# Analysis Report

# AntMover and AWA

# Table of Contents

# 1.0 Requirements

## 1.1 Test Data

The corpus (OA-STM) that being used in this statistic analysis of AntMover and AWA is from Elsevier repository. Elsevier is providing a selection of articles from 10 different STM (Scientific, Technical, and Medical) domains as a freely-redistributable corpus. The articles were selected from our Open Access content and have a Creative Commons CC-BY license so they are free to redistribute and use. The domains are agriculture, astronomy, biology, chemistry, computer science, earth science, engineering, materials science, math, and medicine. Currently we provide 11 articles in each of the 10 domains. So, the total articles in the corpus is 110 with 30930 sentences.

*10 Categories*: Agriculture, Astronomy, Biology, Chemistry, Computer Science, Earth Science, Engineering, Material Science, Mathematics and Medicine


## 1.2 Database

The database used in this research is MySQL database (***will be migrated to postgresSQL in Amazon Web Service***). All the sentences in the 110 articles are pre-processed and stored in the database. The table structures are designed to store all the documents, sentences, annotations and tools in order.


## 1.3 Statistical and programming language

The programming language used in this research is Python. Meanwhile, the statistic computing used R programming language.

# 2.0 AntMover

## 2.1 Background

AntMover 1.0 is a prototype version of a general learning environment that can be applied to the analysis of text structure in any field or discipline, and to any text type. It is a freeware text structure (moves) analysis program.

## 2.2 Drawback

After some testing on this software, I found these are the drawback of this system:

1. The function to separate sentences are not accurate and efficient

→ not able to identify the end of any sentence that end with a character or abbreviation. As a result, the sentence will not be annotated.

*Example:*

The result shows that robots can be constructed with the method shown in Figure A. Therefore, that method is efficient. (these two sentences are not broken into two sentences)

## 2.3 Advantage

Antmover annotates every sentence. However, the test for accuracy on how it annotates a sentence should be conducted by comparing to annotations made my human.

## 2.4 Annotation Scheme/ Moves

| Annotation Id | Annotation Name |
|---:|---|
| 1 | Claiming centrality |
| 2 | Making topic generalization |
| 3 | Indicating a gap |
| 4 | Announcing present research |
| 5 | Announcing principal findings |
| 6 | Evaluation of research |

## 2.5 Test 1: Spread of moves over corpus

This test shows the frequency of each AntMover's moves in the corpus data.

| Annotation Id | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|
| Sentence | 27 | 3725 | 96 | 2680 | 18186 | 6216 | 30930 |
| Proportion (Document) | 0.101 | 13.707 | 0.366 | 8.991 | 64.016 | 22.82 | 110 |

Proportion (Document) formula

= SUM(total_annotation_id_document/total_sentences_document)



AntMover: Frequency (Sentence) of Annotation Scheme



AntMover: Proportion (Document) of Annotation Scheme
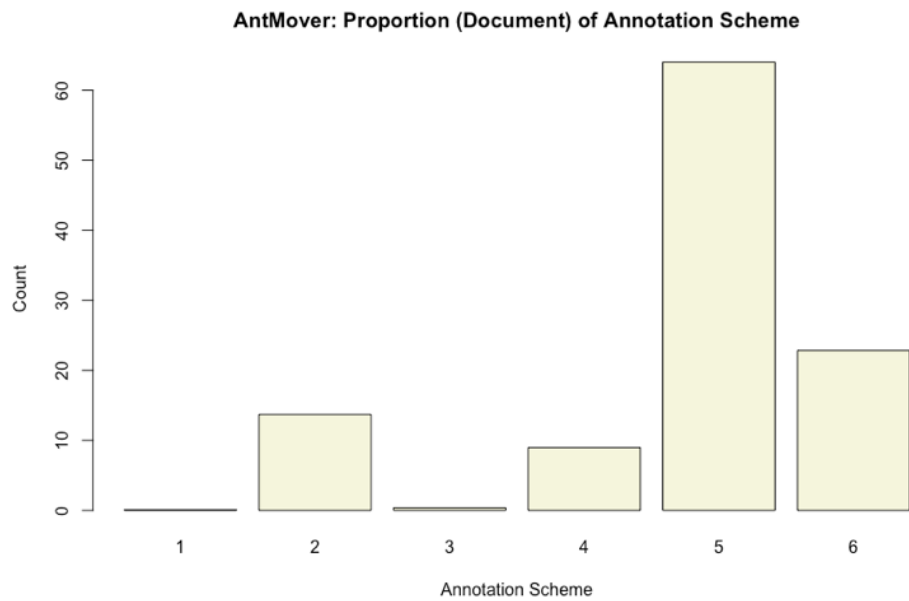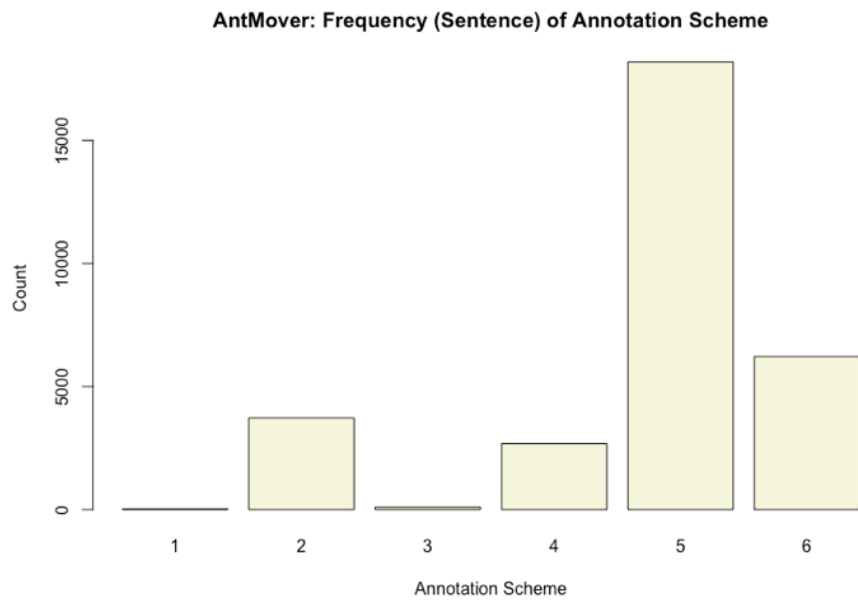
**Result**: It appears that most of the sentences are annotated with annotation 5. Meanwhile, annotations '1' and '3' are very less likely appear in the corpus. It also shows that not all annotations ('1' to '6') can be found in every document in the corpus (110 documents). As an example, only 27 annotation '1' in the whole corpus even though there are 110 documents. One more finding is that each sentence only annotated with at most one move as the total number of moves is 30930 which is equal to the total number of sentences in the corpus.

