Supplementary Information:
Automatic recognition of conceptualisation zones in scientific articles and two life science applications

Maria Liakata, Shyamasree Saha, Simon Dobnik, Colin Batchelor and Dietrich Rebholz-Schuhmann

Contents

1 Definition of categories in the CoreSC scheme
   1.1 Distinguishing between different types of Objective, Approach and Outcome ........................................ 1

2 Examples of annotated categories in the CoreSC scheme ................................................................. 1

3 Information about papers used as training/testing data ................................................................. 1

4 Extra Information about features used in classifying CoreSC categories ....................................... 2
   4.1 Section headings ........................................................................................................................................... 2
   4.2 Verb classes .................................................................................................................................................. 2
   4.3 Examples of significant ngrams and GR triples ....................................................................................... 3

5 Examples of errors in categorisation ..................................................................................................... 4

6 Significance tests for the relation between classifier confidence and agreement between manual annotation & classifiers .............................................................................................................................................. 5

7 Further explanation of boxplots and confusion matrix ......................................................................... 7

8 Combining binary classifiers ..................................................................................................................... 7

References ..................................................................................................................................................... 8

List of Tables

1 Definitions of CoreSC categories ..................................................................................................................... 1
2 Example annotations of CoreSC categories ..................................................................................................... 2
3 Journals, topics and corresponding number of articles in the corpus ..................................................................... 3
4 Heading types and regular expressions ........................................................................................................... 3
5 Verb classes ..................................................................................................................................................... 4
6 Example of significant ngrams and grammatical triples. Here the ranking reflect the usefulness of these features in classifying the Method category, as determined by Liblinear. The feature @@@.@@@ stands for any three digit decimal number. The expression “ncsubj ... obj” signals a passive construction. So “ncsubj observe those obj” correspond to “those were observed”. In the grammatical relations, the first entity denotes the type of relation, the second the predicate and the third the argument which plays the role of subject, direct object etc., depending on the type of relation. Thus, “ncsubj predict we” corresponds to “we predict”, where ‘predict’ is the predicate and ‘we’ is the subject. ............................................................................................................................................. 5
7 Example of categorisation errors ..................................................................................................................... 6
8 Details of significance tests ............................................................................................................................. 7
1 Definition of categories in the CoreSC scheme

For more detailed definitions and examples also see the full 45-page annotation guidelines (Liakata and Soldatova, 2008).

| Category               | Description                                                                 |
|------------------------|-----------------------------------------------------------------------------|
| Hypothesis             | A statement not yet confirmed rather than a factual statement               |
| Motivation             | The reasons behind an investigation                                         |
| Background             | Generally accepted background knowledge and previous work                  |
| Goal                   | A target state of the investigation where intended discoveries are made     |
| Object-New             | An entity which is a product or main theme of the investigation             |
| Object-New-Advantage   | Advantage of an object                                                      |
| Object-New-Disadvantage| Disadvantage of an object                                                   |
| Method-New             | Means by which authors seek to achieve a goal of the investigation          |
| Method-New-Advantage   | Advantage of a Method                                                       |
| Method-New-Disadvantage| Disadvantage of a Method                                                    |
| Method-Old             | A method mentioned pertaining to previous work                               |
| Method-Old-Advantage   | Advantage of a Method                                                       |
| Method-Old-Disadvantage| Disadvantage of a Method                                                    |
| Experiment             | An experimental method                                                      |
| Model                  | A statement about a theoretical model or framework                          |
| Observation            | the data/phenomena recorded in an investigation                            |
| Result                 | factual statements about the outputs of an investigation                    |
| Conclusion             | statements inferred from observations & results relating to research hypothesis|

Table 1: Definitions of CoreSC categories

1.1 Distinguishing between different types of Objective, Approach and Outcome

The CoreSC scheme distinguishes objectives into Hypothesis-Goal-Motivation-Object, the main approach into Method-Model-Experiment and outcomes into Observation-Result-Conclusion.

