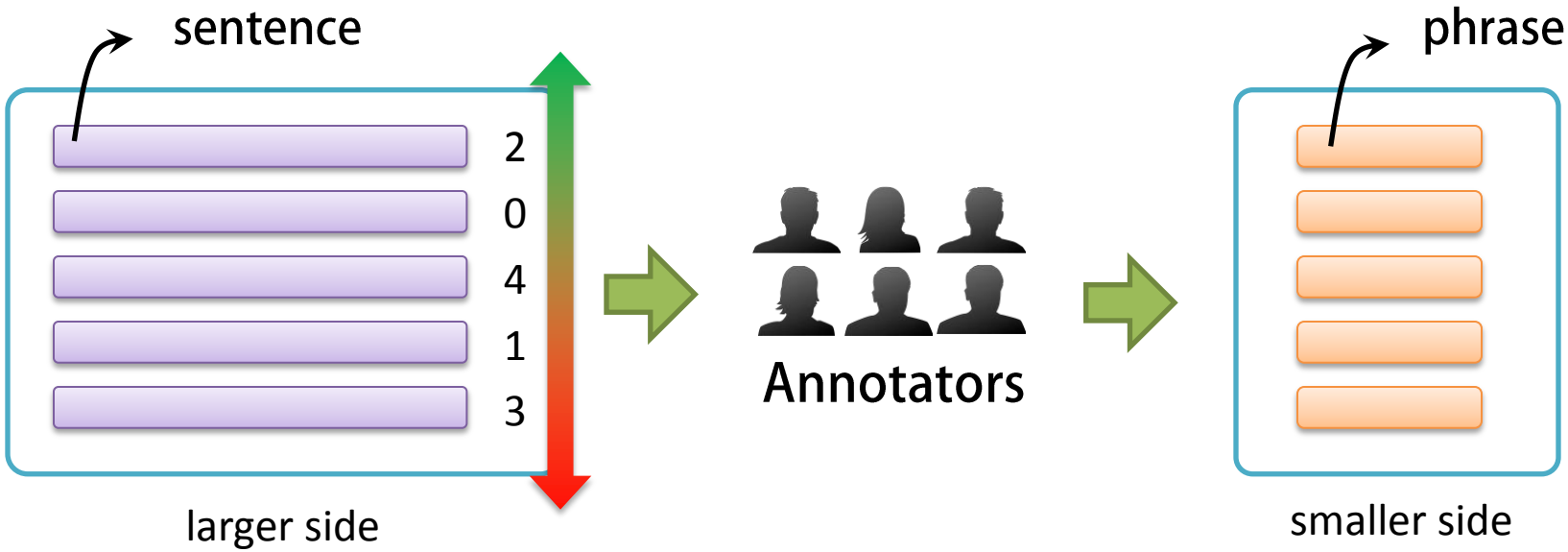
Crafting an idealized similarity distribution



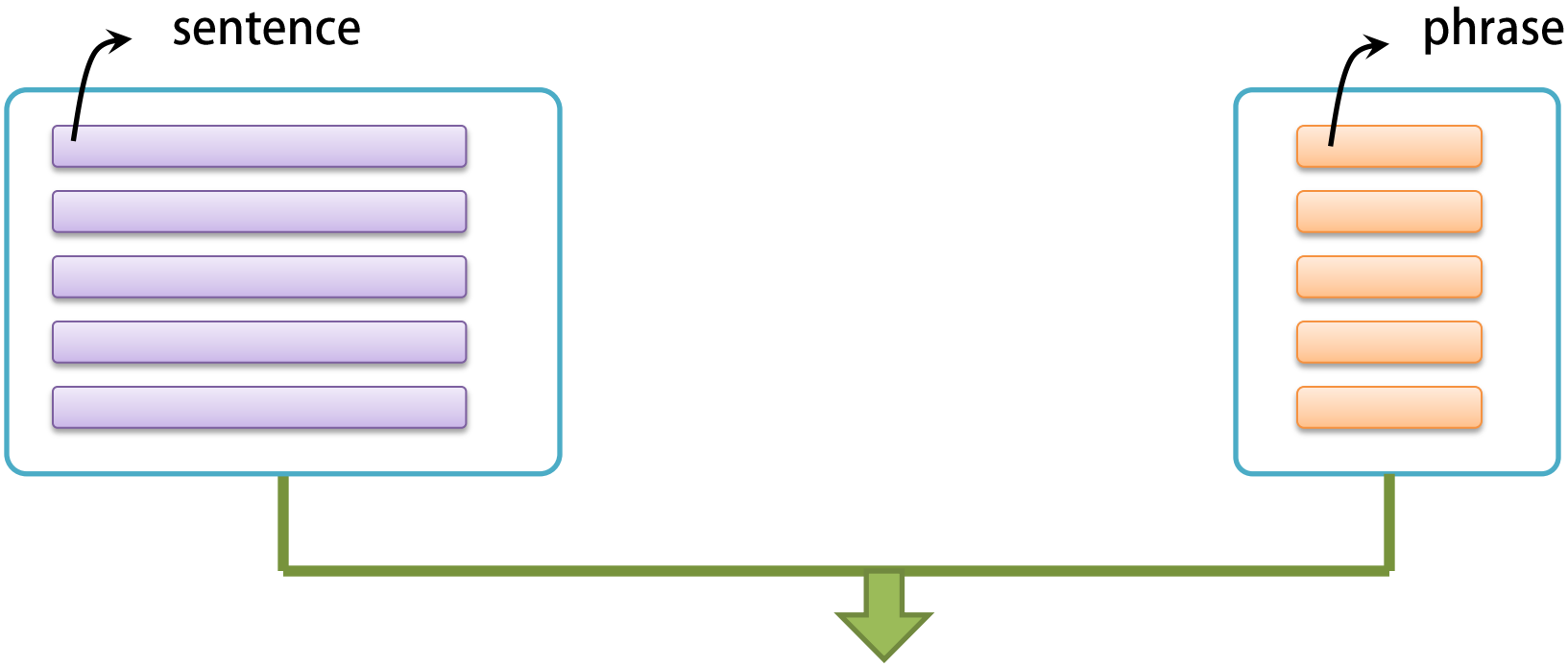
Crafting an idealized similarity distribution



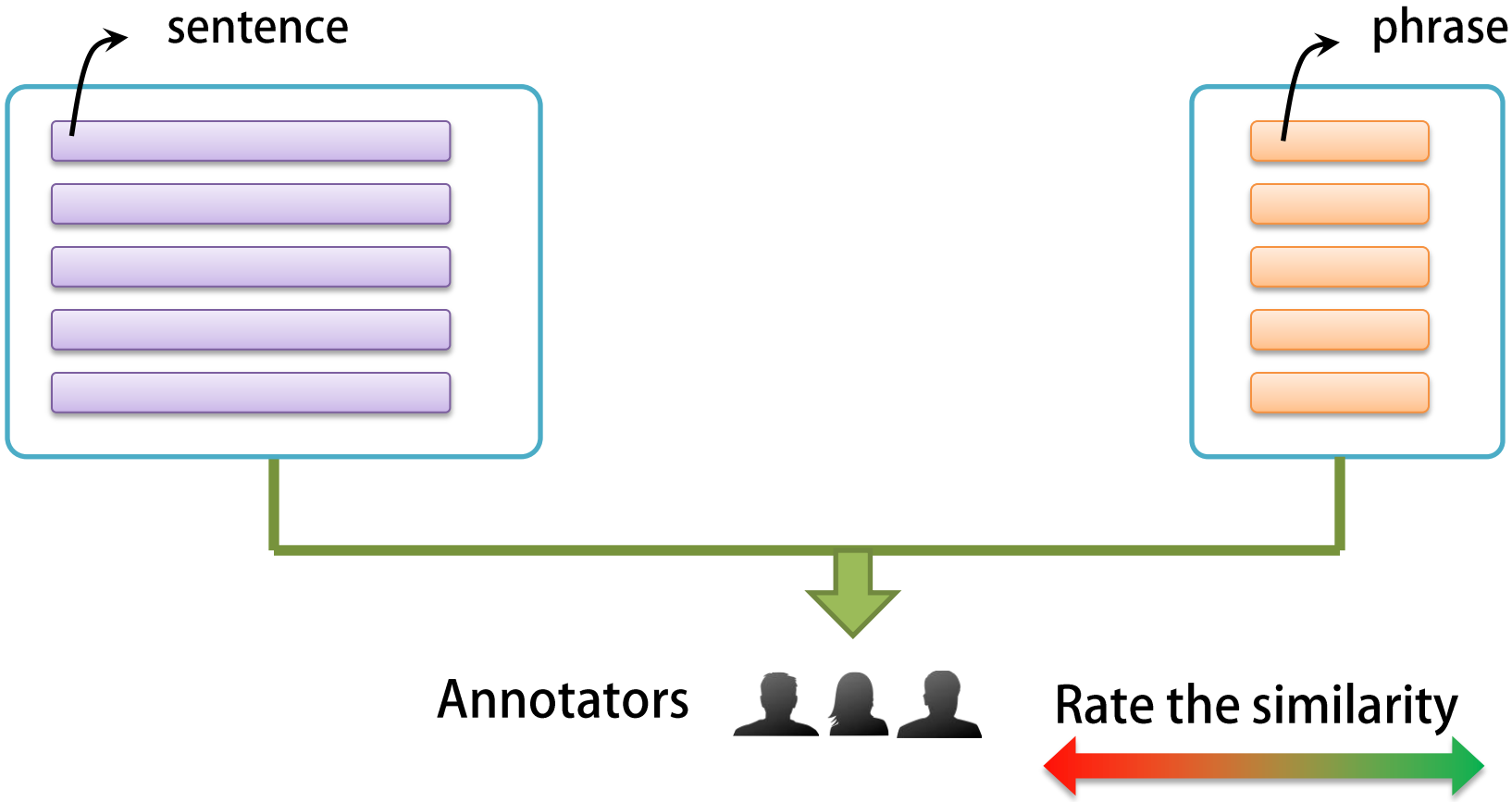
Crafting an idealized similarity distribution