## 2.6 Test 2: Spread of moves over sub-corpora (document category)

This test shows the frequency of occurrences of each annotation scheme in each category of documents of the corpus data.

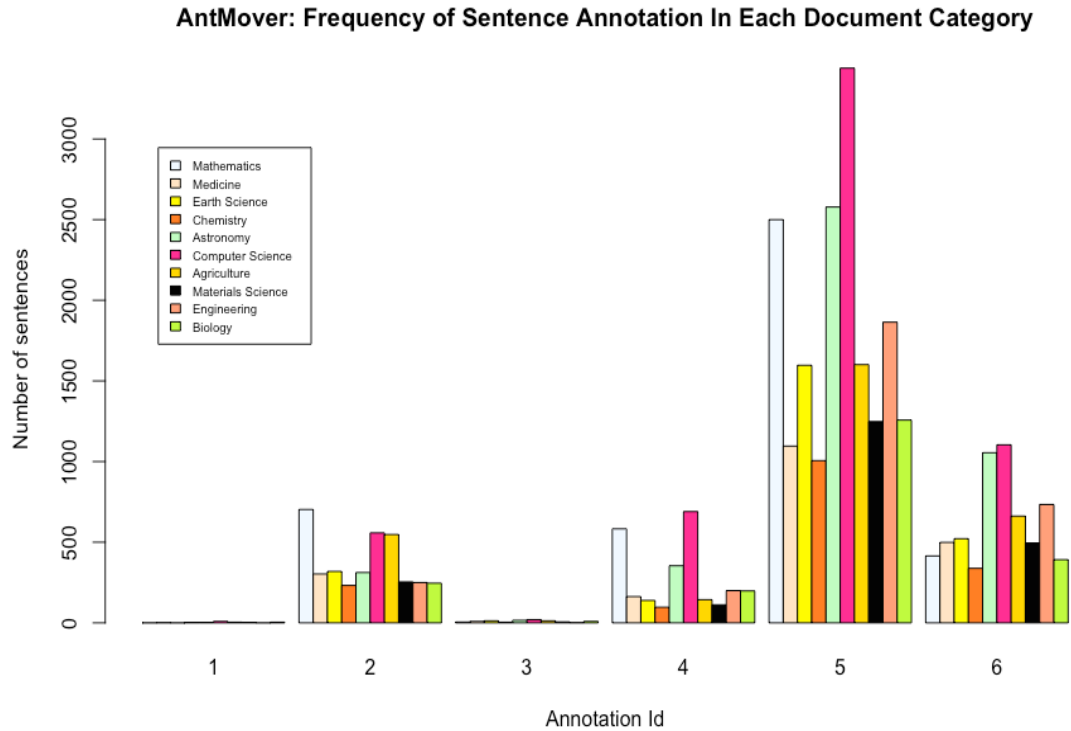| Category | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|
| Mathematics | 0 | 703 | 5 | 583 | 2500 | 415 | 4206 |
| Medicine | 1 | 303 | 9 | 163 | 1096 | 499 | 2071 |
| Earth Science | 0 | 319 | 13 | 139 | 1597 | 522 | 2590 |
| Chemistry | 3 | 232 | 4 | 97 | 1006 | 338 | 1680 |
| Astronomy | 3 | 312 | 16 | 354 | 2578 | 1055 | 4318 |
| Computer Science | 9 | 558 | 20 | 690 | 3439 | 1104 | 5820 |
| Agriculture | 4 | 548 | 13 | 144 | 1601 | 662 | 2972 |
| Materials Science | 3 | 254 | 6 | 112 | 1248 | 495 | 2118 |
| Engineering | 0 | 250 | 1 | 200 | 1864 | 734 | 3049 |
| Biology | 4 | 246 | 9 | 198 | 1257 | 392 | 2106 |
| Total | 27 | 3725 | 96 | 2680 | 18186 | 6216 | 30930 |

**AntMover: Frequency of Sentence Annotation In Each Document Category**

***Result***: Since AntMover annotates every sentence, there is nothing to derived from the above charts. It just appears that most of the annotations fall into annotation '5'. It also shows that some categories of documents do not contain certain annotations at all. Example, documents under category Mathematics do not contain any annotation '1'.