The above distinction is important to expert needs. Hypothesis (a statement to be proven or refuted or an assumption which constitutes a stepping stone of an investigation) provides very different information to a Goal (what the authors aim to achieve), Object (an entity e.g. a gene which constitutes the focal point of an investigation) and the underlying knowledge gap which necessitates the investigation (Motivation). Even though Hypothesis, Goal, Motivation and Object typically cover a small percentage of a paper (2%, 1%, 1% and 3% respectively), they play a crucial role in communicating the essence of a scientific investigation. Thus, the challenge lies in identifying and distinguishing between these categories despite the limited data availability. It is also important to distinguish between different types of outcomes, as they involve different levels of inference. Observations are linked to direct measurement or experience of a phenomenon, while a Result is an outcome inferred from Observations using comparisons or analytical thinking. Conclusions involve a further level of abstraction from Observations and Results, and are expected to contain the answers to Hypotheses. The distinction between Method, Model and Experiment allows one to single out respectively high level mentions of methodology, theoretical frameworks employed and specific experimental steps.

2 Examples of annotated categories in the CoreSC scheme

Examples annotations are given for each category (Table 2). Where possible neighbouring sentences from the same paper were chosen to better illustrate the relationship between categories (for example note the build up from Observation to Conclusion). Square brackets give the paper id and the sentence number, separated by an underscore. More examples are given in Liakata and Soldatova (2008) and also in the annotated online ART/CoreSC corpus:

http://www.ukoln.ac.uk/projects/ART_Corpus/html/

3 Information about papers used as training/testing data

Articles in our training data (the ART/CoreSC corpus) represents 16 different journals and cover 25 topics. Table 3 shows the journals, topics and the number of articles per journal and topic respectively.
Next we asked, whether BOB.1/OBF.1 is able to form ternary complexes together with Oct1 and Oct2 on the murine and human Btk promoters." [N.A.1421503,106]

"To get more insights into the physiological role of BOB.1/OBF.1 the transcriptional network controlled by BOB.1/OBF.1 needs to be elucidated." [N.A.1421503,153]

"The regulation of B cell type-specific gene expression plays an essential role for B cell generation, maturation, development and function. A large number of transcription factors have been shown to be involved in the regulation of B cell fate (60)" [N.A.1421503,150,151]

"Therefore the aim of this study is to demonstrate that existing models can be improved when expanded to a broader range of targets." [b404632e,24]

"In this study we report the synthesis of a new class of biomolecular probe called a SERRS Beacon, which uses the detection technique of surface enhanced resonance Raman scattering, SERRS.” [b506219e,21]

"It offers the option to use fluorescence if desired.” [b506219e,22]

"In an alternative approach the benzotriazole dye maleimide was used as the target for the nucleophilic attack of an alkyl thiol attached to an oligonucleotide (Scheme 1).” [b506219e,89]

"This approach is simple and effective to use and offers a straightforward method for adding benzotriazole dyes to oligonucleotides.” [b506219e,91]

"Microfiltration(20) and dielectrophoresis(11,21) have been used to integrate blood sample preparation with PCR on microfabricated devices.” [b409366f,16]

"The microfilters were prone to clogging by blood samples, while the dielectrophoresis studies used only diluted or resuspended blood cells within a specific separation buffer, that was required to adjust the conductivity.” [b409366f,18]

"The usefulness of the B3LYP/6-31++G** method to describe intra- and intermolecular hydrogen bonds has been demonstrated in recent studies through comparison with the second order Muller-Plesset (MP2) predictions. (31,39)” [b406455e,71]

"A Renishaw inVIA microscope system with a 514.5 nm argon ion laser and a Renishaw microscope system 1000 with a 632.9 nm helium-neon laser were used. Samples were analysed using a 20 objective to focus the laser beam into a microtitre plate containing the sample.” [b506219e,51,52]

"It was possible to obtain SERRS signals from all 12 of the labels in a quantitative manner allowing detection limits to be obtained which were in the order of 1 1012 mol dm3.” [b506219e,76]

"When the detection limits of SERRS, using a commercially available Raman spectrometer, were compared with those obtained using standard quantitative PCR instrumentation, it was found that SERRS was generally at least three orders of magnitude more sensitive than fluorescence (24). The results shown in Table 1 illustrate this.” [b506219e,75,76]

"They indicate that SERRS could be adapted to detect commercially available labels which are already routinely used in methods which currently rely on fluorescence detection, with very high sensitivity.” [b506219e,77]

4 Extra Information about features used in classifying CoreSC categories

4.1 Section headings

A section is taken to be the text between two headings. In the case of nested headings (e.g. title and subtitles) we consider the outermost heading. For our Struct-3 feature we considered the 16 headings in Table 4, which are retrieved by matching the corresponding regular expressions.