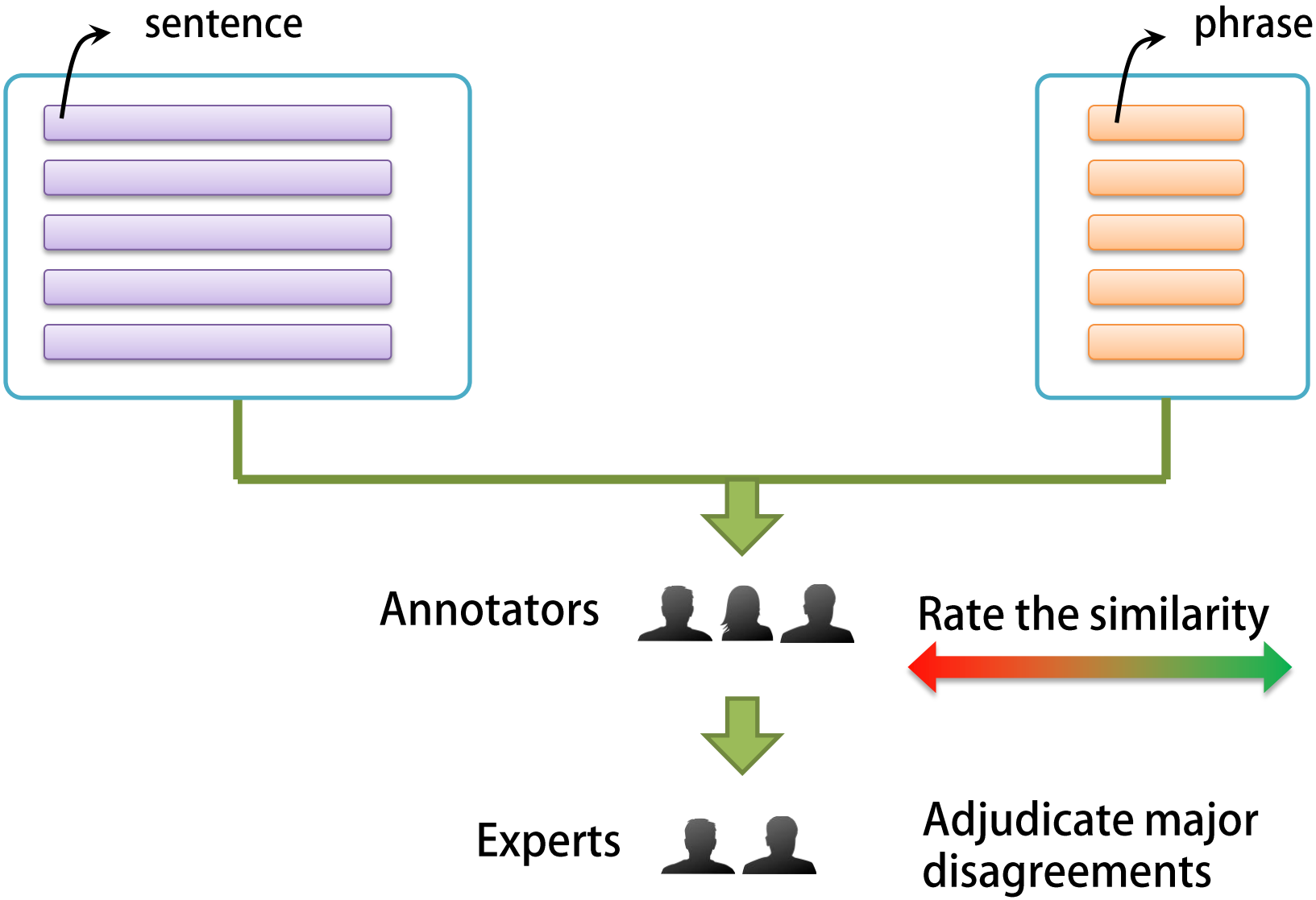
Crafting an idealized similarity distribution



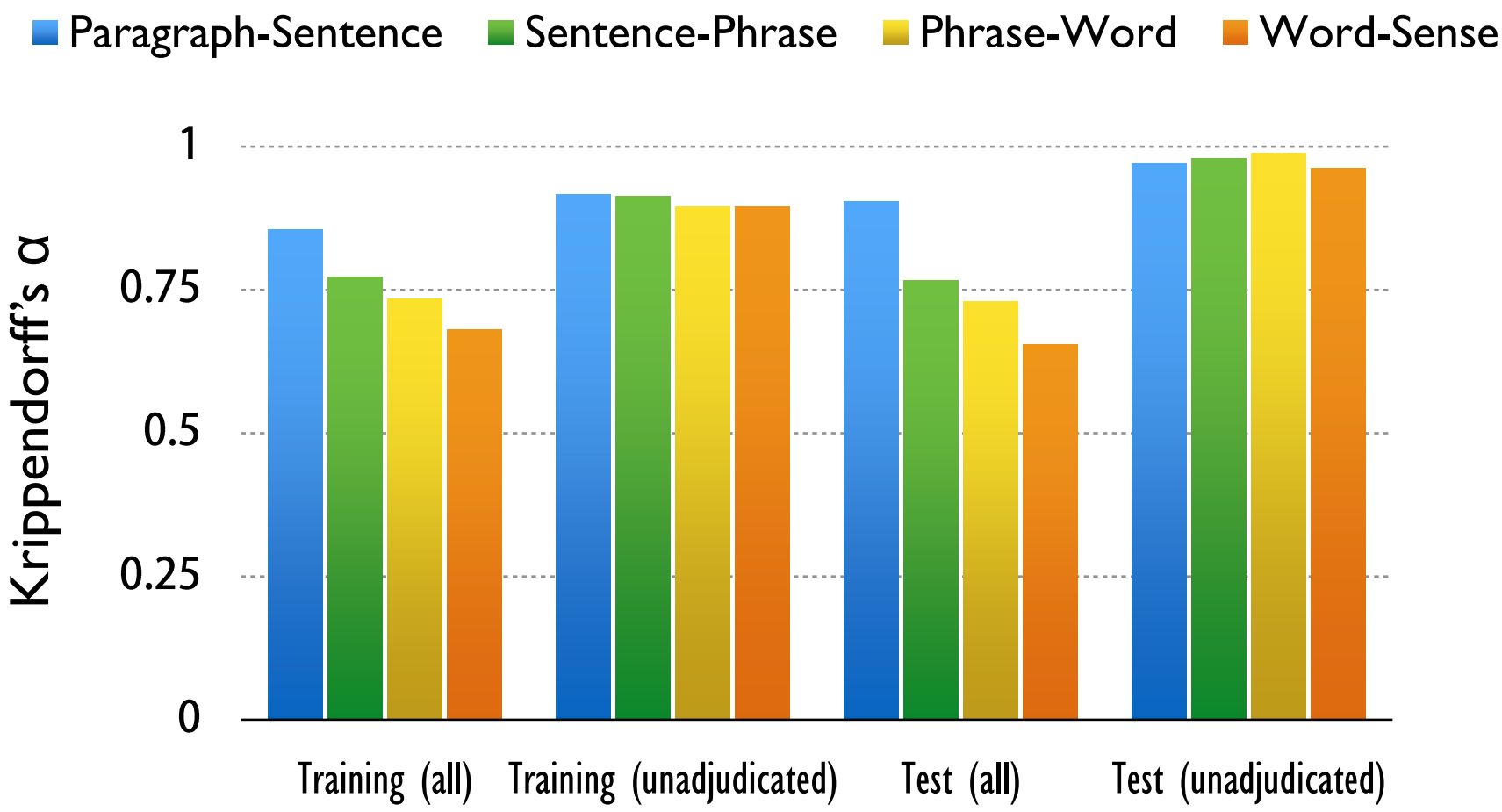
Crafting an idealized similarity distribution