### 2.7 Test 3: Is every sentence annotate by AntMover?

By using the below SQL query, it shows that all sentences are annotated by AntMover.

**Query***:*

*SELECT COUNT(\*) FROM SENTENCE WHERE sentence_id*

*NOT IN(SELECT sentence_id FROM SENTENCE_ANNOTATION WHERE tool_id=1);*

**Query result**: 0

**Result**: All sentences are annotated by AntMover's moves.

## 2.8 Test 4: Is AntMover annotation process remains stable (reliable)?

I scheduled two time of annotations with the same corpus (110 articles) in different session. Then I tried to compare the results of the two sessions.

*Query:*

*#to count the number of sentences that **having same annotation id for two different dates***

*SELECT COUNT(\*) FROM SENTENCE_ANNOTATION sa1, SENTENCE_ANNOTATION sa2*

*WHERE sa1.sentence_id=sa2.sentence_id*

*AND sa1.annotation_id = sa2.annotation_id*

*AND sa1.tool_id = 1 AND sa2.tool_id=1*

*AND sa1.sentence_date='2017-10-08'*

*AND sa2.sentence_date='2017-10-10';*

**Query Result:** 30930 (this is the total number of lines in 110 documents)

#to count number of sentences with **different annotation id for two different dates**

*SELECT COUNT(\*) FROM SENTENCE_ANNOTATION sa1, SENTENCE_ANNOTATION sa2*

*WHERE sa1.sentence_id=sa2.sentence_id*

*AND sa1.annotation_id != sa2.annotation_id*

*AND sa1.tool_id = 1 AND sa2.tool_id=1*

*AND sa1.sentence_date='2017-10-08'*

*AND sa2.sentence_date='2017-10-10';*

**Query Result:** 0

**Result:** Each session produce the same result (same annotation ids on same sentences). Therefore, AntMover's annotation process is reliable.

# 3.0 Academic Writing Analytics (AWA)

## 3.1 Background

Academic Writing Analytics or AWA (pronounced ay-wah) is a web application designed to provide insights to students on their writing. AWA comes in 2 flavours: An analytic version for analysing normal analytical style academic writing, and a reflection version for analysing reflective writing. Both flavours are actively being developed as part of CIC research projects. However, the focus of this research is on the analytical parser.

## 3.2 Drawback

I found that AWA does sentence correction. It may be helpful but it changes the original sentence without notify the users. In this project, it poses a challenge because the uploaded sentences from the corpus are edited and cannot be matched to the original sentences in the database. Therefore, manual matchings have been done to find out which original sentences are being annotated.

## 3.3 Annotation Scheme

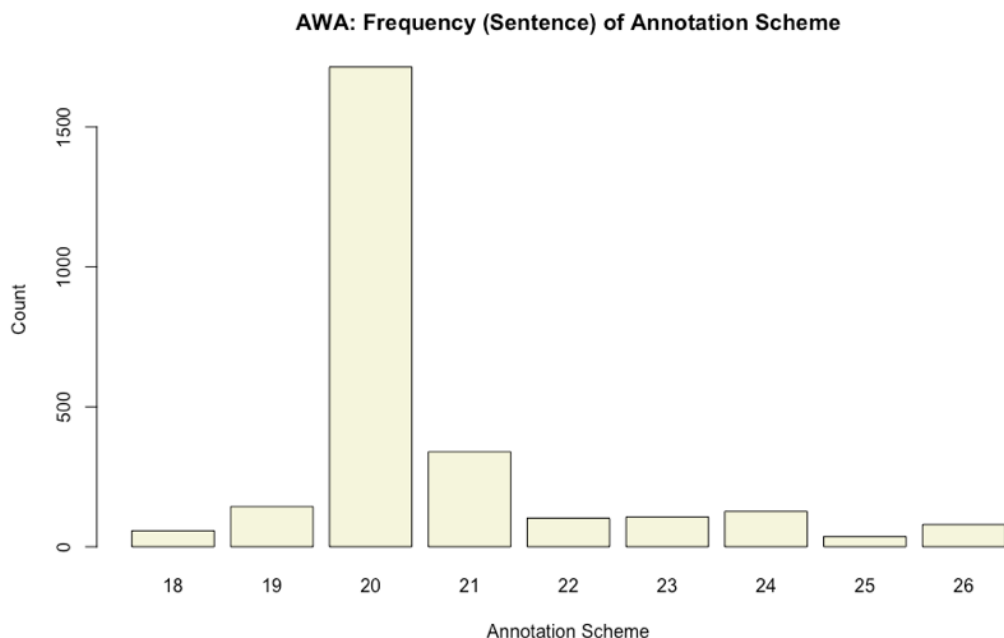| Annotation Id | Annotation Name |
|---------------|-----------------|
| 18 | MainCategory |
| 19 | Background |
| 20 | Contrast |
| 21 | Emphasis |
| 22 | Novelty |
| 23 | Position |
| 24 | Question |
| 25 | Surprise |
| 26 | Trend |