4.2 Verb classes

We employed the verb classes in Table 5 as features for the classification. These were obtained automatically using clustering methods described in Sun and Korhonen (2009):
| Journal                                           | Num.of.articles | Topic          | Num.of.articles.1 |
|--------------------------------------------------|----------------|----------------|-------------------|
| Physical Chemistry Chemical Physics (CP)         | 180            | theoretical    | 57                |
| Faraday Discussions (FD)                         | 18             | spectroscopy   | 43                |
| Chemical Communications (CC)                     | 13             | kinetics       | 26                |
| Photochemical & Photobiological Sciences (PP)    | 11             | biochemistry   | 19                |
| Lab on a Chip (LC)                               | 9              | biophysical    | 13                |
| Dalton Transactions (DT)                         | 8              | electrochemistry| 13                |
| PhysChemComm (QU)                                | 7              | photochemistry | 13                |
| Organic & Biomolecular Chemistry (OB)            | 5              | solid state    | 13                |
| The Analyst (AN)                                 | 3              | catalysis      | 10                |
| Journal of Environmental Monitoring (EM)         | 3              | microfluidics  | 10                |
| New Journal of Chemistry (NJ)                    | 2              | analytical     | 8                 |
| CrystEngComm (CE)                                | 2              | materials      | 7                 |
| Journal of Atomic Analytical Spectrometry (JA)   | 1              | inorganic      | 6                 |
| Journal of the Chemical Society, Perkin Transactions 1 (P1) | 1 | soft matter | 6 |
| Green Chemistry (GC)                             | 1              | synthetic      | 6                 |
| Journal of Materials Chemistry(JM)               | 1              | surfaces       | 5                 |
|                                                 |                | thermodynamics | 2                 |
|                                                 |                | chemoinformatics| 1                 |
|                                                 |                | chemical engineering| 1 |
|                                                 |                | crystallography| 1                 |
|                                                 |                | environmental  | 1                 |
|                                                 |                | imaging        | 1                 |
|                                                 |                | ionic liquids  | 1                 |
|                                                 |                | NMR            | 1                 |
|                                                 |                | statistical mechanics| 1 |

Table 3: Journals, topics and corresponding number of articles in the corpus

| Heading type | Regular Expression       |
|--------------|--------------------------|
| abstract     | /.+abstract.*|i|                         |
| appendix     | /.+appendix.*|.acknowledgement.*|i| |
| introduction | /.+intro.*i|                           |
| background   | /.+background.*|.previous.*|i|                        |
| model        | /.+model.*|i|                            |
| theory       | /.+theor.*i|                                  |
| calculation  | /.+calculation.*|i|                          |
| computational| /.+computational.*|i|                        |
| experiment   | /.+experiment.*|i|                                |
| result       | /.+result.*|i|                                           |
| method       | /.+method.*|.mechanism.*|i|                                      |
| summary      | /.+summar.*|i|                                      |
| conclusion   | /.+conclu.*|i|                                         |
| discuss      | /.+discuss.*|i|                                          |
| specific     | /.+/specific|g|                                         |
| none         | /none/i                  |

Table 4: Heading types and regular expressions

4.3 Examples of significant ngrams and GR triples

In Table 6 we present examples of significant unigrams, bigrams and GR dependency triples, as ranked by the Liblinear algorithm (Fan et al., 2008). Significance was determined according to the absolute value of weights assigned to each feature by Liblinear (|wj|). According to Chang and Lin (2008), ranking results according to |wj| identifies the same features as being important as does considering how the performance is influenced when the features in question are excluded (leave-one-feature-out).
### Table 5: Verb classes

| class1 | class2 | class3 | class4 | class5 | class6 | class7 | class8 | class9 | class10 |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |
|         |        |        |        |        |        |        |        |         |        |
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |
|         |        |        |        |        |        |        |        |         |        |
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |
|         |        |        |        |        |        |        |        |         |        |
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |
|         |        |        |        |        |        |        |        |         |        |
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |
|         |        |        |        |        |        |        |        |         |        |
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |
|         |        |        |        |        |        |        |        |         |        |
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |
|         |        |        |        |        |        |        |        |         |        |
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |
|         |        |        |        |        |        |        |        |         |        |
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |
|         |        |        |        |        |        |        |        |         |        |
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |
|         |        |        |        |        |        |        |        |         |        |
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |
|         |        |        |        |        |        |        |        |         |        |
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |
|         |        |        |        |        |        |        |        |         |        |
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |
|         |        |        |        |        |        |        |        |         |        |
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |
|         |        |        |        |        |        |        |        |         |        |
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |
|         |        |        |        |        |        |        |        |         |        |
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |
|         |        |        |        |        |        |        |        |         |        |
| give   | involve| provide| kcal   | contain| carry  | yield  | stretch| represent| reflect |

#### 5 Examples of errors in categorisation

Table 7 gives examples of common categorisation errors, to better illustrate the challenge in classifying the categories. In the case of Table 7:1(A), The advantage of an old Method is annotated as Background. This is not an unlikely mistake to make as the sentence does pertain to previous work. In the second case, the classifiers assigned this sentence to BAC with a confidence of 0.95. Indeed, this sentence pertains to details of previous work and not actually an old Method.