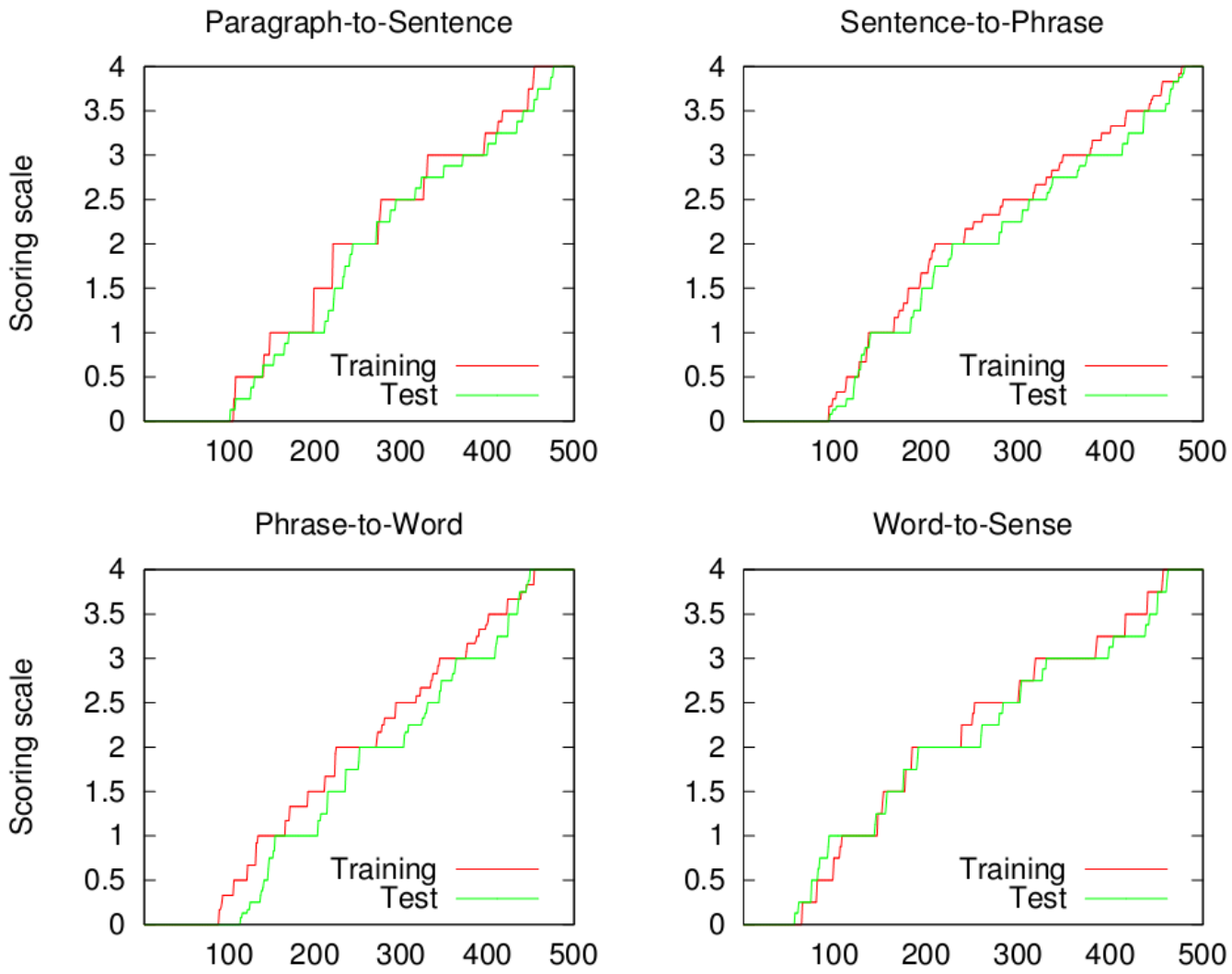
Crafting an idealized similarity distribution



Test and Training data IAA



The annotation procedure produces a balanced rating distribution



Experimental Setup

Baslines:

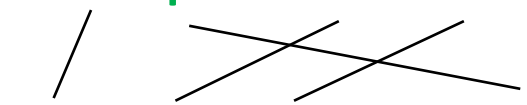
- Longest Common Substring (LCS)

The quick **brown fox**

The **brown fox** was quick

- Greedy String Tiling (GST)

The quick brown fox



The brown foxes were **quick**

Experimental Setup

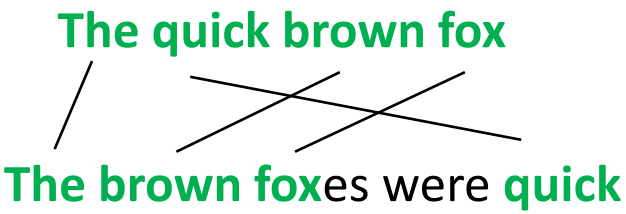
Baslines:

- Longest Common Substring (LCS)