The annotation scheme of AWA is divided into two categories which are 'main category' and 'sub category'. The 'main category' contains 'important', 'summary' and 'important &

summary'. Meanwhile the 'sub category' contains 'background', 'contrast', 'emphasis', 'novelty', 'position', 'question', 'surprise' and 'trend'. **Note that each sentence annotated by AWA must falls into 'main category' and not necessary 'sub category'. Thus, the 'MainCategory' column from the above table refers to the annotated sentences that not fall into any 'sub category'. The others annotation labels refer to sentence being annotated with 'main category' as well as the specified 'sub category'. Each sentence can be annotated by more than one 'sub category'.**

### 3.4 Test 1: Spread of annotations over corpus

This test shows the frequency of occurrences of each annotation scheme in the corpus data.

| Annotation Id | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Sentence | 57 | 143 | 1714 | 339 | 102 | 106 | 126 | 36 | 79 | 2702 |
| Proportion (Document) | 0.264 | 0.572 | 6.205 | 1.511 | 0.387 | 0.405 | 0.495 | 0.159 | 0.299 | 10.297 |



AWA: Frequency (Sentence) of Annotation Scheme

AWA: Proportion (Document) of Annotation Scheme
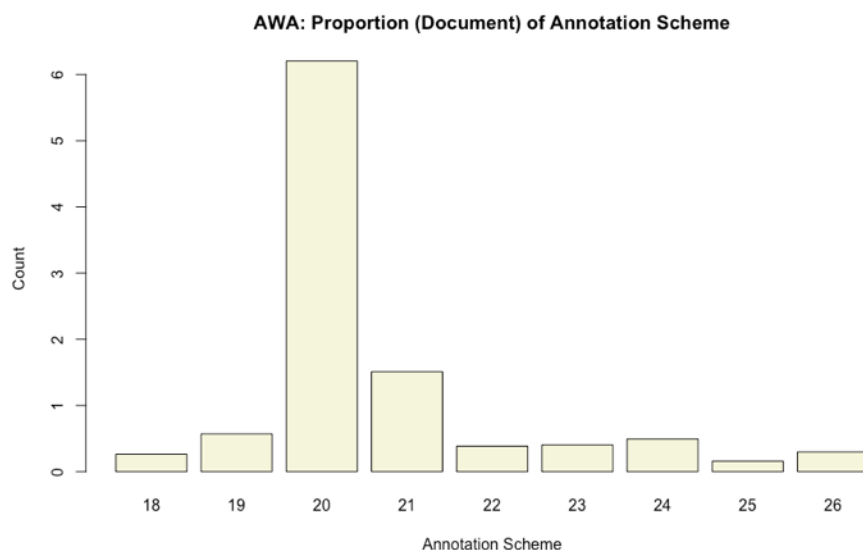
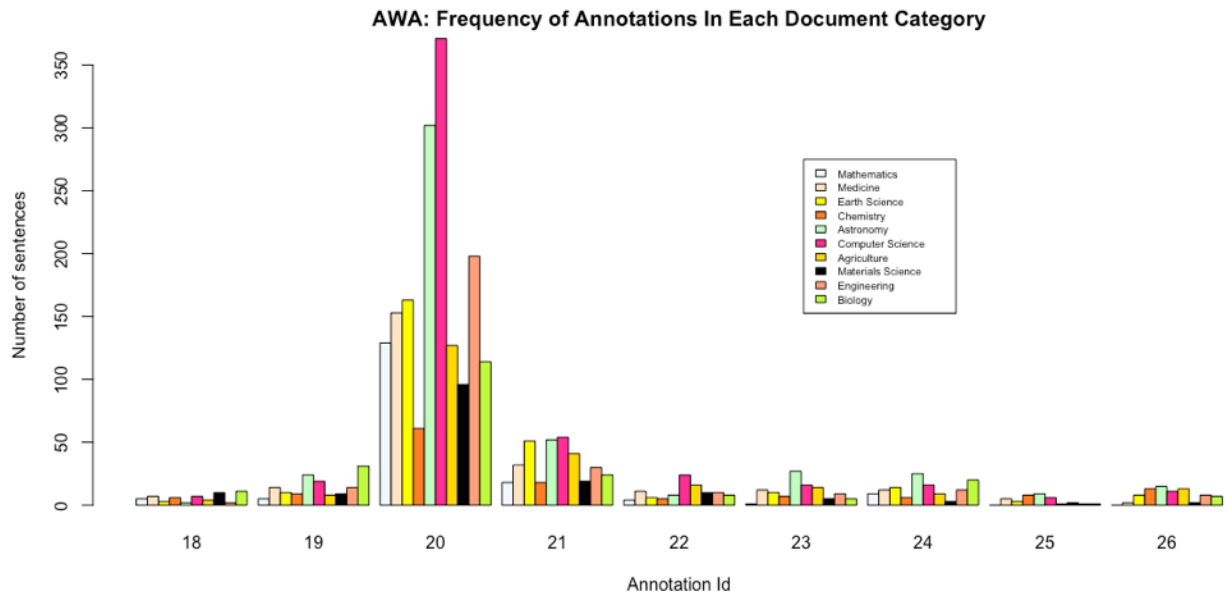**Result**: Most of the annotated sentences fall into annotation '20' follow by annotation '21'. Meanwhile, annotation '25' is least likely to appear. Total number of annotated sentences is 2702. It means that AWA is not annotating every sentence. This is contradicted to AntMover that annotates every sentence with moves.