Another common error arises from the fact that Goals and Methods are often conveyed together in the same sentences. This is shown in examples Table 7:2 (A) and 2 (B). The classifiers captured the Method part of these sentences rather than the Goal part. Allowing multiple labels per sentences would rectify this.

In Table 7:3 (A) a sentence containing both a Goal and an Object is classified as Object rather than Goal by the classifiers. For the sentence in 3(B), however, to which the classifiers assigned Object with a high confidence of 0.84, this is indeed an Object as it refers to what the investigation is about, not what the authors aim to achieve.

In Table 7:4 (A) the classifiers label the sentence as Result instead of Conclusion, whith a low confidence of 0.53. So this case is considered difficult by the classifiers, presumably because it contains mixed indications such as the verb ‘conclude’ but also a figure, usually associated with Result or Observation. For 4(B), however, the classifiers assigned Result with a high confidence of 0.94 while the human expert annotated it as Conclusion. This sentence contains mention of results and a comparison rather than a more elaborate conclusion. In the guidelines human annotators were told that a comparison is usually an indication of a Result rather than a Conclusion.

In Table 7:5 (A) the classifiers label the sentences as Observation rather than Result with a low confidence of 0.45. Indeed, this sentence contains a comparison so it should be a Result rather than a direct Observation. 5 (B) which was assigned to Observation with a high confidence of 0.81 contains both elements of an Observation (e.g. measurement units and the word ‘observed’) but it is primarily about a Result, the assignment of these observations to particular elements.

In Table 7:6 (A) both classifiers agreed in the annotation of the sentence as Method with a medium confidence (0.69) but the label assigned by annotators was Object. Here the sentence has actually both an Object part (the interconnectivity between different Xe adsorption regions) and a Method part (examined by 2D EXSY spectroscopy). In the second example, 6(B), classifier confidence was low (mean of 0.45) in assigning Method to this instance. This sentence is indeed an Object as it focusses on the...
Table 6: Example of significant ngrams and grammatical triples. Here the ranking reflects the usefulness of these features in classifying the Method category, as determined by Liblinear. The feature @@@.@@@ stands for any three digit decimal number. The expression “ncsubj ... obj” signals a passive construction. So “ncsubj observe those obj” corresponds to “those were observed”. In the grammatical relations, the first entity denotes the type of relation, the second the predicate and the third the argument which plays the role of subject, direct object etc., depending on the type of relation. Thus, “ncsubj predict we” corresponds to “we predict”, where ‘predict’ is the predicate and ‘we’ is the subject.

As an example of Hypothesis wrongly annotated as Conclusion in Table 7: 7(A), the classifiers agreed in annotating this sentence as a Conclusion with high classifier confidence (mean of 0.9) and indeed this is a weak Conclusion drawn from some outcome mentioned previously. The human annotator must have been confused by the use of “may” in this sentence, which is often used to express Hypotheses. The example in 7(B) was annotated by both classifiers as Conclusion with medium confidence (0.75) but was in fact a Hypothesis, as the statement was followed by further investigations. However, had it been at the end of the document it could well have been a weak Conclusion.

An other common error is Hypothesis wrongly annotated as Result. In Table 7, 8 (A) was annotated as Result rather than Hypothesis with high confidence (0.78). This sentence indeed contains elements that point to a result (e.g. the word ‘result’) but its role in the paper was that of a stepping stone, as it resulted in more investigations. Viewed on its own, however, it is difficult to ascertain if this is a Hypothesis or a weak Conclusion, and part of it definitely includes Result. Example 8 (B) was assigned Result by the classifiers also with high confidence (0.75) but is a Hypothesis, consisting of various conditions and parts.