The quick **brown fox**

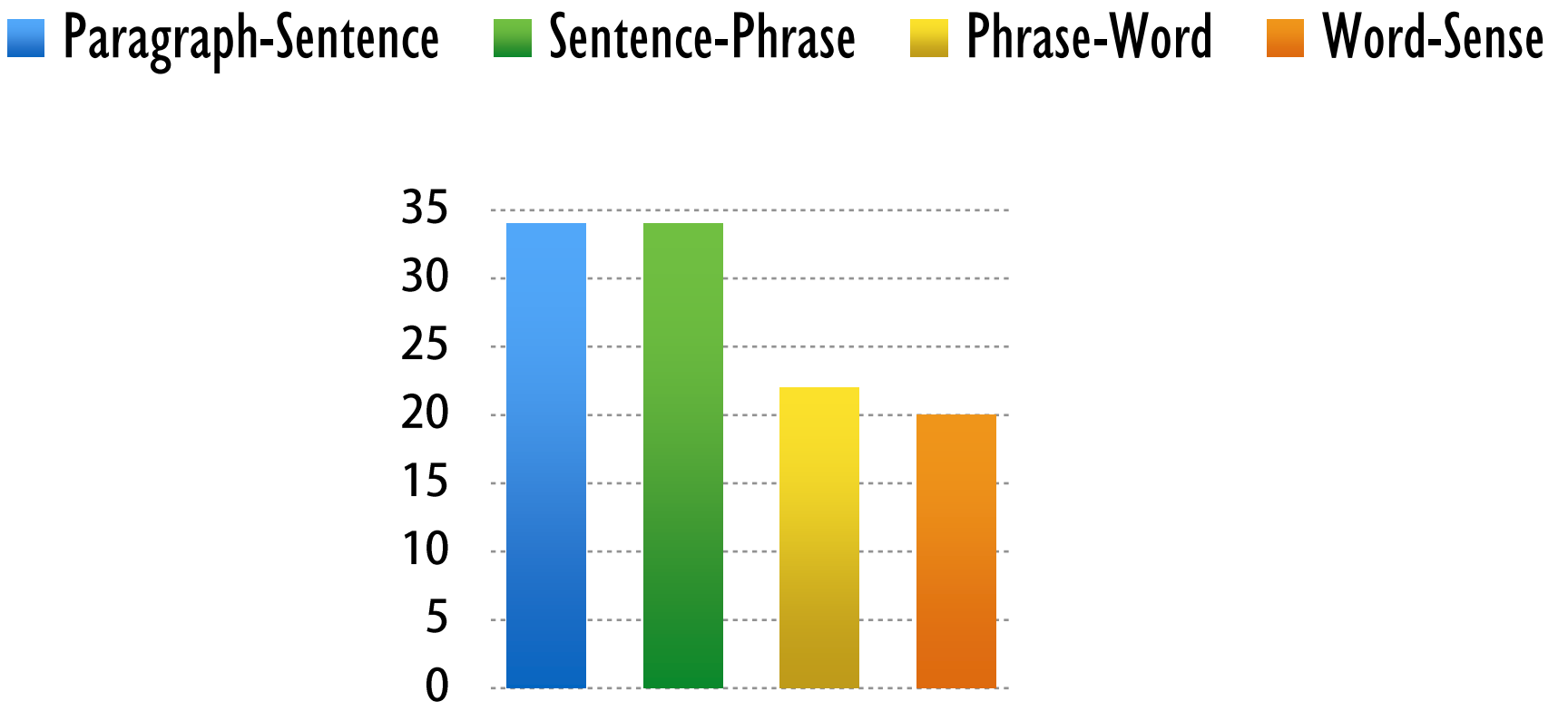
The **brown fox** was quick

- Greedy String Tiling (GST)



