

# Rough Analysis of marked\_and\_linked\_test.json

## General Information

- Num. Entries: 22
- Tokenizer used: dmis-lab/biobert-v1.1

### Definitions:

- Words represent the number of entities that have been considered distinct during the process of assigning tags to part of the text, it is essentially the text split on spaces
- Tokens are the actual pieces of data going into the model, punctuation is split from words and some large words are split down into smaller parts
- Total tags is the total number of tags across all words (If two or more words are part of the same tag this counts as a tag for each word)
- Unique tags is the total number of tags that are unique (not part of same tag)
- Max tags is the most tags a single word has
- Tagged words is the number of words that have at least one tag
- Tagged words % is the percentage of words that are tagged, it's given in brackets for individual entries.
- Avg tags is the average number of tags per word that is tagged
- MC words is the number of words with tags across multiple categories
- Total links is the total number of links across all tags (If two or more tags are linked this counts as a link for each one)
- Unique links is the total number of links that are unique (not part of the same link)
- Max links is the most amount of links a single tag or word has associated with it
- Linked tags, words is the number of tags or words that have links
- Linked % tags, words is the percentage of tags or words with links
- Avg links per tag, word is the average number of links per tag or word
- The schema used is BIES as it is all encompassing and models can convert to a lower resolution if wanted (Beginning, Inner, End, Singleton)

## Sizes

- Average words: 210.0
- Maximum words: 304
- Minimum words: 119
- Average tokens: 404.6
- Maximum tokens: 589
- Minimum tokens: 228
- Entries with over 512 tokens: 5/22, 22.73%

## Tags

### Maximums

- Total tags: 85
- Unique tags: 28

- Max tags: 4
- Tagged words: 78
- Tagged words %: 39.2%
- Avg tags: 1.22
- MC words: 1

### Averages

- Total tags: 35.0
- Unique tags: 12.0
- Max tags: 1.82
- Tagged words: 32.86
- Tagged words %: 14.83%
- Avg tags: 1.05
- MC words: 0.05

### Links

#### Maximums

- Total links: 143
- Unique links: 14
- Max links per tag, word: 10, 10
- Linked tags, words: 63, 57
- Linked % tags, words: 100.0%, 28.64%
- Avg links per tag, word: 3.67, 3.67

#### Averages

- Total links: 32.73
- Unique links: 3.64
- Max links per tag, word: 2.5, 2.5
- Linked tags, words: 18.77, 17.64
- Linked % tags, words: 61.6%, 8.34%
- Avg links per tag, word: 1.47, 1.55

### Schema

- Maximums (B, I, E, S): 21, 32, 21, 18
- Averages (B, I, E, S): 9.45, 10.45, 9.45, 5.64

### Labels

- Label: Total, Maximum per entry, Average per entry

### Categories

- Perturbing\_Action: 292, 35, 13.27
- Context: 490, 62, 22.27
- Effect: 466, 58, 21.18

### Description

- Knock Out: 44, 17, 2.0
- Conditional Knock Out: 8, 6, 0.36
- Knock In: 0, 0, 0.0
- Conditional Knock In: 0, 0, 0.0
- Knock Down: 68, 13, 3.09
- Plasmid Vector: 0, 0, 0.0

- Viral Vector: 8, 8, 0.36
- Pharmacological Inhibition: 31, 9, 1.41
- Decreased Expression Level: 17, 5, 0.77
- Increased Expression Level: 26, 9, 1.18
- Other: 90, 19, 4.09

### **Experiment\_Type**

- Organism: 58, 10, 2.64
- Tumour: 6, 2, 0.27
- Cells: 63, 14, 2.86
- Cell Line: 14, 6, 0.64
- Transformed Cell Line: 61, 17, 2.77
- Primary Culture: 0, 0, 0.0
- Organoid Culture: 0, 0, 0.0
- Cell Transplantation: 0, 0, 0.0
- Xenotransplantation: 43, 14, 1.95
- Other: 0, 0, 0.0
- Not Stated: 0, 0, 0.0