## 3.5 Test 2: Spread of annotations over sub-corpora (document category)

This test shows the frequency of occurrences of each annotation scheme in each category of documents of the corpus data. Notice that one sentence with two annotations will be counted as two as this test focus on number of annotations and not annotated sentences.

| Category | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Mathematics | 5 | 5 | 129 | 18 | 4 | 1 | 9 | 0 | 0 | 171 |
| Medicine | 7 | 14 | 153 | 32 | 11 | 12 | 12 | 5 | 2 | 248 |
| Earth Science | 3 | 10 | 163 | 51 | 6 | 10 | 14 | 3 | 8 | 268 |
| Chemistry | 6 | 9 | 61 | 18 | 5 | 7 | 6 | 8 | 13 | 133 |
| Astronomy | 2 | 24 | 302 | 52 | 8 | 27 | 25 | 9 | 15 | 464 |
| Computer Science | 7 | 19 | 371 | 54 | 24 | 16 | 16 | 6 | 11 | 524 |
| Agriculture | 4 | 8 | 127 | 41 | 16 | 14 | 9 | 1 | 13 | 233 |
| Materials Science | 10 | 9 | 96 | 19 | 10 | 5 | 3 | 2 | 2 | 156 |
| Engineering | 2 | 14 | 198 | 30 | 10 | 9 | 12 | 1 | 8 | 284 |
| Biology | 11 | 31 | 114 | 24 | 8 | 5 | 20 | 1 | 7 | 221 |
| Total | 57 | 143 | 1714 | 339 | 102 | 106 | 126 | 36 | 79 | 2702 |

AWA: Frequency of Annotations In Each Document Category

**Result**: In each category, most of its sentences fall into annotation '20'. Computer Science category has the highest number annotations (524 annotations) follow by Astronomy category (464 annotations). Meanwhile the lowest number of annotations is in Chemistry (133 annotations) follow by Material Science (156 annotations).

### 3.6 Test 3: Spread of annotated sentences over sub-corpora (document category)

This test shows the percentage of annotated sentences (opposite to Test 2 above) in each category of documents of the corpus data. Notice that one sentence with two annotations will be counted as one as this test focus on number of annotated sentences and not number of annotations.

| Category | Annotated | Not Annotated | Total | Annotated (%) | Not Annotated (%) |
|---|---|---|---|---|---|
| Mathematics | 164 | 4042 | 4206 | 3.899 | 96.101 |
| Medicine | 219 | 1852 | 2071 | 10.575 | 89.425 |
| Earth Science | 238 | 2352 | 2590 | 9.189 | 90.811 |
| Chemistry | 106 | 1574 | 1680 | 6.31 | 93.69 |
| Astronomy | 410 | 3908 | 4318 | 9.495 | 90.505 |
| Computer Science | 488 | 5332 | 5820 | 8.385 | 91.615 |
| Agriculture | 199 | 2773 | 2972 | 6.696 | 93.304 |
| Materials Science | 141 | 1977 | 2118 | 6.657 | 93.343 |
| Engineering | 255 | 2794 | 3049 | 8.363 | 91.637 |
| Biology | 189 | 1917 | 2106 | 8.974 | 91.026 |
| Total | 2409 | 28521 | 30930 | | |



AWA:Spread of Annotations (%) Over Sub-corpora (Annotated Sentences Only)