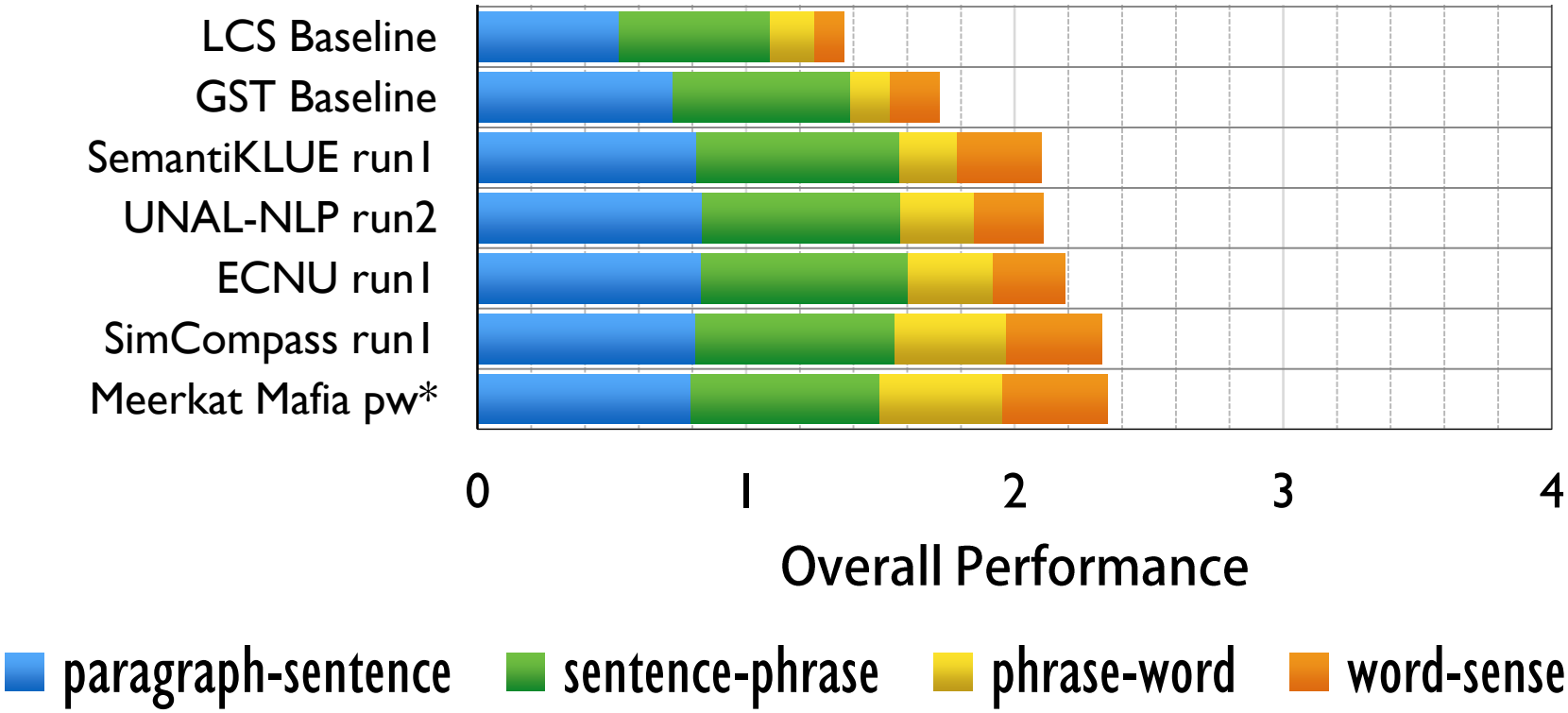
Evaluation Measure: Pearson correlation

Number of participants

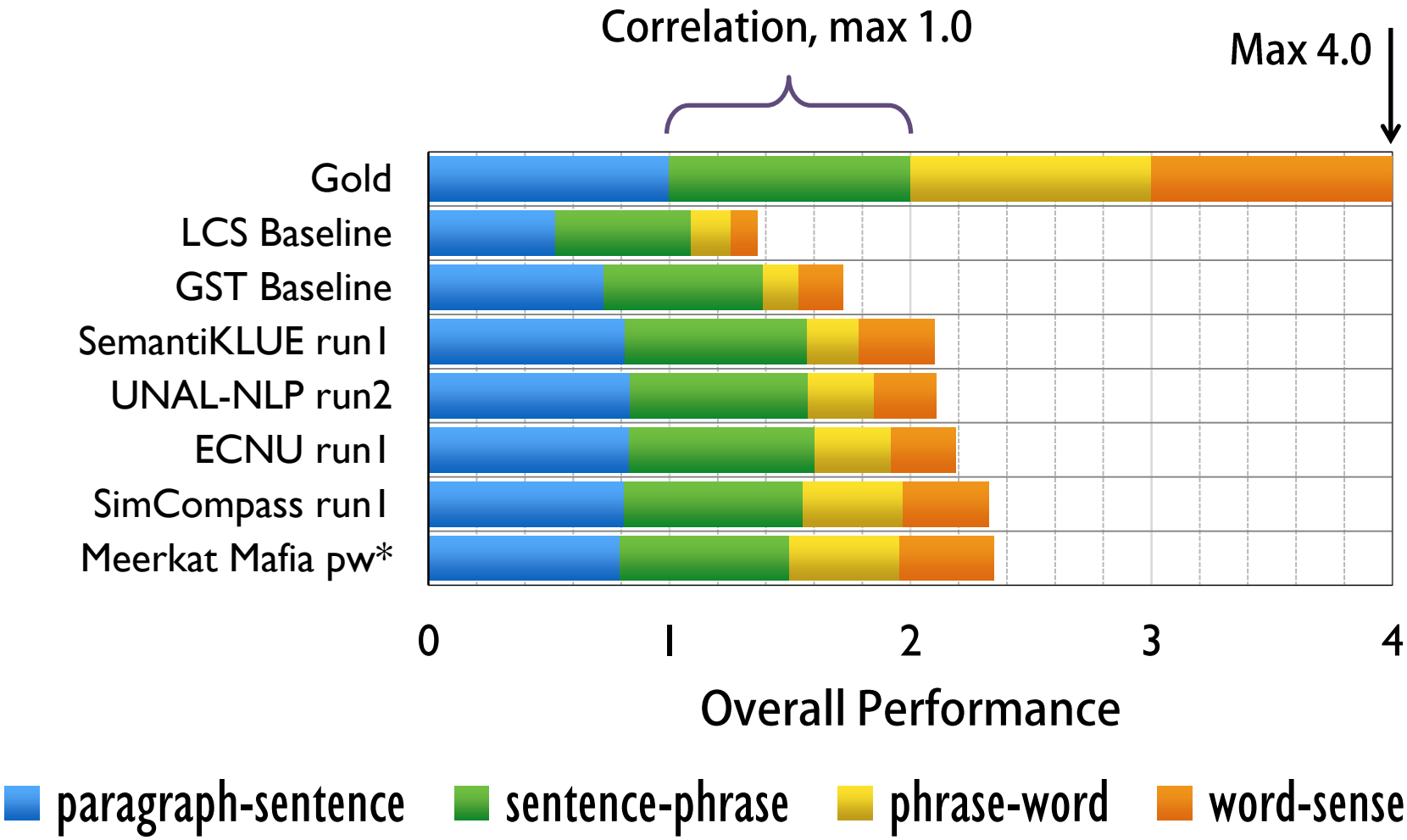


38 Systems total from 19 teams

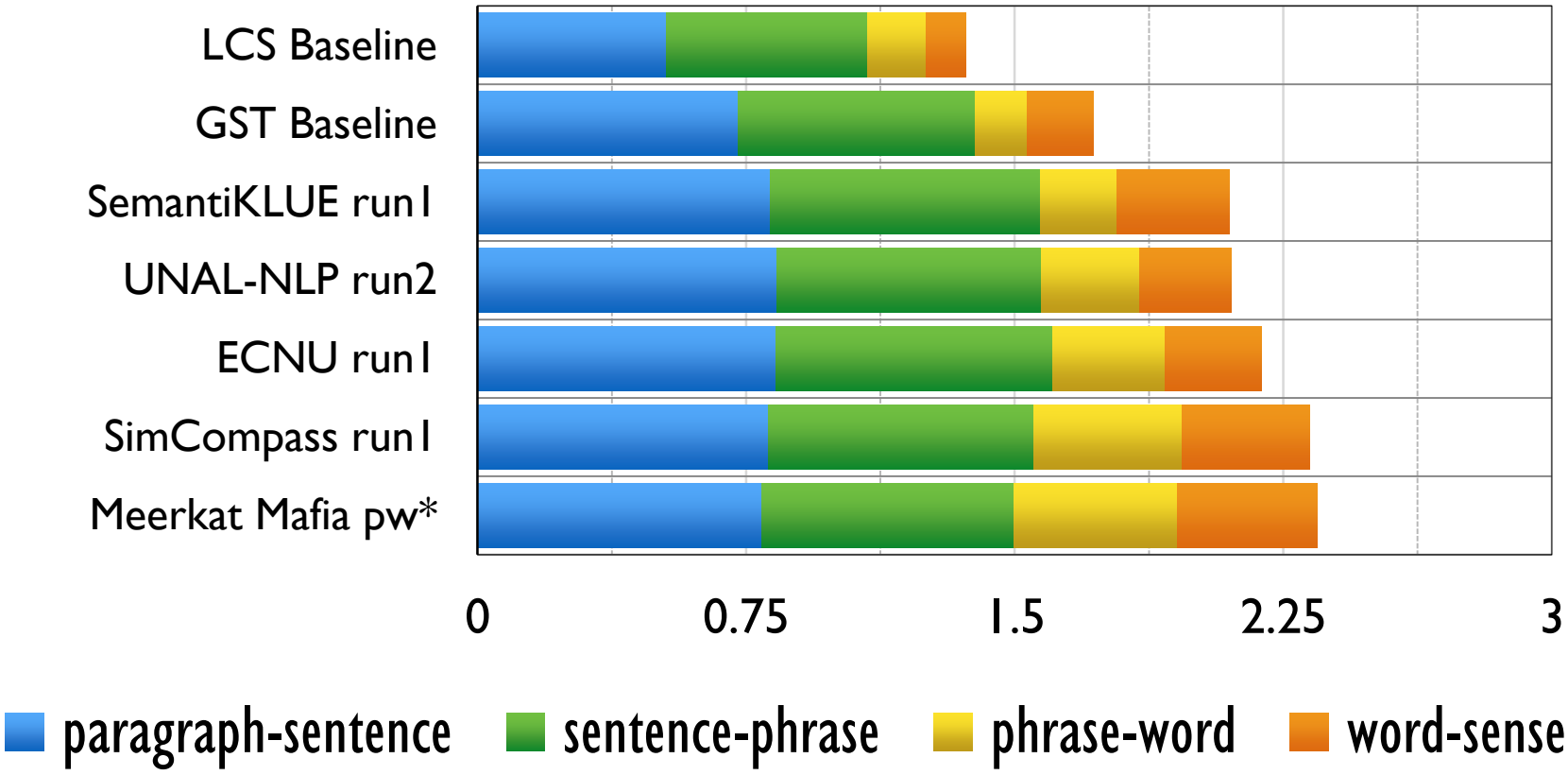