### **Species**

- Human: 14, 6, 0.64
- Mouse: 49, 10, 2.23
- Rat: 1, 1, 0.05
- Fish: 0, 0, 0.0
- Fly: 0, 0, 0.0
- Yeast: 0, 0, 0.0
- Bacterium: 0, 0, 0.0
- Other: 0, 0, 0.0
- Not Stated: 181, 26, 8.23

### **Phenotype**

- Adhesion: 0, 0, 0.0
- Apoptosis: 22, 10, 1.0
- Anoikis: 0, 0, 0.0
- Autophagy: 43, 19, 1.95
- Cell Cycle Arrest: 11, 8, 0.5
- Cell Death: 17, 12, 0.77
- Cell Growth: 17, 5, 0.77
- Cell Survival: 5, 3, 0.23
- Colony Formation: 3, 3, 0.14
- Differentiation: 5, 5, 0.23
- Entosis: 0, 0, 0.0
- Epithelial-Mesenchymal Transition: 0, 0, 0.0
- Ferroptosis: 0, 0, 0.0
- Invasion: 4, 4, 0.18
- Metastasis: 10, 5, 0.45
- Migration: 6, 4, 0.27
- Mitophagy: 11, 9, 0.5
- Necroptosis: 0, 0, 0.0
- Necrosis: 0, 0, 0.0
- Netosis: 0, 0, 0.0
- Oncosis: 0, 0, 0.0
- Proliferation: 22, 8, 1.0
- Pyroptosis: 2, 2, 0.09

- Quiescence: 0, 0, 0.0
- Self-Renewal: 0, 0, 0.0
- Senescence: 15, 8, 0.68
- Transformation: 0, 0, 0.0
- Tumour Growth: 23, 6, 1.05
- Tumourigenesis: 17, 5, 0.77

## Activity

- Causes: 52, 10, 2.36
- Inhibits: 48, 11, 2.18
- Increases: 46, 7, 2.09
- Decreases: 70, 13, 3.18
- Regulates: 8, 8, 0.36
- No Effect: 9, 5, 0.41
- Not Stated: 0, 0, 0.0

## Information on each entry

**Title:** 725\_Pmid31679823.txt\_CC.xml, Words: 123, Tokens: 233

Tag data: Total: 10, Unique: 4, Max tags: 1, Tagged words: 10 (8%), Avg tags: 1.0, MC words: 0

Link data: Total: 17, Unique: 2, Max links (tag, word): (2, 2), Linked (tags, words): (10 (100.0%), 10 (8.13%))

Avg links (tag, word): (1.7, 1.7), Schema counts (B, I, E, S): (3, 2, 3, 2)

**Title:** 178\_Pmid32999000.txt\_CC.xml, Words: 228, Tokens: 404

Tag data: Total: 18, Unique: 6, Max tags: 2, Tagged words: 16 (7%), Avg tags: 1.12, MC words: 0

Link data: Total: 20, Unique: 2, Max links (tag, word): (2, 2), Linked (tags, words): (14 (77.78%), 12 (5.26%))

Avg links (tag, word): (1.43, 1.67), Schema counts (B, I, E, S): (7, 3, 7, 1)

**Title:** 267\_Pmid32675324.txt\_CC.xml, Words: 203, Tokens: 352

Tag data: Total: 33, Unique: 8, Max tags: 1, Tagged words: 33 (16%), Avg tags: 1.0, MC words: 0

Link data: Total: 16, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (16 (48.48%), 16 (7.88%))

Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (6, 19, 6, 2)

**Title:** 423\_Pmid32203171.txt\_CC.xml, Words: 180, Tokens: 256

Tag data: Total: 13, Unique: 6, Max tags: 1, Tagged words: 13 (7%), Avg tags: 1.0, MC words: 0

Link data: Total: 6, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (6 (46.15%), 6 (3.33%))

Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (5, 1, 5, 2)