AWA: Spread of Annotations (%) Over Sub-corpora (All Sentences)
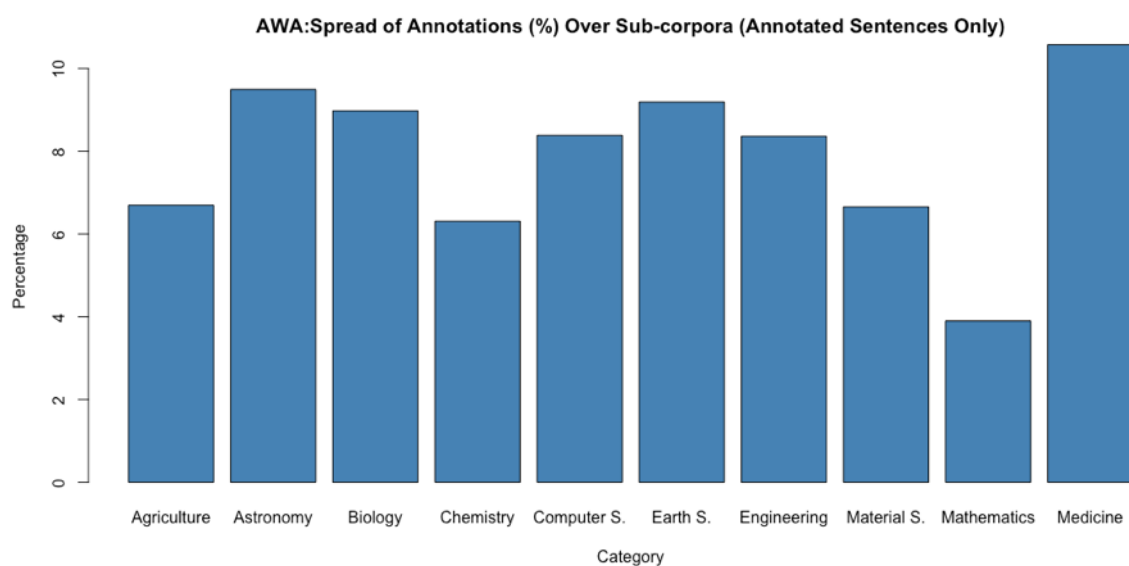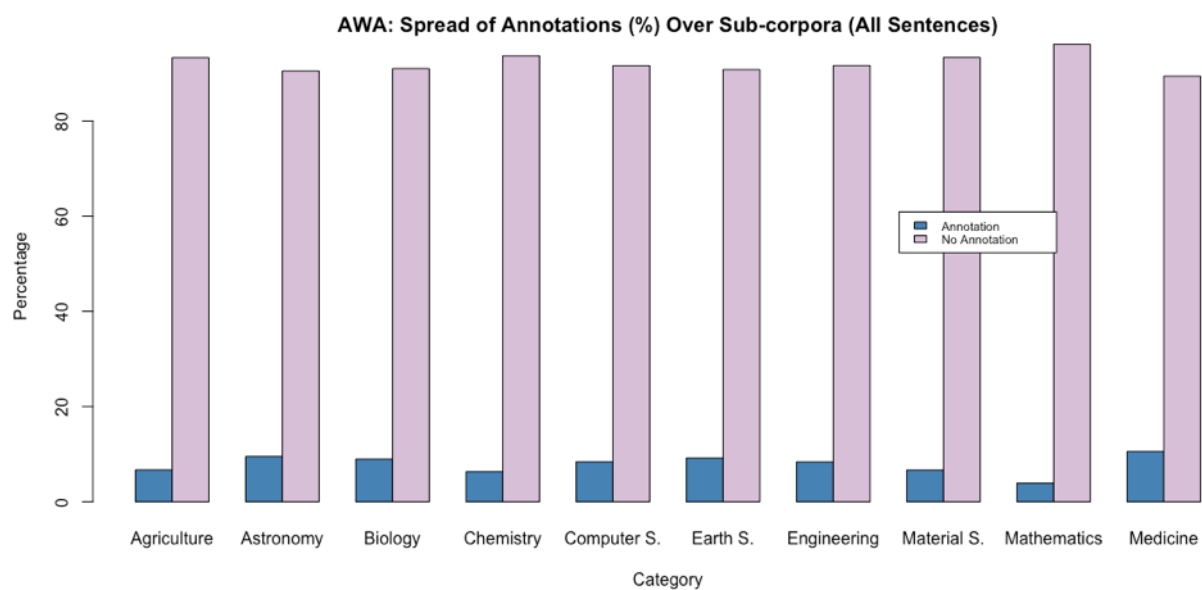
**Result**: It appears that Mathematics sub-corpora has the least number of annotated sentences. It means AWA less likely to annotate sentences in Mathematics documents. Meanwhile, the sub-corpora that works well with AWA are medicine, astronomy and earth science.

# 4.0 AntMover and AWA reliability

This section will compare the reliability of AntMover and AWA. In this test, AntMover and AWA will act as 'raters' and the number of sentences will be the subjects. The agreement between AntMover and AWA to annotate same sentences will be evaluated.

## 4.1 Test 1: Percentage agreement in annotating sentence

Subjects = 30930 (sentences)

Raters = 2 (AntMover and AWA)

%-agree = 7.79

**Result**: There is 7.79% similarity in agreement of annotating the same sentences.

## 4.2 Test 2: Krippendorf's alpha and Cohen's Kappa for reliability

Based on Krippendorff (2004) and Fulcher (2010), the minimum acceptable inter-rater agreement value for Krippendorf's alpha is 0.80 while Cohen's kappa is 0.61.

| AntMover and AWA | Krippendorf's alpha | Cohen's kappa |
|---|---|---|
| Value | -0.855 | 0 |

**Result**: The negative value of Krippendorf's alpha (-0.855) shows that there is no agreement between AntMover and AWA. On the other hand, both have strong disagreement. In addition, the Cohen's kappa value is 0 and this proof that the agreement in annotating same sentence is happened purely by chance (no agreement). In short, there is no inter-rater reliability between these tools.

## 4.3: Test 3: Krippendorf's alpha and Cohen's Kappa for reliability over sub-corpora

This test the reliability of AntMover and AWA over each document category

| Category | Subject | Agreement (%) | Krippendorf's alpha | Cohen's kappa |
|---|---|---|---|---|
| Mathematics | 4206 | 3.9 | -0.925 | 0 |
| Medicine | 2071 | 10.6 | -0.808 | 0 |
| Earth Science | 2590 | 9.19 | -0.831 | 0 |
| Chemistry | 1680 | 6.31 | -0.881 | 0 |
| Astronomy | 4318 | 9.5 | -0.826 | 0 |
| Computer Science | 5820 | 8.38 | -0.845 | 0 |
| Agriculture | 2972 | 6.7 | -0.874 | 0 |
| Materials Science | 2118 | 6.66 | -0.875 | 0 |
| Engineering | 3049 | 8.36 | -0.845 | 0 |
| Biology | 2106 | 8.97 | -0.835 | 0 |
| Total | 30930 | | | |

**Result**: In general, all categories have negative values in Krippendorf's alpha and 0 values in Cohen's kappa. This means both tools are in disagreement in annotating sentences in each category. The category with highest disagreement is Mathematics which accounted with -0.925 in Krippendorf's alpha value. Based on the percentage agreement, Medicine category has the highest agreement (10.6%) follow by Astronomy (9.5%).

# 5.0 AntMover and AWA association

## 5.1 Test 1: Association between AntMover's moves and AWA's annotations

The test will compare the association between AntMover's moves and AWA's annotations.