6 Significance tests for the relation between classifier confidence and agreement between manual annotation & classifiers

We conducted significance tests to assess if classifier confidence and agreement between classifiers and manual annotation are correlated. We considered:

(A) A Welch Two Sample t-test for the values of classifier confidence in cases of agreement between classifiers and manual annotation and in cases of disagreement.
| Type of error | Actual sentence |
|---------------|-----------------|
| 1. MET annotated as BAC: | (A) “Unlike conventional $^{129}$Xe NMR in which the nuclear spin polarization is governed by Boltzmann equilibrium, laser-polarized $^{129}$Xe NMR facilitates a metastable polarization transfer from the alkali-metal electronic spin to the $^{129}$Xe nucleus by the spin exchange optical pumping method, thus capable of enhancing signal sensitivity by 103105 folds even at dilute Xe loading.” [5062806,17]  
(B) “Examples of sensing principles explored in the measurement of microfluidic flow are heat transfer detection, 24 molecular sensing, (5) atomic emission detection, (6) streaming potential measurements, (7) electrical impedance tomography, (8) ion-selective field-effect transistors (9) and periodic flapping motion detection.” [510282c,11] |
| 2. GOA annotated as MET: | (A) “In order to investigate if the in situ deposit is metallic or not, current/voltage tunneling spectroscopy was conducted.” [502789f,138]  
(B) “The purpose of this paper is to reduce the uncertainty of the theoretical predictions for the above-discussed spectroscopic and thermochemical properties of the HCCI and CCl2 radicals by using sophisticated theoretical techniques.” [506790a,75] |
| 3. GOA annotated as OBJ: | (A) “In order to investigate the role of the photo-induced reduction of $[\text{RuIII(edta)}\text{NO}]$ to $[\text{RuII(edta)}\text{NO}]$, the reaction of $\text{RuII}$ with NO was studied under the same experimental conditions.” [b108236c,145]  
(B) “Here we report preliminary results from NMR-spectroscopy on PsbO from Thermosynechococcus elongatus (T. elongatus) describing its structural and dynamic properties.” [b407316a,23] |
| 4. CON annotated as RES: | (A) “Consequently, it can be concluded that the exchange rates between different environments follow the following trend: $\text{kgm} > \text{kmo} > \text{kgo}$, where $\text{kmo}$ denotes the exchange rate between mobile Xe in the mesopores and Xe residing in the organic phase and $\text{kgm}$ and $\text{kgo}$ denote exchange rates between gaseous-mobile Xe and gaseous-organic phase Xe, respectively, as illustrated in Fig. 5.” [5062806,76]  
(B) “Furthermore, the uncertainties of the present results are smaller by a factor of 25 than the uncertainties of even the best of the former theoretical values, and more than an order of magnitude smaller than $\Delta f_{H0}$ given in the available thermochemical databases.” [506790a,245] |
| 5. RES annotated as OBS: | (A) “In fact, $\chi_{xx}$ and $\chi_{yy}$ differ by only a few percent.” [b314320a,123]  
(B) “The observed three diagonal peaks at ca. 0, 87, and 98 ppm in the vertical (F2) axis were assigned (vide supra) to gaseous Xe, mobile Xe adsorbed in the mesopores, and Xe residing in the organic phase, respectively.” [b5062806,71] |
| 6. OBJ annotated as MET: | (A) “The interconnectivity between different Xe adsorption regions in the mesopores of AEP-MP was further examined by 2D EXSY NMR spectroscopy (Fig. 4ac).” [b506808b,68]  
(B) “We have also calculated the energies for H adsorption on the FeMoco for the same configurations, although they already have been calculated elsewhere. A,B,C.” [b310850c,223] |
| 7. HYP annotated as CON: | (A) “Such a behaviour may be expected for larger oxygenated alkoxy radicals.” [b313251j,164]  
(B) “This could be due to the fact that, because of the low solubility of NO in water, only a small range of NO concentrations could be investigated.” [b108236c,133] |
| 8. HYP annotated as RES: | (A) “In view of our results (G2 and CCSD(T)//MP2 in Table 2), this 1,6-isomerisation channel through a 7-membered transition state may be energetically competitive with the $\text{-CH scission.”}” [b313251j,100]  
(B) “If the motion of the HBr sub-unit in the xy plane is assumed to be circular in reasonable approximation and if the electric field gradient tensor at Br in free HBr is assumed to be unaffected by the presence of the thiophene sub-unit (i.e. any electronic redistribution in HBr is ignored and the electric field gradient (EFG) at Br arising from the presence of the thiophene electric charge distribution is negligible), a rough estimate of the extent of the zero-point angular excursion of the HBr sub-unit from its equilibrium position can be made.” [b314320a,124] |