**Title:** 51\_Pmid31007149.txt\_CC.xml, Words: 267, Tokens: 568

Tag data: Total: 79, Unique: 23, Max tags: 2, Tagged words: 75 (28%), Avg tags: 1.05, MC words: 0

Link data: Total: 88, Unique: 8, Max links (tag, word): (4, 4), Linked (tags, words): (24 (30.38%), 24 (8.99%))

Avg links (tag, word): (3.67, 3.67), Schema counts (B, I, E, S): (19, 32, 19, 9)

**Title:** 114\_Pmid32312833.txt\_CC.xml, Words: 176, Tokens: 367

Tag data: Total: 24, Unique: 9, Max tags: 2, Tagged words: 22 (12%), Avg tags: 1.09, MC words: 0

Link data: Total: 58, Unique: 6, Max links (tag, word): (4, 4), Linked (tags, words): (24 (100.0%), 22 (12.5%))

Avg links (tag, word): (2.42, 2.64), Schema counts (B, I, E, S): (5, 6, 5, 8)

**Title:** 504\_Pmid31175486.txt\_CC.xml, Words: 260, Tokens: 568

Tag data: Total: 39, Unique: 13, Max tags: 2, Tagged words: 38 (15%), Avg tags: 1.03, MC words: 0

Link data: Total: 13, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (13 (33.33%), 13 (5.0%))

Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (10, 14, 10, 5)

**Title:** 67\_Pmid30661440.txt\_CC.xml, Words: 225, Tokens: 409

Tag data: Total: 27, Unique: 9, Max tags: 1, Tagged words: 27 (12%), Avg tags: 1.0, MC words: 0

Link data: Total: 9, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (9 (33.33%), 9 (4.0%))

Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (7, 8, 7, 5)

**Title:** 360\_Pmid33257688.txt\_CC.xml, Words: 199, Tokens: 387  
Tag data: Total: 85, Unique: 28, Max tags: 4, Tagged words: 78 (39%), Avg tags: 1.09, MC words: 0  
Link data: Total: 143, Unique: 14, Max links (tag, word): (10, 10), Linked (tags, words): (63 (74.12%), 57 (28.64%))  
Avg links (tag, word): (2.27, 2.51), Schema counts (B, I, E, S): (21, 25, 21, 18)

**Title:** 405\_Pmid31804607.txt\_CC.xml, Words: 198, Tokens: 422  
Tag data: Total: 31, Unique: 11, Max tags: 2, Tagged words: 27 (14%), Avg tags: 1.15, MC words: 0  
Link data: Total: 12, Unique: 2, Max links (tag, word): (2, 2), Linked (tags, words): (10 (32.26%), 8 (4.04%))  
Avg links (tag, word): (1.2, 1.5), Schema counts (B, I, E, S): (10, 7, 10, 4)

**Title:** 331\_Pmid33040078.txt\_CC.xml, Words: 304, Tokens: 546  
Tag data: Total: 58, Unique: 20, Max tags: 3, Tagged words: 55 (18%), Avg tags: 1.05, MC words: 0  
Link data: Total: 87, Unique: 10, Max links (tag, word): (6, 6), Linked (tags, words): (31 (53.45%), 28 (9.21%))  
Avg links (tag, word): (2.81, 3.11), Schema counts (B, I, E, S): (15, 18, 15, 10)

**Title:** 549\_Pmid32564202.txt\_CC.xml, Words: 248, Tokens: 506  
Tag data: Total: 19, Unique: 11, Max tags: 1, Tagged words: 19 (8%), Avg tags: 1.0, MC words: 0  
Link data: Total: 20, Unique: 4, Max links (tag, word): (2, 2), Linked (tags, words): (13 (68.42%), 13 (5.24%))  
Avg links (tag, word): (1.54, 1.54), Schema counts (B, I, E, S): (4, 3, 4, 8)

**Title:** 438\_Pmid31320749.txt\_CC.xml, Words: 274, Tokens: 536  
Tag data: Total: 60, Unique: 20, Max tags: 4, Tagged words: 49 (18%), Avg tags: 1.22, MC words: 0  
Link data: Total: 72, Unique: 9, Max links (tag, word): (4, 4), Linked (tags, words): (42 (70.0%), 35 (12.77%))  
Avg links (tag, word): (1.71, 2.06), Schema counts (B, I, E, S): (15, 13, 15, 17)