Top 5 Systems and Baselines



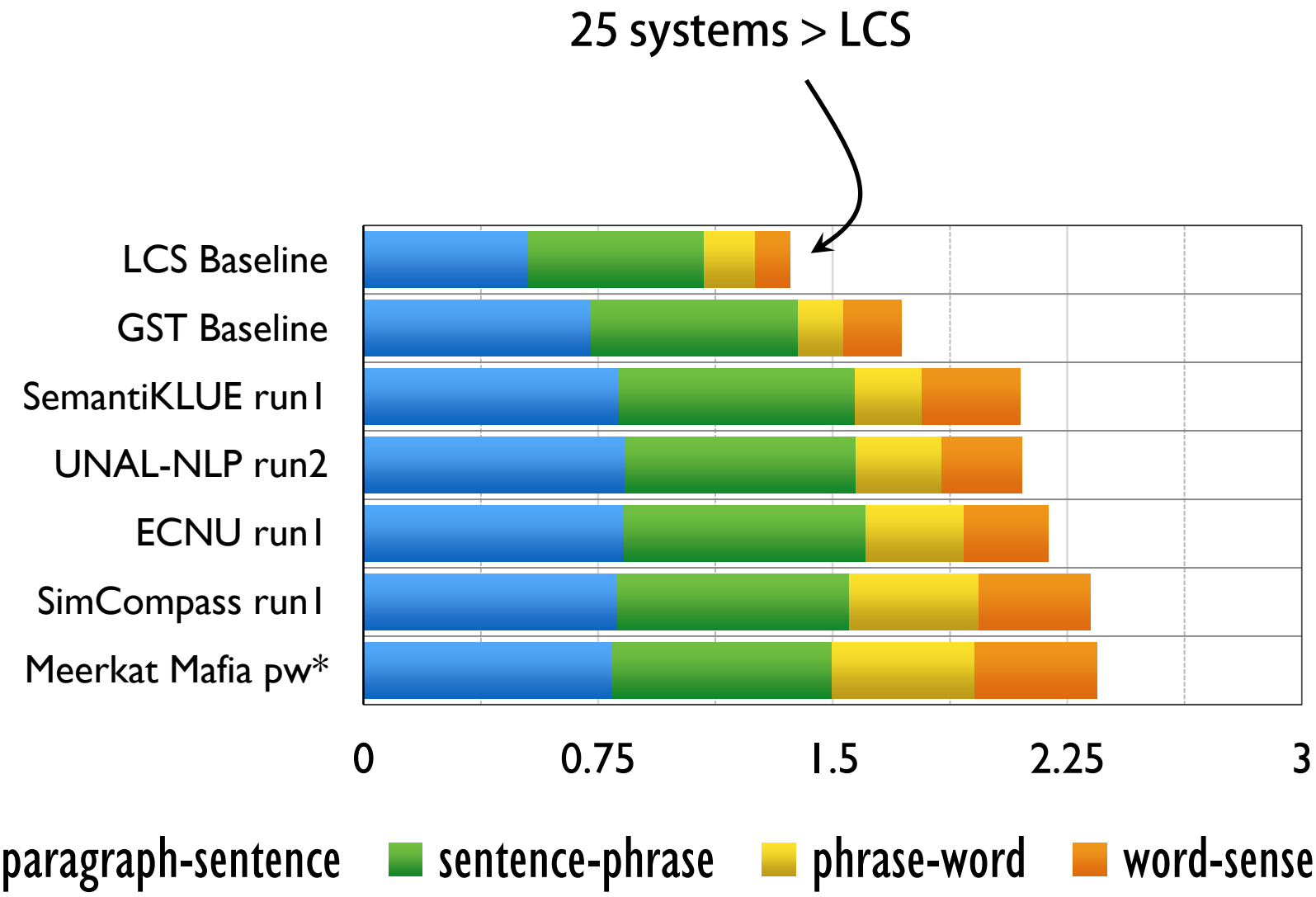
Top 5 Systems and Baselines



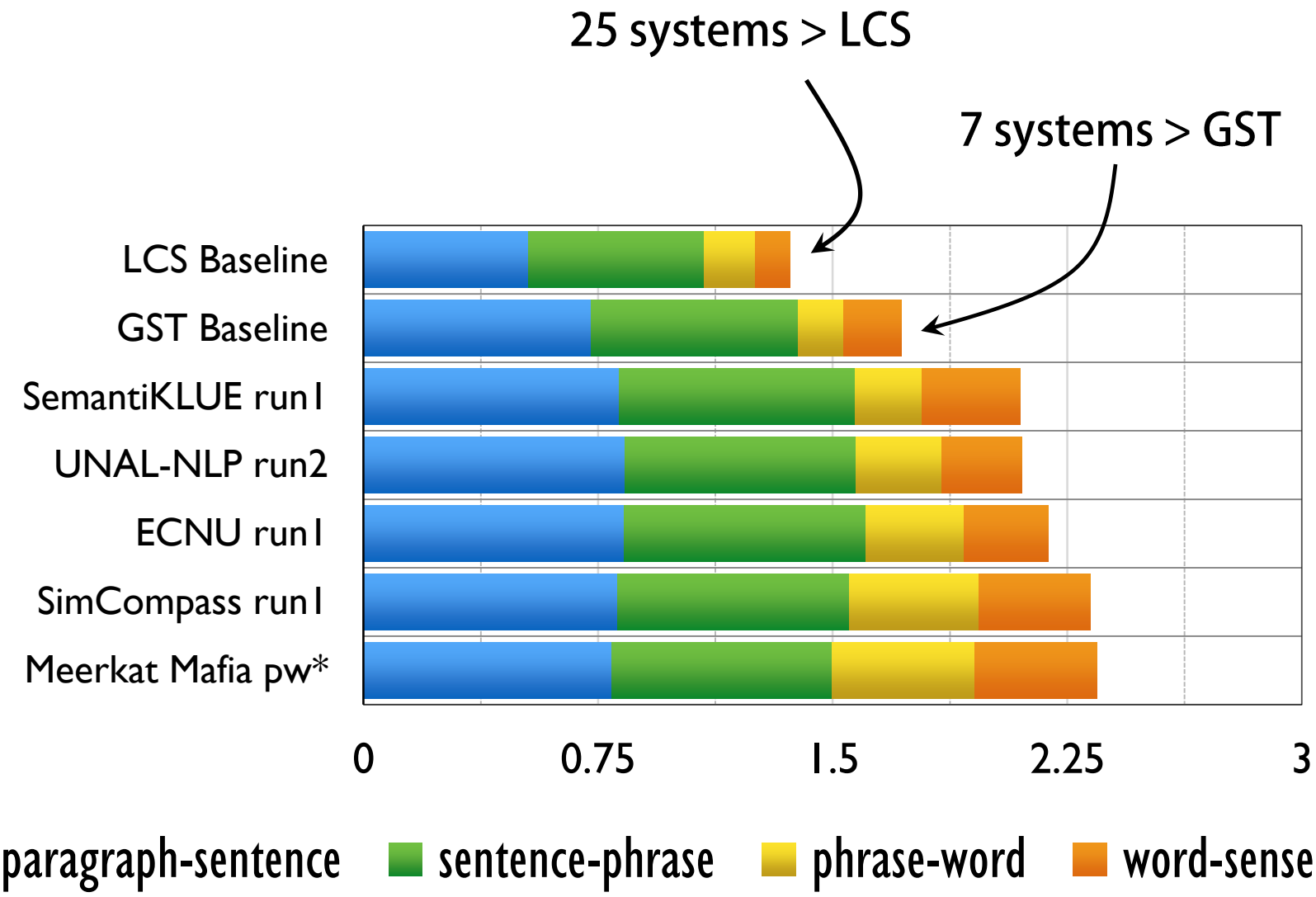
Where do the baselines stand?



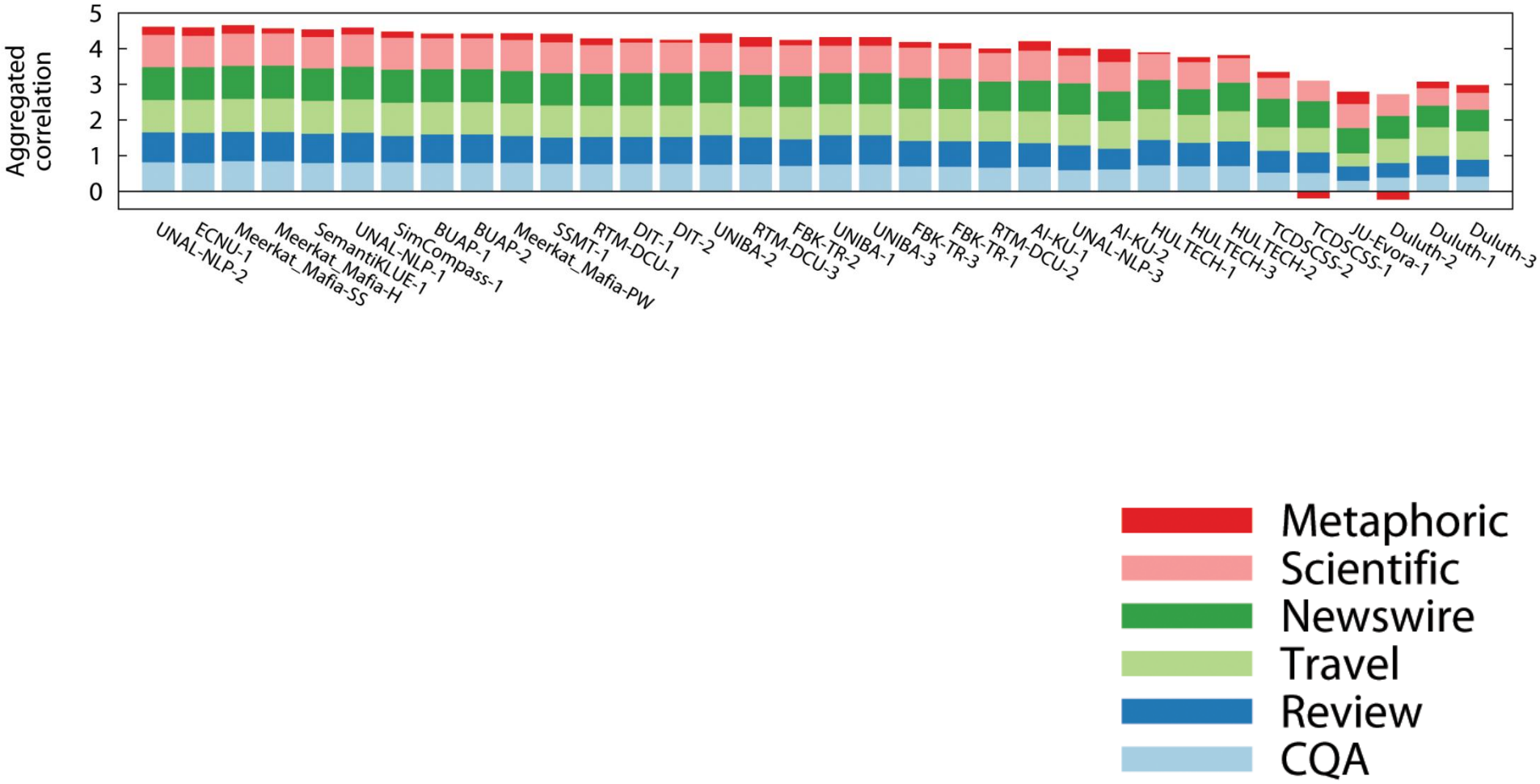
Where do the baselines stand?