Table 7: Example of categorisation errors

(B) a direct Pearson’s product-moment correlation test between classifier confidence values and levels of agreement. We considered the following levels of agreement as integer values:
2: no agreement between classifiers or manual annotation
3: one classifier agrees with manual annotation but not both
4: classifiers agree with each other, but not with manual annotation
5: classifiers agree with each other and with manual annotation
Both tests were conducted using the R statistics package and showed significance at 99% level with a \( p-value < 2.2e-16 \). The Welch Two Sample t-test showed that in cases of agreement, the mean value of classifier confidence (represented as a probability) is at least 0.25 greater than in cases of disagreement. The Pearson’s product moment correlation showed a significant positive correlation between agreement and classifier confidence, with a coefficient of at least 0.54.

Details for both tests are available in Table 8.

### Further explanation of boxplots and confusion matrix

In the confusion matrix in the manuscript (Figure 7), rows indicate correct category values, as assigned in the ART/CoreSC corpus (Liakata and Soldatova, 2009), while columns show predicted categories as identified by LibSVM.

The boxplots in Figures 4-6 of the manuscript show classifier confidence in different scenarios: When the classifiers agree with the manual annotation (Figure 4), when the classifiers don’t agree with the manual annotation (Figure 5) and classifier confidence for the entire corpus (Figure 6). It is important to note that the trend in Figure 6 is more similar to the trend in Figure 4 (cases of agreement).

The lines in the middle of the boxes are median values, while the upper edge of the boxplot is the upper quartile (25% of instances have value greater than this) and the lower edge is the lower quartile (25% of data have value less that this). Circles in the boxplots denote outliers, that is values that are less than 3/2 of the lower quartile. The whiskers show minimum and maximum values with the exclusion of outliers.

### Combining binary classifiers

To provide a direct comparison with the multi-class classifier (which only assign one category label per sentence), the result from binary classifiers is combined to give a single label each time.

We follow the rules below in combining the binary classifiers:

- If a sentence receives 0 from all classifiers \( \rightarrow \) assign to ‘BAC’, the most wide-ranging category
- If a sentence receives 0 from all classifiers apart from one category \( \rightarrow \) assign that category
- If a sentence receives multiple categories \( \rightarrow \) apply priority of concepts and assign the prevailing category

---

| Welch Two Sample t-test | Pearson's product-moment correlation |
|-------------------------|---------------------------------------|
| \( t = 108.0809, df = 6335.361, p-value < 2.2e-16 \) | \( t = 128.0734, df = 39913, p-value < 2.2e-16 \) |
| alternative hypothesis: true difference in means is not equal to 0 | alternative hypothesis: true correlation is greater than 0 |
| 99 percent confidence interval: | 99 percent confidence interval: |
| 0.2525231 0.2648572 | 0.5313846 1.0000000 |
| sample estimates: | sample estimates: |
| mean of x mean of y | cor 0.5396893 |
| 0.8042673 0.5455772 | |

Table 8: Details of significance tests
The priority of concepts is the same as given to human annotators in the guidelines of [Liakata and Soldatova, 2008]. They favour categories over more frequent ones, in the case where more than one category is assigned to a sentence. The list of priorities from highest to lowest is the following:

\(<\text{Hypothesis}>\)
\(<\text{Goal}>\)
\(<\text{Motivation}>\)
\(<\text{Conclusion}>\)
\(<\text{Result}>\)
\(<\text{Model}>\)
\(<\text{Experiment}>\)
\(<\text{Method}>\)
\(<\text{Observation}>\)
\(<\text{Background}>\)

References

Chang, Y.-W. and Lin, C.-J. (2008). Feature ranking using linear svm. In *JMLR Workshop and Conference Proceedings: Causation and Prediction Challenge (WCCI 2008)*, volume 3, pages 53–64.

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J. (2008). LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9, 1871–1874.

Liakata, M. and Soldatova, L. (2008). Guidelines for the annotation of general scientific concepts. Aberystwyth University, JISC Project Report http://ie-repository.jisc.ac.uk/88/.

Liakata, M. and Soldatova, L. (2009). The ART Corpus. Technical report, Aberystwyth University.

Sun, L. and Korhonen, A. (2009). Improving verb clustering with automatically acquired selectional preferences. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 638–647, Singapore. Association for Computational Linguistics.