**Title:** 370\_Pmid33277461.txt\_CC.xml, Words: 264, Tokens: 589  
Tag data: Total: 67, Unique: 24, Max tags: 2, Tagged words: 66 (25%), Avg tags: 1.02, MC words: 1  
Link data: Total: 9, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (9 (13.43%), 9 (3.41%))  
Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (19, 21, 19, 8)

**Title:** 737\_Pmid31526760.txt\_CC.xml, Words: 119, Tokens: 228  
Tag data: Total: 14, Unique: 5, Max tags: 1, Tagged words: 14 (12%), Avg tags: 1.0, MC words: 0  
Link data: Total: 10, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (10 (71.43%), 10 (8.4%))  
Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (5, 3, 5, 1)

**Title:** 32\_Pmid30290714.txt\_CC.xml, Words: 255, Tokens: 462  
Tag data: Total: 39, Unique: 16, Max tags: 2, Tagged words: 37 (15%), Avg tags: 1.05, MC words: 0  
Link data: Total: 35, Unique: 5, Max links (tag, word): (4, 4), Linked (tags, words): (27 (69.23%), 25 (9.8%))  
Avg links (tag, word): (1.3, 1.4), Schema counts (B, I, E, S): (15, 7, 15, 2)

**Title:** 751\_Pmid30205049.txt\_CC.xml, Words: 128, Tokens: 236  
Tag data: Total: 15, Unique: 7, Max tags: 1, Tagged words: 15 (12%), Avg tags: 1.0, MC words: 0  
Link data: Total: 6, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (6 (40.0%), 6 (4.69%))  
Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (6, 1, 6, 2)

**Title:** 150\_Pmid32107212.txt\_CC.xml, Words: 150, Tokens: 290  
Tag data: Total: 15, Unique: 3, Max tags: 1, Tagged words: 15 (10%), Avg tags: 1.0, MC words: 0  
Link data: Total: 15, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (15 (100.0%), 15 (10.0%))  
Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (2, 10, 2, 1)

**Title:** 789\_Pmid29805077.txt\_CC.xml, Words: 132, Tokens: 280  
Tag data: Total: 6, Unique: 3, Max tags: 1, Tagged words: 6 (5%), Avg tags: 1.0, MC words: 0  
Link data: Total: 6, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (6 (100.0%), 6 (4.55%))  
Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (2, 0, 2, 2)

**Title:** 430\_Pmid29317762.txt\_CC.xml, Words: 239, Tokens: 438  
Tag data: Total: 42, Unique: 16, Max tags: 2, Tagged words: 41 (17%), Avg tags: 1.02, MC words: 0  
Link data: Total: 55, Unique: 7, Max links (tag, word): (4, 4), Linked (tags, words): (42 (100.0%), 41 (17.15%))  
Avg links (tag, word): (1.31, 1.34), Schema counts (B, I, E, S): (14, 8, 14, 6)

**Title:** 421\_Pmid30341421.txt\_CC.xml, Words: 211, Tokens: 382  
Tag data: Total: 18, Unique: 5, Max tags: 1, Tagged words: 18 (9%), Avg tags: 1.0, MC words: 0

Link data: Total: 14, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (14 (77.78%), 14 (6.64%))  
Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (3, 8, 3, 4)

**Title:** 395\_PMID33243998.txt\_CC.xml, Words: 237, Tokens: 443

Tag data: Total: 58, Unique: 17, Max tags: 3, Tagged words: 49 (21%), Avg tags: 1.18, MC words: 0

Link data: Total: 9, Unique: 1, Max links (tag, word): (1, 1), Linked (tags, words): (9 (15.52%), 9 (3.8%))

Avg links (tag, word): (1.0, 1.0), Schema counts (B, I, E, S): (15, 21, 15, 7)