Null hypothesis, $H_0$ : AntMover's moves and AWA's annotations are independent

Alternate hypothesis, $H_1$ : AntMover's moves and AWA's annotations are dependent

| AntMover's moves | AWA's annotations | | | | | | | | | Not annotated by AWA but have Antmover's move |
|---|---|---|---|---|---|---|---|---|---|---|
| | Background | Contrast | Empahsis | MainCategory | Novelty | Position | Question | Surprise | Trend | |
| Claiming_centrality | 0 | 2 | 3 | 0 | 0 | 1 | 0 | 0 | 0 | 23 |
| Announcing_principal_findings | 50 | 798 | 134 | 28 | 43 | 37 | 56 | 16 | 37 | 17095 |
| Evaluation_of_research | 45 | 526 | 151 | 12 | 26 | 58 | 41 | 12 | 19 | 5460 |
| Making_topic_generalization | 42 | 175 | 35 | 3 | 10 | 7 | 21 | 4 | 18 | 3446 |
| Indicating_a_gap | 0 | 9 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 86 |
| Announcing_present_research | 6 | 204 | 15 | 14 | 23 | 3 | 8 | 4 | 4 | 2411 |

Chisquare test result:

**X-squared = 176.7, df = 40, p-value < 2.2e-16**

**Result**: The p-value is smaller than 0.05, so reject null hypothesis. There is a significant association between AntMover's moves and AWA's annotations. The chisquare score is 176.7. The values of contributing cells to the total Chi-square score is shown at section 5.2.

## 5.2 Test 2: Positive and negative association between AntMover and AWA

Cells with the highest absolute standardized residuals contribute the most to the total Chi-square score. The negative sign means negative association.

| AntMover's moves | AWA's annotations | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Background | Contrast | Empahsis | MainCategory | Novelty | Position | Question | Surprise | Trend |
| Claiming_centrality | -0.564 | -0.926 | 2.59 | -0.356 | -0.476 | 1.576 | -0.529 | -0.283 | -0.419 |
| Announcing_principal_findings | -1.689 | 1.357 | -1.34 | 0.538 | -0.336 | -1.463 | 0.012 | 0.006 | 0.328 |
| Evaluation_of_research | -0.306 | -1.623 | 3.723 | -1.564 | -1.311 | 3.907 | -0.078 | 0.041 | -1.376 |
| Making_topic_generalization | 6.204 | -1.756 | -0.719 | -1.414 | -0.548 | -1.524 | 1.647 | -0.096 | 2.896 |
| Indicating_a_gap | -0.763 | 0.766 | -0.324 | -0.842 | -0.644 | -0.657 | -0.716 | -0.383 | 1.196 |
| Announcing_present_research | -2.301 | 1.929 | -3.411 | 3.315 | 3.805 | -2.417 | -1.41 | 0.132 | -1.471 |
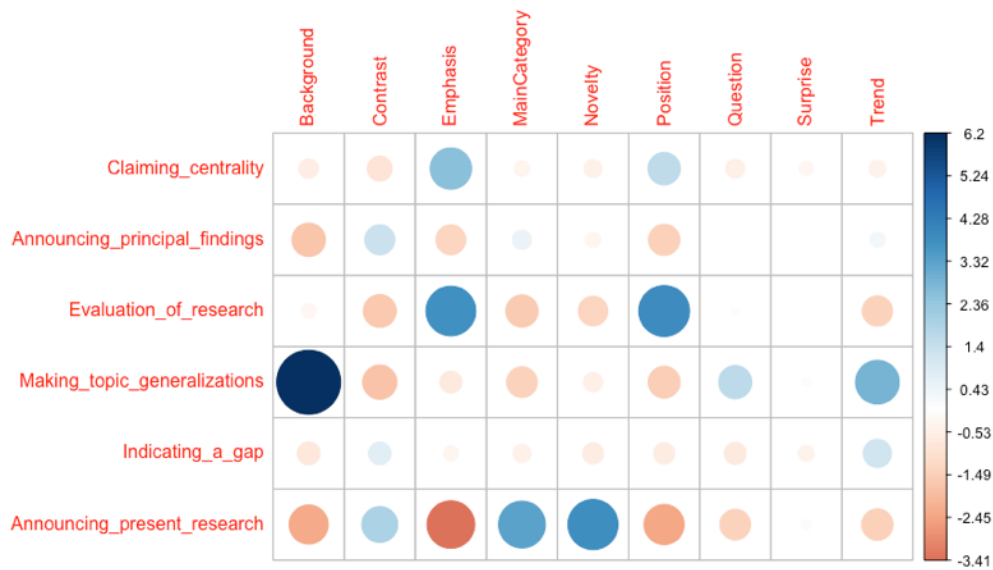
**Result**: The blue colour represents positive association while orange colour represents negative association (no association in this context). The volume of the 'round' shape represents the strength of the association. From the picture, the following can be derived

a. Strong positive association:

- 'making topic generalization' → 'background'
- 'trend' → 'background'
- 'evaluation of research' → 'emphasis'
- 'evaluation of research' → 'position'
- 'annonucing present research' → 'novelty'
- 'annonucing present research' → 'mainCategory
- 'claiming centrality' → 'emphasis'

b. Strong negative association

- 'announcing present research' → 'emphasis'
- 'announcing present research' → 'novelty'
- 'announcing present research' → 'background'

## 5.3 Test 3: Degree of association (%) between AntMover and AWA

The contribution in percentage just shows the 'positive' contribution.