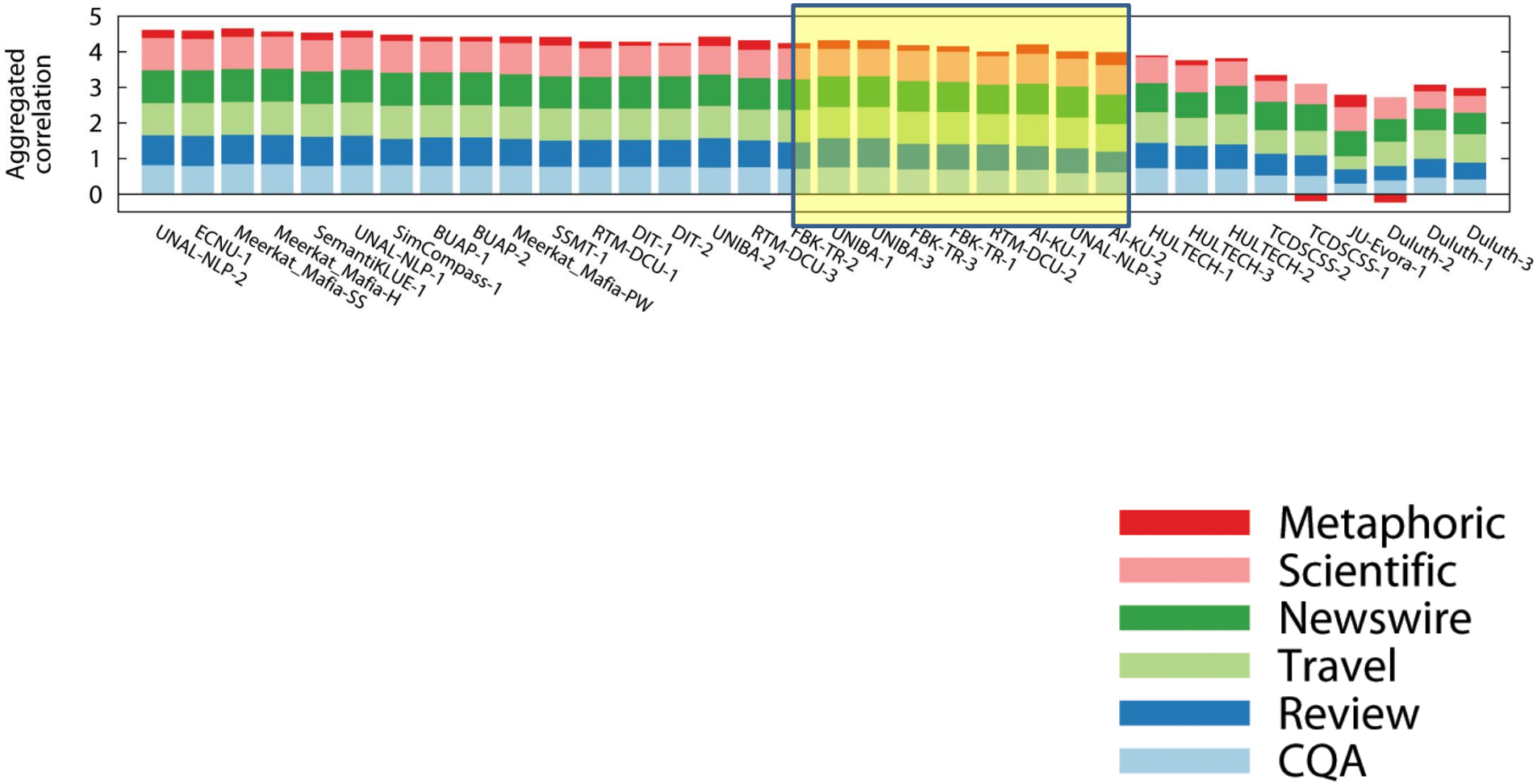
Where do the baselines stand?



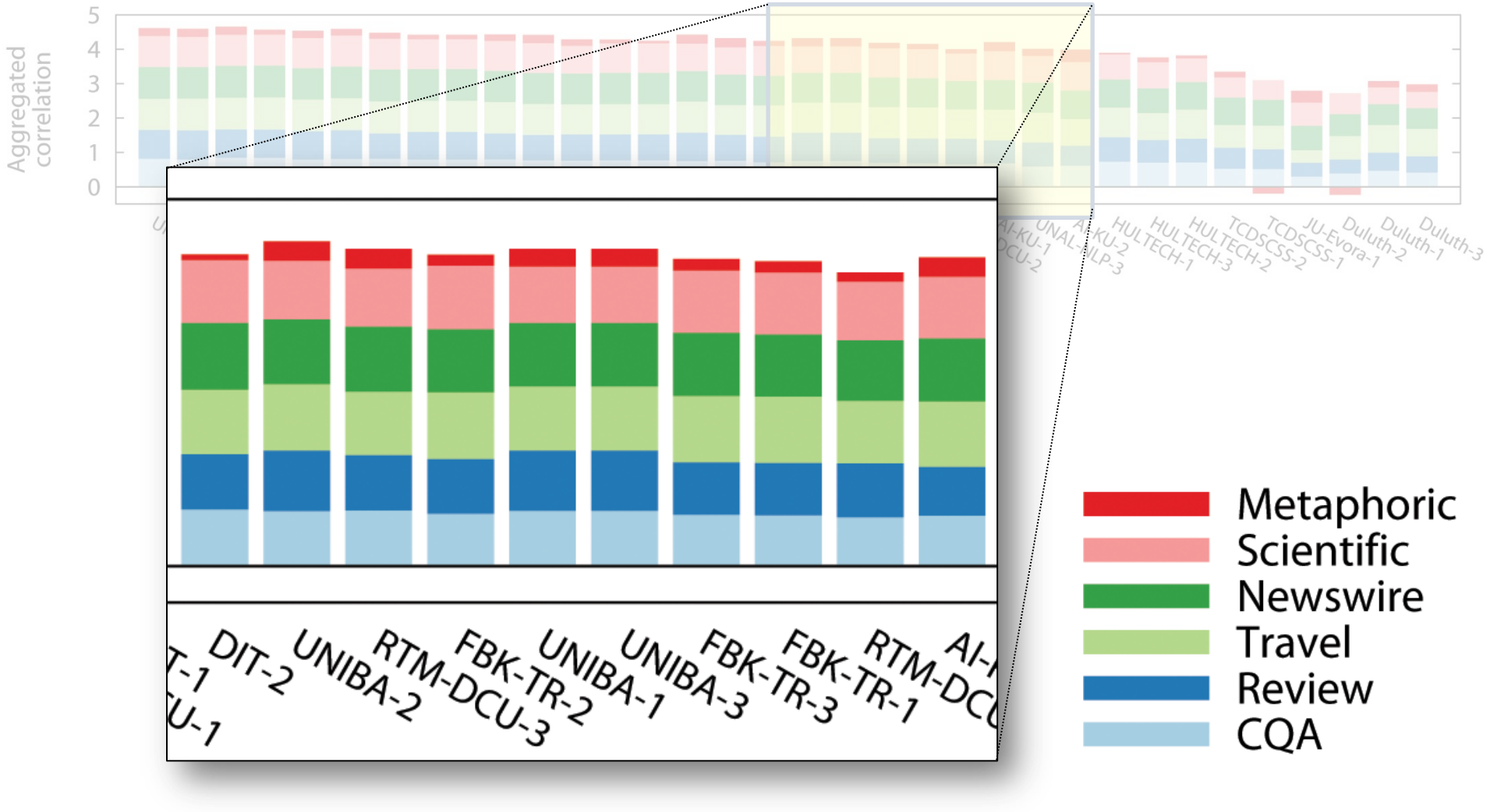
Correlation per genre
paragraph-to-sentence