| AntMover's moves | AWA's annotations (%) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Background | Contrast | Empahsis | MainCategory | Novelty | Position | Question | Surprise | Trend |
| Claiming_centrality | 0.18 | 0.485 | 3.797 | 0.072 | 0.128 | 1.406 | 0.158 | 0.045 | 0.099 |
| Announcing_principal_findings | 1.615 | 1.042 | 1.016 | 0.0164 | 0.064 | 1.212 | 0 | 0 | 0.061 |
| Evaluation_of_research | 0.053 | 1.491 | 7.843 | 1.384 | 0.972 | 8.638 | 0.003 | 0.001 | 1.072 |
| Making_topic_generalization | 21.779 | 1.745 | 0.293 | 1.132 | 0.17 | 1.315 | 1.534 | 0.005 | 4.748 |
| Indicating_a_gap | 0.329 | 0.332 | 0.059 | 0.131 | 0.235 | 0.244 | 0.29 | 0.083 | 0.81 |
| Announcing_present_research | 2.995 | 2.105 | 6.586 | 6.221 | 8.193 | 3.305 | 1.125 | 0.01 | 1.224 |
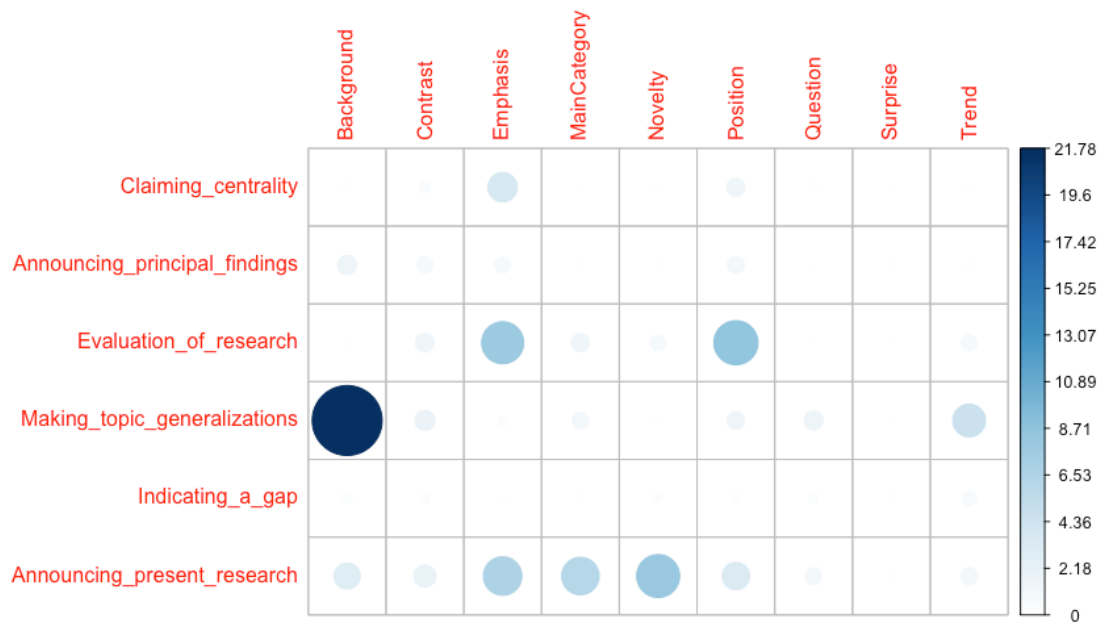


**Result**: This picture can be interpreted the same way as above. 'Making topic generalization' is significantly associated with 'background'. In general sense, this is true as most of the content in a 'background' category of an article generalizes or the topic of the article. On the other hand, the 'question' and 'surprise' categories have zero association with 'announcing principal findings'. 'Announcing principal findings' should have strong association with 'emphasis' or 'background' in AWA annotation. As an example, a sentence with 'principal findings' should be recognized by AWA as 'emphasis' or 'background' of certain research/facts. Furthermore, another category to be concerned is the 'indicating a gap'. This category has very weak association in all AWA annotations. It supposes to have high association in 'question', 'emphasis', 'position' or 'trend'.

## 5.4 Test 4: Association between AntMover and AWA over sub-corpora

This test is similar to test 5.1 but it restricted to category of document (sub-corpora). The result of the chisquare for each category of documents is as follow.

| Category | Total | Chisquare p-value | Null Hypothesis | AntMover and AWA Independent (Yes/No) |
|---|---|---|---|---|
| Mathematics | 171 | 0.261 | accept | Yes |
| Medicine | 248 | 0.004 | reject | No |
| Earth Science | 268 | 0.293 | accept | Yes |
| Chemistry | 133 | 0.459 | accept | Yes |
| Astronomy | 464 | 6.85E-05 | reject | No |
| Computer Science | 524 | 0.002 | reject | No |
| Agriculture | 233 | 0.424 | accept | Yes |
| Materials Science | 156 | 0.342 | accept | Yes |
| Engineering | 284 | 0.015 | reject | No |
| Biology | 221 | 0.345 | accept | Yes |

**Result**: Even though in general AntMover and AWA are dependent based on test result 5.1, the result is different categories. As an example, AntMover and AWA are not associated in 'Mathematics', 'Earth Science', 'Chemistry', 'Agriculture', 'Material Science' and 'Biology'. This implies that some categories of documents are not working well with AntMover or AWA.

## References

Fulcher, G. (2010). Practical Language Testing. London: Hodder Education.

Krippendorff, K. (2004). Content analysis: An introduction to its methodology (2nd ed.). Thousand Oaks, CA: Sage.