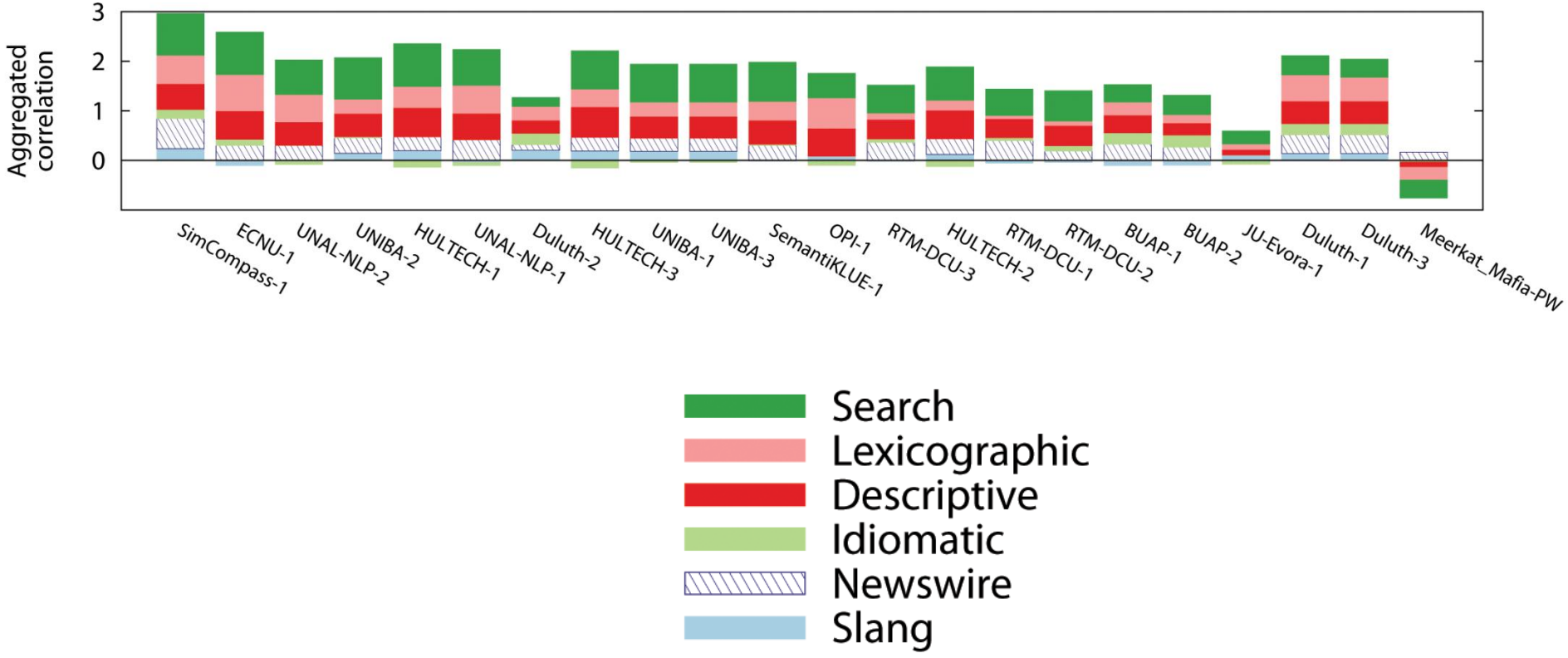
Correlation per genre
paragraph-to-sentence



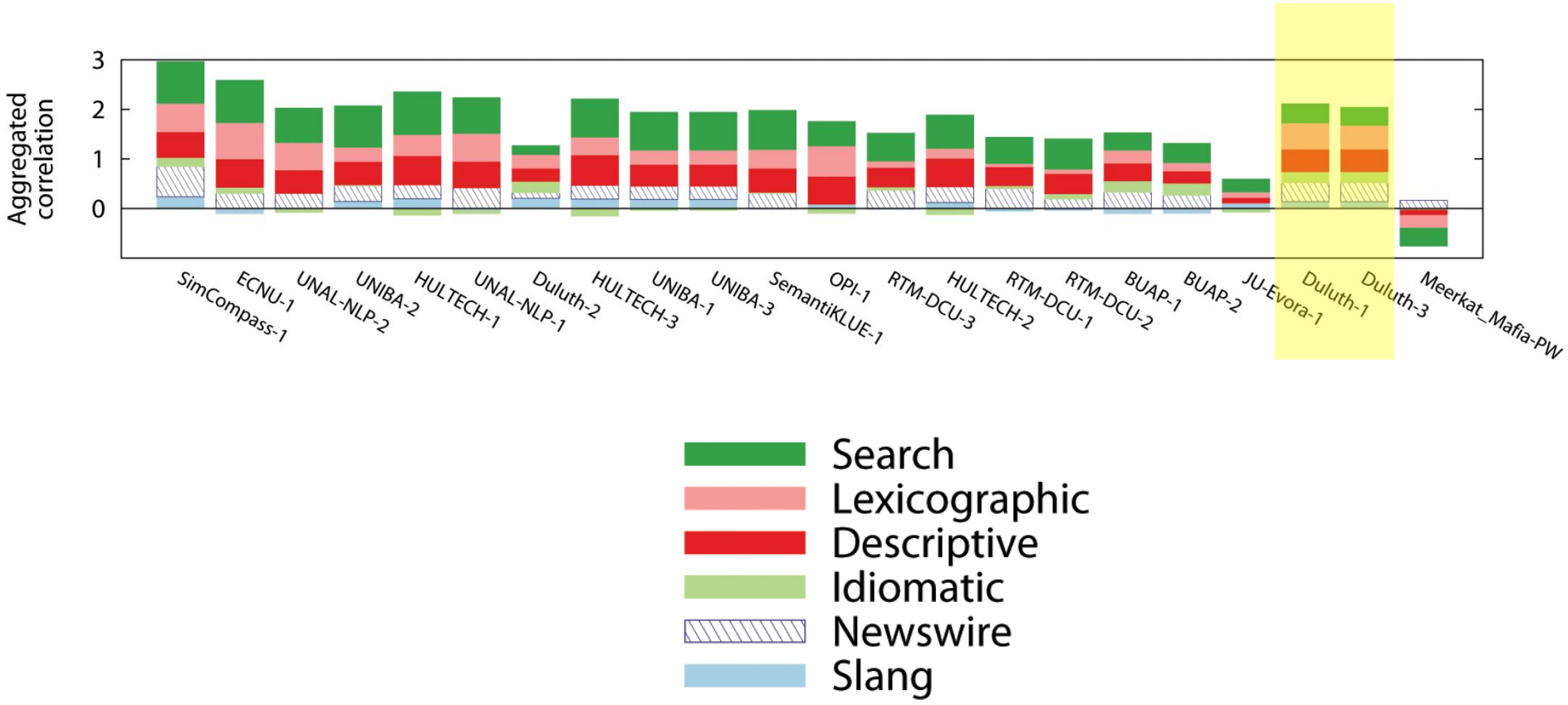
Correlation per genre
paragraph-to-sentence



Correlation per genre
phrase-to-word



Correlation per genre
phrase-to-word



What makes the task difficult?

Handling OOV words and novel usages

How often do draik eggs come in Merifoods in Meridell?

frequency of draik eggs in Merifoods

Hard feelings

grudge

Dealing with *social media* text

can i watch 4od bbc iplayer etc with 10GB useage allowence?

online television streaming for bbc

Can d Internet companies see which websyts ive bin visiting?

internet provider's knowledge of my actions

CLSS: Cross-Level Semantic Similarity

Similarity of different types of lexical items

High-quality dataset: 4000 pairs for four comparison types

38 systems from 19 teams

http://alt.qcri.org/semeval2014/task3/

Thank you!

SAPIENZA
UNIVERSITÀ DI ROMA

MultiJEDI
ERC 259234

David Jurgens

Mohammad Taher Pilehvar

Roberto Navigli

http://alt.qcri.org/semeval2014/task3/